ANALOG AND MIXED-SIGNAL CIRCUITRY FOR SYSTEM-ASSISTED HIGH-SPEED I/O LINKS

BY

ADAM CLAY FAUST

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2012

Urbana, Illinois

Doctoral Committee:

Professor Elyse Rosenbaum, Chair
Professor Milton Feng
Professor Naresh Shanbhag
Professor José Schutt-Ainé

ABSTRACT

The state-of-the-art design methodology for high-speed I/O links is to specify component-level design requirements to achieve high-fidelity component-level performance. While designing each component in the link with high fidelity guarantees a reliable link, it does not inherently optimize the link for metrics such as the power, design complexity, or bit error rate performance. Recently, due to the increased demand for data bandwidth in backplane I/O, a system-assisted design methodology has been developed to optimize the system for a given set of metrics. By optimizing on the system level rather than the component level, the performance at the component level can be reduced from high quality to sufficient when the component is deployed within the I/O link. The new system-level design methodology encourages the utilization of novel circuit architectures. In this dissertation, novel analog and mixed-signal circuitry for system-assisted high-speed I/O links is presented. The novel circuitry expands upon traditional analog and mixed-signal circuit architectures in order to achieve system-level design goals and requirements without significant power or area overhead.

To my family and friends

ACKNOWLEDGMENTS

When I look back at the student I was upon entering graduate school and compare to the student I have become, I am astounded at the change. Along the path, I have been aided by so many, and I thank all who have been there for me. At this time, I would like to thank a few individuals who have had a significant impact on my studies.

I would like to thank my adviser, Professor Elyse Rosenbaum. Next, I would like to thank Professor Shanbhag for his dedicated involvement in the system-assisted mixed-signal (SAMS) project which I have been fortunate to participate in. For their friendship and assistance through the years, I would like to thank my colleagues Dr. Karan Bhatia, Dr. James Di Sarro, Dr. Farzan Farbiz, Dr. Nicholas Olson, Ankit Srivastava, Jeffrey Lee, Nathan Jack, Vrashank Shukla, Arjun Kripanidhi, Min-Sun Keel, Nick Thomson, Kuo-Hsuan Meng, Robert Mertens, Dr. Rajan Narasimha, Chhay Kong, Yingyan Lin, Pavle Milosevic, Wen-Yi Chen, and Dr. Richard Tseng. I would especially like to thank Karan and Ankit for guiding me through my first two years of learning how to design analog circuits. I would like to thank Dr. Hyeon-Min Bae and Jonathan Ashbrook for donating their time and sharing their industrial experience.

Thank you to the Semiconductor Research Corporation, Texas Instruments, the Department of Electrical and Computer Engineering, and the Micron Foundation for their financial support during my studies. Thank you to NVIDIA, Micron Technologies, Advanced Micro Devices, and Texas Instruments for hosting me for summer internships.

TABLE OF CONTENTS

CHAPTER 1

HIGH-SPEED I/O BACKGROUND

1.1 Motivation for High-Speed I/O

In today's media-rich environment, the amount of data being transmitted within modern electronic products has been increasing at a steady rate, increasing the throughput required in chip-to-chip backplane communication links. Due to limitations on board real estate and the number of package pins, the number of transmission lines between board components has been prevented from growing to meet the increased throughput demand. Therefore, to increase the data throughput, the data rate on the transmission lines has been increased. At these high-data rates, the signal integrity of the received signal is degraded due to the non-idealities of the channel and other parasitic elements, requiring additional circuitry with significant design effort and power to insure an acceptable bit error rate (BER) performance. Over the years, desirable component-level performance requirements were developed in order to insure that specific components did not limit the BER performance.

1.2 Serial I/O Links

A simplified block diagram of a high-speed serial I/O link can be seen in Figure 1.1; the link is composed of components located on two chips, chip A and chip B, and on the printed circuit board (PCB) on which the transmission line that connects the chips is located. Chip A and chip B may be physically connected to the board through the use of solder, through-hole mounting, or a socket. In many instances, chip A and chip B may be on separate PCBs, using connectors to complete the physical connection to the motherboard. Regardless of how the chips are physically connected, a transmission line is used to electrically connect chip A to chip B. In this simplified block diagram, the serial I/O link is not bi-directional, in that data are only sent in this link from chip A to chip B. As such, chip A has the transmitter and chip B has the receiver. The transmitter is composed of circuits which are specifically designed to convert a data stream into electrical

signals which are sent over the PCB transmission line to the receiver of chip B. The receiver is composed of circuits which are specifically designed to convert the received electrical signals into a reconstructed data stream.

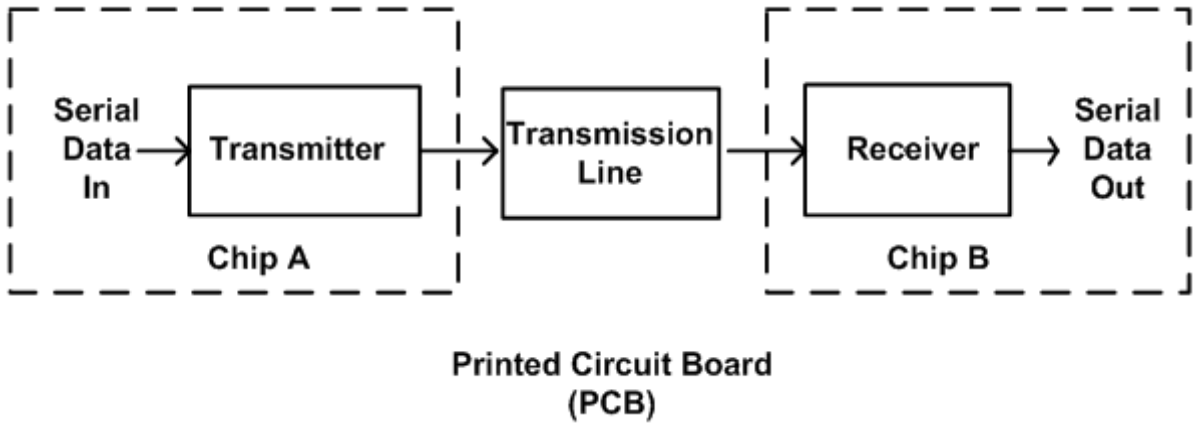


Figure 1.1: Basic high-speed serial I/O link block diagram.

Ideally, the data stream generated at the receiver matches the data stream at the transmitter, but the physical limitations discussed in this chapter may corrupt the process. Corruption of data in serial I/O links is measured in terms of BER, which is defined as the rate at which incorrect data bits appear in the reconstructed data stream generated by the receiver. For example, a BER of 10^{-12} signifies that one bit out of every trillion data bits is corrupt. While the BER requirement of a link varies from application to application, this dissertation defines highly reliable serial links as links with BER targets under 10^{-12}. For a 10 Gbps link operating at capacity, a BER of 10^{-12} results in an average of one bit error every 100 seconds. For the high-speed I/O links presented in this dissertation, a BER of 10^{-12} is the reliability performance requirement.

1.3 NRZ Signal Power Spectral Density

In serial I/O links, the simplest and most common signaling scheme is non-return-to-zero (NRZ). NRZ is the classical digital representation of signals; bits are represented with one of two discrete levels and transitions between levels only occur between symbols (bits). An example of NRZ signaling is shown in Figure 1.2 [1]. All data transmission in this dissertation is done with NRZ signaling.

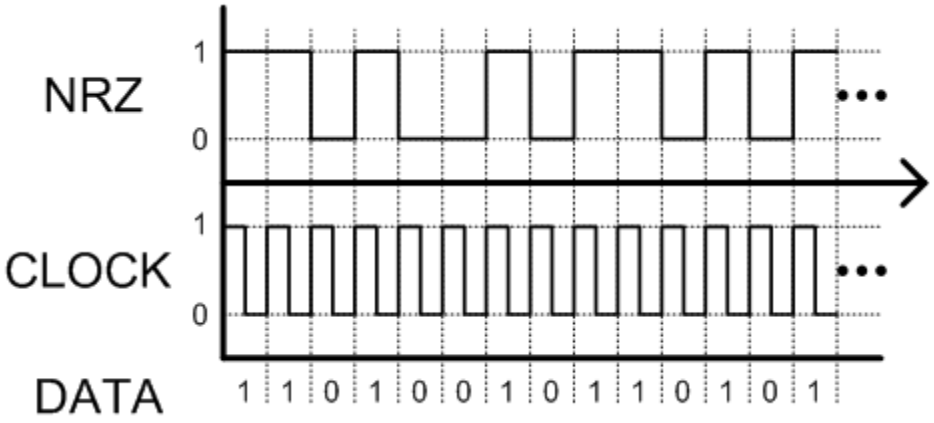


Figure 1.2: NRZ signaling [1].

For random NRZ data, the power spectral density is a sinc(*f*) function with nulls at integer multiples of the data (baud) rate. An example power spectral density for a pseudorandom bit sequence (PRBS) is shown in Figure 1.3 [1]. The results show that the signal is wideband and that a significant percentage of the energy is located at low frequencies.

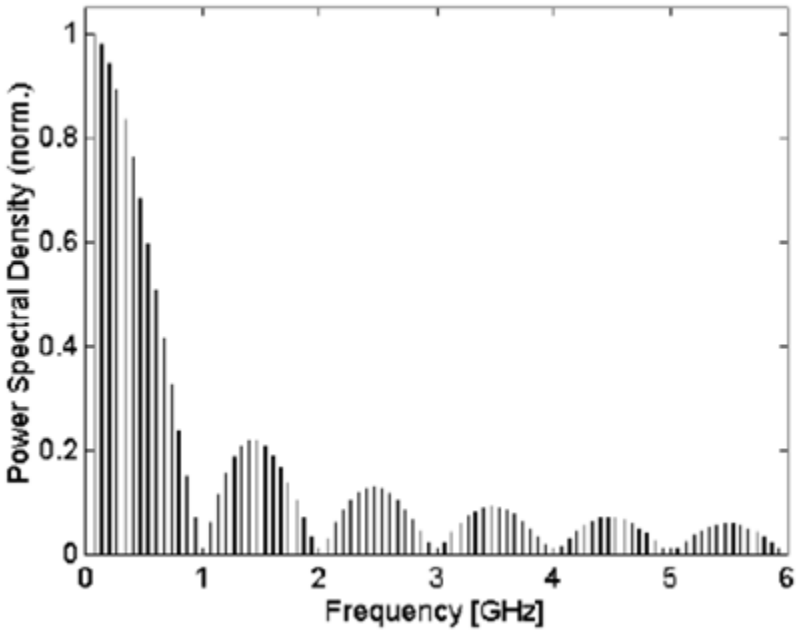


Figure 1.3: NRZ power spectral density for a 1 Gbps, 2^4-1 PRBS sequence [1].

1.4 Backplane Transmission Lines

Chip-to-chip communication over a backplane (a circuit board on which electric components and connectors to other circuit boards can be placed) utilizes a transmission line (or lines)

composed of a trace which is electromagnetically coupled to a return path. The two most common PCB backplane trace geometries are stripline and microstrip. Stripline, whose geometry can be seen in Figure 1.4, is a copper trace located between two copper reference planes which constitute the return path. Microstrip, whose geometry can be seen in Figure 1.5, is a copper trace on an exterior PCB layer that is coupled to a single copper reference plane which constitutes the return path. Basic PCBs use FR4, a dielectric, to separate the trace from the reference plane(s). Advanced dielectrics can be used to reduce high-frequency loss. Since the microstrip trace is located on an exterior PCB layer, the trace and adjacent dielectric are covered with a protective solder mask layer.

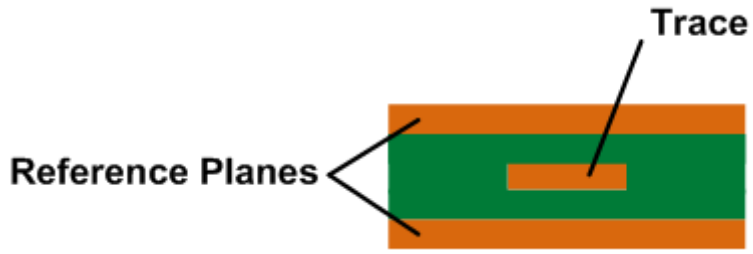


Figure 1.4: Stripline cross section.

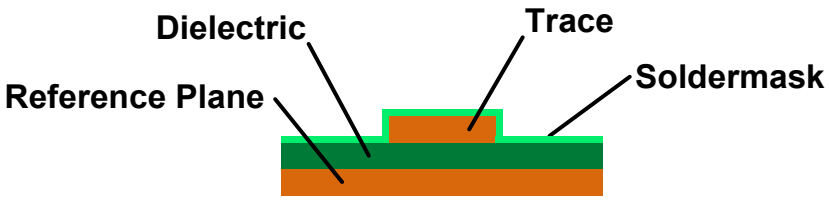


Figure 1.5: Microstrip cross section.

In this dissertation, grounded coplanar transmission lines, as shown in Figure 1.6, are utilized to minimize variation. Since the trace is coupled to multiple reference conductors, variation in dielectric thickness or spacing has a reduced effect on the electrical characteristics.

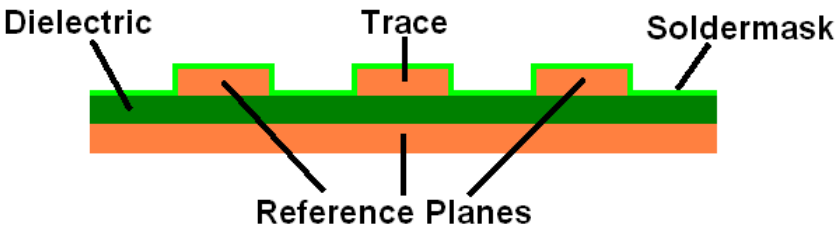


Figure 1.6: Grounded coplanar transmission line cross section.

For the stripline geometry, signals travel in the transverse electromagnetic (TEM) mode since the non-TEM modes have very high cut-off frequencies (>100 GHz) [2]. As such, we can generate a distributed model of the transmission line with the circuit model in Figure 1.7. The transmission line parameters for the circuit model can be found in Table 1.1.

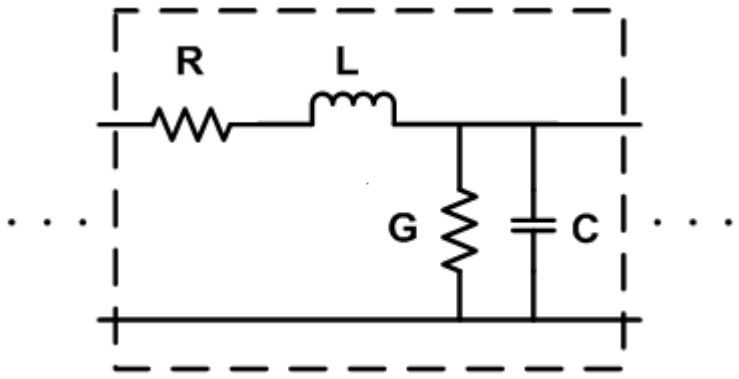


Figure 1.7: Transmission line circuit model.

Table 1.1: Transmission Line Parameter Descriptions

Parameter	Description
R	Resistance per Unit Length
L	Inductance per Unit Length
G	Conductance per Unit Length
C	Capacitance per Unit Length

The distributed model is used to generate the well-known telegrapher's equations in the frequency domain. The telegrapher's equations define the voltage and current waves at a location z along the line.

$$\frac{\partial V(z,j\omega)}{\partial z} = -(R + j\omega L) \cdot I(z,j\omega) \tag{1.1}$$

$$\frac{\partial I(z,j\omega)}{\partial z} = -(G + j\omega C) \cdot V(z,j\omega) \tag{1.2}$$

From (1.1) and (1.2), we generate several expressions which describe the response of the channel in the frequency domain. The characteristic impedance Z_o of the channel is the ratio of the forward propagating voltage wave to the forward propagating current wave.

$$Z_o = \sqrt{\frac{(R+j\omega L)}{(G+j\omega C)}} \tag{1.3}$$

When a propagating wave encounters a discontinuity or load, a reflected wave is generated with a magnitude scaled by the reflection coefficient

$$\Gamma = \frac{Z_L - Z_o}{Z_L + Z_o}, \tag{1.4}$$

where Z_L is the impedance of the discontinuity or load.

For a serial I/O, the impedances at the ends of the transmission line are those of the receiver and transmitter in a uni-directional link. In a bi-directional link, a transmitter and receiver are located at both ends of the transmission line. An increase in the magnitude of the reflection coefficient corresponds to an increase in the power reflected by the discontinuity or load. The propagation coefficient γ of the channel describes the wave propagation and attenuation.

$$\gamma = \sqrt{(R + j\omega L)\,(G + j\omega C)} \tag{1.5}$$

$$= \alpha + j\beta \tag{1.6}$$

For a lossy channel ($R \neq 0$ or $G \neq 0$), γ has a real component α, known as the attenuation coefficient, and an imaginary component β, known as the phase coefficient, as seen in (1.6). The term α describes the attenuation (in nepers per unit length) and β describes the phase change (in radians per unit length). In a lossless channel ($R = 0$ and $G = 0$), $\alpha = 0$ and β is a linear function of the angular frequency, ω. Therefore, the phase velocity (1.7) is constant in a lossless channel and frequency dependent in a lossy channel.

$$v_p = \frac{\omega}{\beta} \tag{1.7}$$

Unlike the stripline geometry, in the microstrip and grounded coplanar geometries the electromagnetic fields are not confined within the dielectric layer between the signal trace and the reference plane [2]. The field configuration therefore allows non-TEM propagation at significantly lower frequencies when compared to the stripline geometry [2]. This non-TEM propagation produces a frequency dependent phase shift which is referred to as microstrip dispersion.

At the frequencies which contain the vast majority of the signal power, which will hereafter be referred to as the frequencies of interest, three physical phenomena impact the channel response for both stripline and microstrip transmission lines at high frequencies: the skin effect, the proximity effect, and the dielectric loss [2].

The skin effect alters the current distribution within a conductor at high frequencies. As the frequency of the current increases, the current is confined to a decreasing depth δ below the surface of the conductor.

$$\delta = \sqrt{\frac{2}{\omega \mu \sigma}} \tag{1.8}$$

The decrease in the cross section of the current flow effectively increases the resistance per unit length at higher frequencies.

The change in current distribution also generates an inductive component whose impedance equals the increase in resistance [2]. The proximity effect also alters the current distribution within the conductors [2]. As alternating current flows in the conductors, a changing magnetic field is generated around the conductors as a result of Ampere's law. When the magnetic field of the trace interacts with the return path, the current of the return path is attracted toward the trace. Similarly, the current of the trace is attracted toward the return path. This change in current distribution results in an additional resistance per unit length which increases as the frequency increases.

Dielectric loss is the dissipation of signal power due to conduction current through the dielectric at high frequencies [2]. The dielectric loss is modeled as a shunt conductance which increases with frequency.

In high-speed I/O traces, the skin effect, the proximity effect, and dielectric loss change the attenuation coefficient α and vary the phase coefficient β. As a result of these phenomena, the attenuation coefficient becomes frequency dependent and increases as the frequency increases. For frequencies above 10 MHz, α increases proportionally to the square root of the frequency due to the skin effect [2]. Above 1 GHz, the attenuation coefficient increases proportionally to the frequency due to dielectric loss [2]. From (1.7), the variation of β due to these phenomena results in a dispersion of the signal since the phase velocity is not constant. The dispersion results in a spread of the signal energy in the time and spatial domains.

The effects of the attenuation and dispersion distort the received signal as it travels within the transmission line. The distortion of the signal at high data rates is significant and may limit the BER performance of the link if left unmitigated.

1.5 ESD Protection

1.5.1 Circuits

Electrostatic discharge (ESD) protection circuits and devices are utilized to protect integrated circuits against ESD events which occur during packaging, product assembly, and use. These ESD events can produce considerable voltages and currents. In integrated circuits without adequate ESD protection, the voltages and currents result in permanent damage to the devices. Since the transmitter and receiver have an external interface, they are subject to damage from ESD events. For example, without ESD protection circuitry, events such as human handling may damage the transmitter or receiver. The level of protection the ESD devices provide is dependent on many factors, including ESD device type, ESD device size, and the routing of discharge paths.

The drawback of the protection devices is the off-state parasitics they present to the external interfaces they protect. The dominating parasitic of ESD protection devices is a shunt capacitance from the external interface to the substrate in the range of several hundred femtofarads to a couple of picofarads. At high frequencies, the parasitic capacitance modifies the effective impedance at the end of the transmission line, degrading the termination match. From (1.4), this change will cause the reflection coefficient to deviate from zero, resulting in reflection of the signal energy from the receiver. Consequently, the received signals for high-speed I/Os will be distorted, with the magnitude of the distortion dependent on the magnitude of the parasitic capacitance. To provide a higher level of protection, additional devices must be added or the sizes of the current devices must be increased. As a result, the magnitude of the parasitic capacitance increases as the protection level of the ESD devices increases. Additionally, in some ESD protection schemes, a series resistor may be present in the signal path. The impact of the series resistor on the signal integrity is investigated in Chapter 3.

A basic ESD protection scheme, consisting of a pair of primary dual-diodes, for a high-speed I/O is shown in Figure 1.8. The diodes and rail clamp (which can conduct bi-directionally) protect against ESD events which may occur at any of the pins by providing current conduction between the pins and clamping the voltage at the pins. Due to the possibility of positive and negative polarity stresses at the pins, the network is designed to conduct current in and out of the

pin as necessary. It is important to note that during standard operation, the diodes and rail clamp are inactive and the current conduction is limited to leakage, such that normal operation may occur.

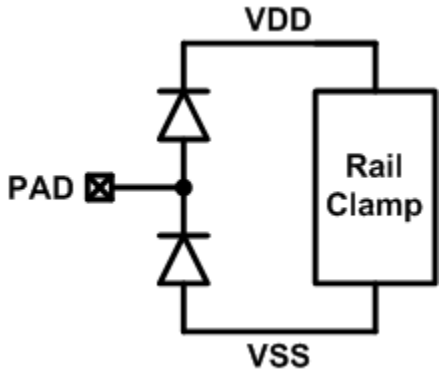


Figure 1.8: Dual-diode ESD protection scheme.

1.5.2 ESD Protection Event Standards and Testing

While many ESD protection event standards exists, the JEDEC and ESD Association (ESDA) standards and standard test methods are commonly used by industry when determining the ESD resiliency of integrated circuits. The Industry Council on ESD Target Levels investigated ESD related failures for billions of shipped parts and utilized the information to derive protection level recommendations as a function of the ESD controls utilized during packaging and assembly of the integrated circuits [3], [4]. Specifically, the Industry Council has resiliency requirements for both the human body model (HBM) and charged device model (CDM) [3], [4].

The human body model (HBM) is utilized to model ESD events which occur when a charged individual discharges through the pin of an IC while another IC is connected to a fixed potential, such as ground. A circuit model of a HBM simulator is shown in Figure 1.9 [5]. The event is modeled as a 100 pF capacitor which is discharged through a 1.5 kΩ resistor. The capacitor models the charge stored on the individual, and the resistor models the impedance of the individual and the spark resistance of the HBM event. In the figure, the 100 pF capacitor is charged up before the event using a high-voltage power supply with a series impedance of greater than 1 MΩ. The event occurs when the switch is closed and the capacitance discharges through the IC. A typical HBM event has a 10 ns risetime and a 100 ns duration. The HBM event

level is defined as the precharge voltage of the capacitor before the event. The Industry Council recommends a 1 kV protection level to provide adequate protection with margin when basic ESD control methods are utilized [3]. The peak current during a 1 kV HBM event is 0.66 A.

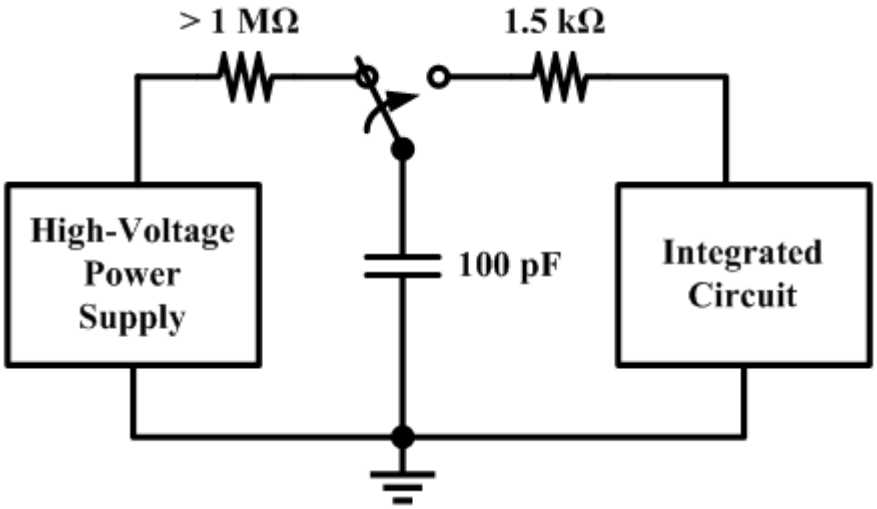


Fig 1.9: Human body model circuit model [5].

Due to triboelectric charging, an IC can be charged such that potential is developed [4]. The charged device model (CDM) is utilized to model the discharge of the IC through a single pin in contact with a fixed potential, such as ground. The CDM event level is defined as the precharge voltage of the integrated circuit before the event. A CDM event has a significantly shorter risetime and duration in comparison with a HBM event. Typically, a CDM event has a risetime on the order of a few hundred ps and an approximate duration of one ns [4]. The Industry Council recommends a 250 V protection level to provide adequate protection with margin when basic ESD control methods are utilized [4]. The peak current during a 250 V CDM event can be several amperes, and is dependent on the IC packaging [4].

1.6 Intersymbol Interference and Equalization

Due to the signal integrity degradation of the channel and other parasitic elements of the I/O link, intersymbol interference (ISI) is produced [6]. ISI occurs when signal distortion results in the energy of one symbol (one bit) entering another symbol (another bit).

If significant ISI is present, the BER of the link may be too high to support reliable data transmission. State-of-the-art high-speed I/O links which are ISI dominated implement

equalization which reduces ISI [6]. By reducing the ISI, the equalizer assists in the proper reconstruction of a bit. A common design is to implement feed-forward equalization (FFE) and decision-feedback equalization (DFE) to reduce ISI [6]. An example FFE architecture can be seen in Figure 1.10 and an example DFE architecture can be seen in Figure 1.11. The FFE is utilized to reduce the precursor ISI, which is ISI resulting from bits transmitted after the bit being resolved. The DFE is utilized to reduce the post-cursor ISI, which is ISI resulting from bits transmitted before the bit being resolved. In general, equalization requires active circuitry which dissipates power. It can be assumed that the more equalization effort exerted to reduce the ISI, the more complex the equalizer will be, resulting in higher power consumption.

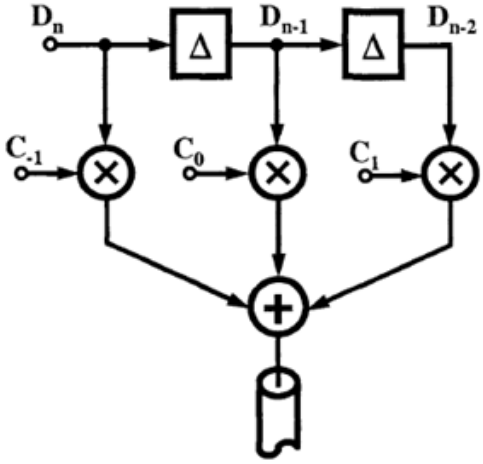


Figure 1.10: Transmit pre-emphasis finite impulse response (FIR) filter which implements feed-forward equalization [6].

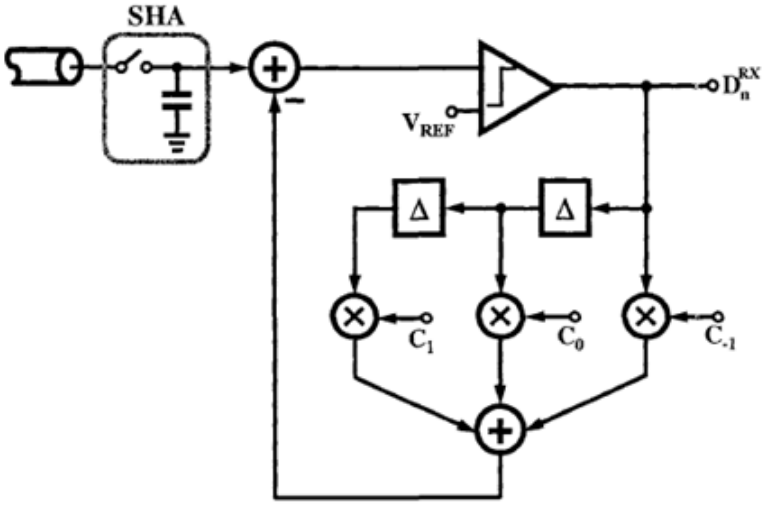


Figure 1.11: Sample and hold followed by a DFE equalizer [6].

Instead of attempting to reduce the ISI through the addition of mitigation signals at the transmitter or receiver, a frequency shaping filter can be utilized to flatten the channel response as shown in Figure 1.12 [6]. One such frequency shaping filter is transmit pre-emphasis, as utilized in Chapter 7, where the low-frequency content is attenuated through RC source degeneration of the transmit driver.

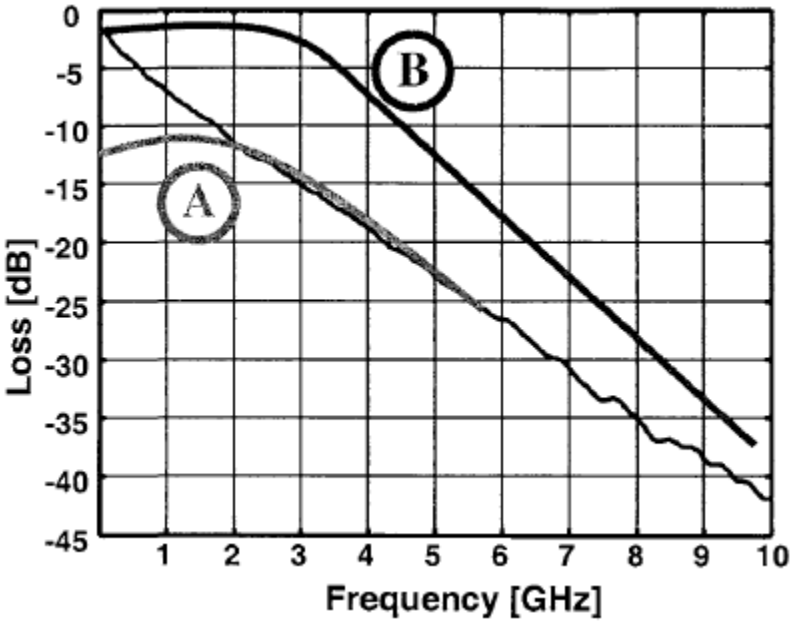


Figure 1.12: Two ways of equalization through a frequency shaping filter [6]. (a) Attenuation of low frequencies. (b) Boosting high frequencies.

1.7 Signal Integrity and Eye Diagrams

Signal integrity is a qualitative term which refers to the degradation of the signal due to ISI and noise. If a signal is said to have excellent signal integrity, then the signal has not been significantly impacted by ISI and noise. Therefore, the transmitted data stream can be reconstructed quite easily. On the other hand, if a signal is said to have poor signal integrity, then the signal has been significantly impacted by ISI and noise. Therefore, reconstruction of the data stream is quite difficult. When a signal is transitioning away from excellent signal integrity toward poor signal integrity, the transition can be described as degrading the signal integrity.

Often an eye diagram is utilized to determine the signal integrity of a high-speed I/O link. An eye diagram, example shown in Figure 1.13, is a measurement plot of a PRBS signal where the signal waveform has been overlapped upon itself at a specified interval, usually the period of the

symbol (bit period for NRZ signaling). The quality of the signal integrity is determined through a qualitative assessment of several factors including zero-crossing jitter, eye opening (height), and the shape of the inner-most (worst-case) eye.

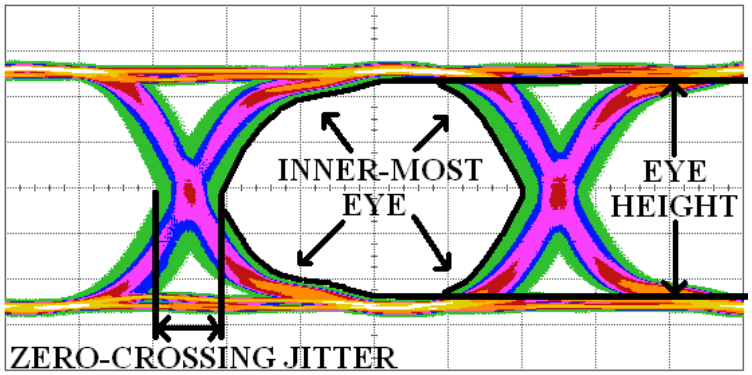


Figure 1.13: Eye diagram of a 4 Gb/s, 0.5 V_{ppd} PRBS sequence generated with a commercial transmitter (horizontal scale = 50 ps/div. and vertical scale = 100 mV/div.).

CHAPTER 2

SYSTEM-ASSISTED I/O DESIGN STUDIES

2.1 Introduction

The state-of-the-art design methodology for high-speed I/O links is to specify component-level design requirements to achieve high-quality component-level performance. The components are then unified and the high-quality component performance is used to guarantee link performance, specifically the bit error rate. While the methodology guarantees a reliable link, it does not inherently optimize the link for metrics such as the power, design complexity, or bit error rate performance. Recently, due to the increased demand for data bandwidth in backplane I/O, a system-assisted design methodology has been developed to optimize the system for a given set of metrics. By optimizing on the system level rather than the component level, the performance at the component level can be reduced from high-quality to sufficient when utilized in the co-designed I/O link system. The reduction in component-level performance requirements provides additional design margin which allows for optimization of the desired metrics. These novel system-level approaches may require the design of analog and mixed-signal circuits with unique functionality requirements. The unique functionality requirements encourage the utilization of novel circuit architectures in the context of high-speed I/O links.

2.2 Objectives

In this dissertation, novel analog and mixed-signal circuitry for system-assisted high-speed I/O links is presented. In the context of system-assisted design methodology, the circuits do not present ground-breaking fidelity. Instead, the circuit architectures were designed to provide unique characteristics which enable novel system architectures. The novel circuit architectures were developed by expanding traditional analog and mixed-signal circuit architectures to achieve system-level design goals while exploiting the advantages of the design methodology. Due to the

fundamental goals of the system-assisted and state-of-the-art commercial IC design, the power and area were also important design characteristics.

In this dissertation, three system-level design studies were investigated. In the first system-level study, the transmitter output and receiver input bandwidth was investigated. If codesign of the I/O link shows that the input bandwidth of the receiver limits the bit error rate performance of the link, then negative capacitance circuits can be implemented to provide bandwidth extension to mitigate the system-level limitations. In the second system-level study, the utilization of a variable-reference analog-to-digital converter (ADC) for receive sampling was investigated. Since the performance metric for high-speed I/O links is BER rather than sampling fidelity, the design requirements for an ADC in a high-speed I/O link may be modified from those of traditional ADCs. The system-level solution included an optimization of the reference levels to reduce the complexity of the ADC. Therefore, a digital-to-analog converter (DAC) was designed to generate the variable reference levels for the ADC. In the third system-level study, the utilization of coding in high-speed backplane links was investigated. The utilization of coding relaxes the BER requirement at the input of the receive slicer, allowing for a reduction of inner-transceiver power and design complexity.

2.3 Receiver Input Bandwidth Extension

The non-idealities of backplane I/O links limit the signal integrity at the input of the receiver, as noted in Chapter 1. The degraded signal integrity may limit the bit error rate performance of the link. The current design methodology for high-speed I/O links minimizes or mitigates the effect of each non-ideality with a component view of the problem: minimizing the sizing of the ESD protection, implementing equalization to negate the channel induced ISI, etc. In a system-level codesign of the link, attempting to minimize or mitigate all of the non-idealities may be unnecessary. For example, in [7], the effect of ESD parasitic capacitance on the BER performance is shown to be dependent on the characteristics of the link. In some links, the BER is insensitive to changes in the ESD parasitic capacitance, thereby relaxing the design constraint on the ESD protection circuitry. For links where the BER is sensitive to the parasitic capacitance, [7] investigates the usage of equalization to mitigate the impact of the parasitic capacitance on the BER.

In this system-assisted study, a lumped element model is developed to capture the effect of parasitics at the output of the transmitter and the input of the receiver. The lumped element model is reduced to a first- or second-order system to generate compact equations for the bandwidth at the output of the transmitter or the input of the receiver. The compact equations for the bandwidth can be analyzed to provide an understanding of how the parasitic elements interact to limit the bandwidth. The first- or second-order system can be used within the system-assisted codesign of the link to determine the effect of the transmitter output and receiver input bandwidth on the BER performance of the link.

Since the first- or second-order system does not sufficiently describe how the receiver input poles change as a function of the parasitic elements, a root locus analysis is performed on the receiver input, with the magnitude of the parasitic capacitance as the variable. The root locus analysis also provides information on the pole location limitations and maximum achievable bandwidth. Additionally, the analysis is utilized to formulate parasitic capacitance and inductance budgets.

If the receiver input bandwidth is shown to limit the bit error rate performance of the link, then several system-level and component-level solutions exist. The component-level solution is to restrict the parasitic elements to magnitudes that do not affect the performance of the link; i.e., they set a maximum level of parasitic capacitance for the ESD protection circuitry. On the system-level, equalization [7] or bandwidth extension [8], [9], [10], circuits can be utilized to improve the BER performance to an acceptable level. In this dissertation, the usage of negative capacitance circuits to extend the bandwidth is investigated. Additionally, a novel, g_m-boosted negative capacitance circuit is implemented and measured for the first time. The g_m-boosted negative capacitance, in a system-assisted approach, utilizes the inherent gain of the receive amplifier to enhance the characteristics of the traditional negative capacitance circuit.

2.4 Receive Sampling via Variable-Reference ADC

In the past five years, the usage of high-speed ADCs in backplane links to sample the receive signal has been presented [11], [12], [13]. The advantage of utilizing a high-speed ADC is that it allows for the use of digital receiver equalization. In prior work, the required ADC resolution to

effectively equalize the link has been in the range of 4.5 to 6 bits. Due to power constraints of high-precision ADCs, precisions higher than 6 bits have not been implemented.

In [14], a BER-optimal ADC-based receive architecture is presented. The optimal BER performance for an ADC does not necessarily occur for uniformly distributed quantization levels [14]. Therefore, a variable-reference ADC is required for the system-assisted design of a BER-optimal ADC-based receiver. In fact, [14] showed that the number of thresholds for a variable-level ADC can be reduced while maintaining a BER performance equal to or exceeding that of a uniform ADC with a higher number of thresholds. A reduction in the number of thresholds can result in a significant power savings in the ADC, especially if a flash ADC architecture is utilized [14].

In the second system-assisted study, a variable-reference ADC for receive sampling was implemented in 90 nm CMOS. The variable-reference ADC maintains the same core structure as a uniform ADC, but the reference levels provided to the threshold decision circuits are modified to obtain the desired BER-optimal thresholds. As such, the DAC, which provides the reference levels, must be modified to generate the optimum thresholds. In Chapter 5, a novel single-core multiple-output DAC architecture is presented which fulfills the requirements of the BER-optimal ADC receiver without negating the power advantage from the reduction in the number of threshold levels.

2.5 Backplane High-Speed I/O Utilizing Coding

Multi-Gb/s backplane serial links strive to achieve a bit error rate (BER) < 10^{-12} in the presence of significant channel loss and dispersion. Traditionally, such low BERs are achieved exclusively through the application of transmit side pre-distortion (i.e., pre-emphasis) and/or receive side equalization [11], [15]. The combination of channel loss and pre-emphasis equalization reduces the received signal swing such that circuit noise, clock jitter and crosstalk can limit the BER. In such a scenario, forward error correction (FEC) can be an effective technique for enhancing BER and/or reducing energy.

Though FEC is ubiquitous in wireless and long-haul wireline communication links [16], it has not found application in backplane links thus far. In [17] and [18], the application of FEC to 10 Gb/s signaling over channels with 12 to 15 dB loss at the Nyquist frequency via analysis and

simulations was studied, and it was concluded that binary Bose and Ray-Chaudhuri (BCH) codes along with receive equalization provides a coding gain of 3 to 6 dB at a BER of 10^{-15}. The reduced signal-to-noise ratio (SNR) requirements represented by the coding gain may be employed to: (*a*) reduce transmit swing, (*b*) reduce equalizer complexity, (*c*) enhance jitter-tolerance, and (*d*) improve comparator offset tolerance.

The third system-assisted study presents the results of a 90nm CMOS implementation of a 4 Gb/s line rate backplane transceiver employing FEC. The measurement results provide a quantitative assessment of how coding gain affects the trade-off between transmit swing, jitter tolerance, BER, and power. The results show that much of the promise of FEC identified in [17] and [18] can indeed be fulfilled in practice.

CHAPTER 3

TRANSMITTER AND RECEIVER BANDWIDTH ESTIMATION

3.1 Introduction

This chapter investigates the effect of ESD protection and other parasitics on the signal integrity of multi-Gb/s wireline transmitters and receivers. A lumped element model of the transmitter output or receiver input is used to generate a compact approximation for the bandwidth. The resulting expressions elucidate the sensitivity of the frequency response to the ESD protection devices and other parasitics. For the receiver input, Chapter 5 will present test chip measurement data to quantify the effect of ESD protection on the signal integrity at the receiver input as a function of the data rate, and also to validate the lumped model analysis.

3.2 Transmitter Schematic and Lumped Element Model

In this section, a lumped model is developed for the signal path in a system consisting of a wireline transmit driver and its load, i.e., the termination, transmission line, package, and all on-chip elements at the driver output. The parasitic elements in the package and on-chip, including those associated with the ESD protection circuit, will be shown to deteriorate the termination match of the transmitter to the transmission line at high frequencies, placing an upper bound on the transmitter output bandwidth.

First, we model the impedance at the input of the transmission line. For this analysis, we do not explicitly follow the traditional frequency-domain analysis of the transmission line impedance for two reasons. First, a periodic steady-state does not exist for I/O links transmitting random data. Second, any signal arriving at the transmitter due to reflection is generally considered ISI due to transit time considerations and, therefore, will be treated as noise. As such, we model the transmission line with its characteristic impedance, Z_O. The signal propagated down the transmission line is therefore equal to V_{BEGIN}, the voltage across the resistor Z_O.

The dominant packaging-related parasitic element is often a series inductance, L, located between the transmission line and the input pads. On-chip, a shunt termination resistance R_{TERM} is placed at the output of the driver, with the pad also contributing parasitic capacitance $C_{BONDPAD}$. The other on-chip parasitic elements vary with the ESD protection scheme; the dual-diode based protection scheme of Figure 3.1 is commonly used as primary protection for high-speed I/O links, due to its relatively low parasitic capacitance [19]. This circuit contributes a parasitic capacitance C_{ESD} that is in parallel with R_{TERM}, $C_{BONDPAD}$, and C_{TX}, the capacitance of the driver output transistors. Driving our output node, we model the transconductance transistors of our driver as a current source, I_{DRV}. The complete lumped element model is shown in Figure 3.2.

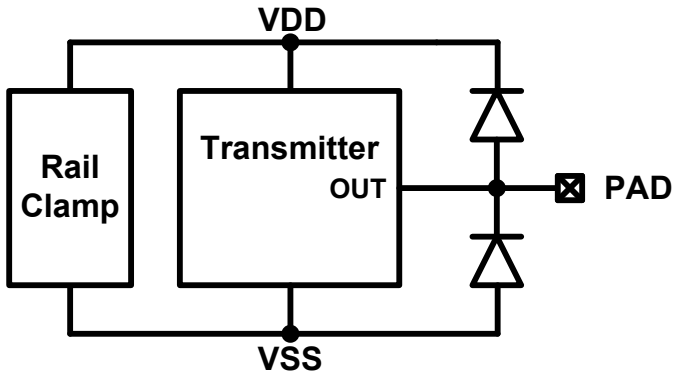


Figure 3.1: Transmitter dual-diode ESD protection scheme.

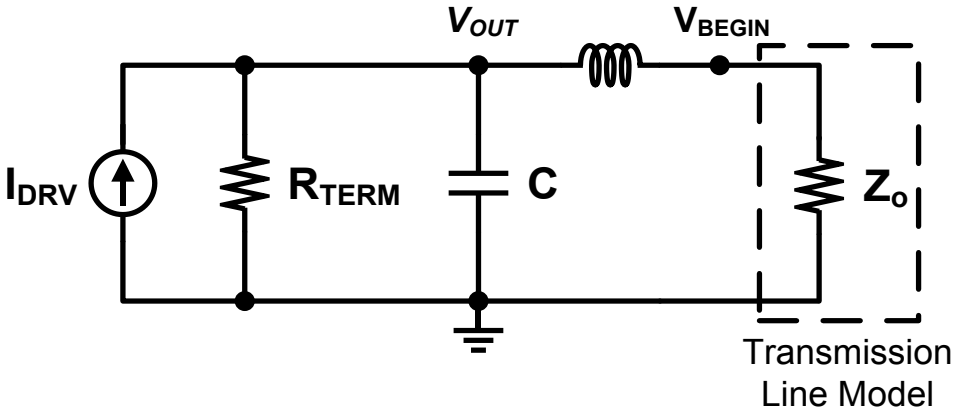


Figure 3.2: Transmitter input lumped element model for the dual-diode ESD protection scheme of Figure 3.1. For this circuit, $V_{PAD} = V_{OUT}$ and $C = C_{BONDPAD} + C_{ESD} + C_{TX}$.

3.3 Receiver Schematics and Lumped Element Models

In this section, lumped models are developed for the signal path in a system consisting of a wireline receiver and its feed, i.e., the transmission line, package, and all on-chip elements up to the input transistor. The parasitic elements in the package and on-chip, including those associated with the ESD protection circuit, will be shown to deteriorate the termination match of the receiver to the transmission line at high frequencies, placing an upper bound on the receiver bandwidth.

First, a Thevenin equivalent model of the transmission line is constructed, based on the time-domain transmission line equations

$$V_{END}(t) = V^{+}(t) \cdot (1 + \Gamma_L) \tag{3.1}$$

and

$$\Gamma_L \equiv \frac{Z_L - Z_o}{Z_L + Z_o}, \tag{3.2}$$

where, V_{END} is the voltage at the end of the transmission line where it abuts the package, V^{+} is the signal incident on this end point, Γ_L is the reflection coefficient at the end of the transmission line, Z_L is the impedance seen at the end of the transmission line, and Z_O is the characteristic impedance of the transmission line. Inserting (3.2) into (3.1) produces

$$V_{END}(t) = 2V^{+}(t) \frac{Z_L}{Z_o + Z_L}. \tag{3.3}$$

The dominant packaging-related parasitic element is often a series inductance, L, located between the transmission line and the input pads; this is the case for the high-speed QFN package used in Chapter 5. On-chip, a termination resistance R_{TERM} is placed at the input pad, with the pad also contributing parasitic capacitance $C_{BONDPAD}$. The other on-chip parasitic elements vary with the ESD protection scheme; the dual-diode based primary protection scheme of Figure 3.3 is commonly used on high-speed I/O links, due to its relatively low parasitic capacitance [19], and is employed in this dissertation. This circuit contributes a parasitic capacitance C_{ESD} that is in parallel with R_{TERM}, $C_{BONDPAD}$, and C_{RX}, the capacitance of the receiver input. The complete lumped element model is shown in Figure 3.4.

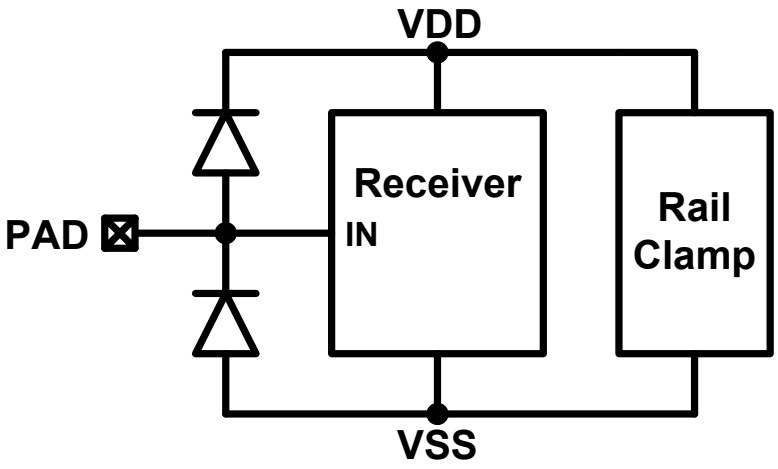


Figure 3.3: Receiver dual-diode ESD protection scheme.

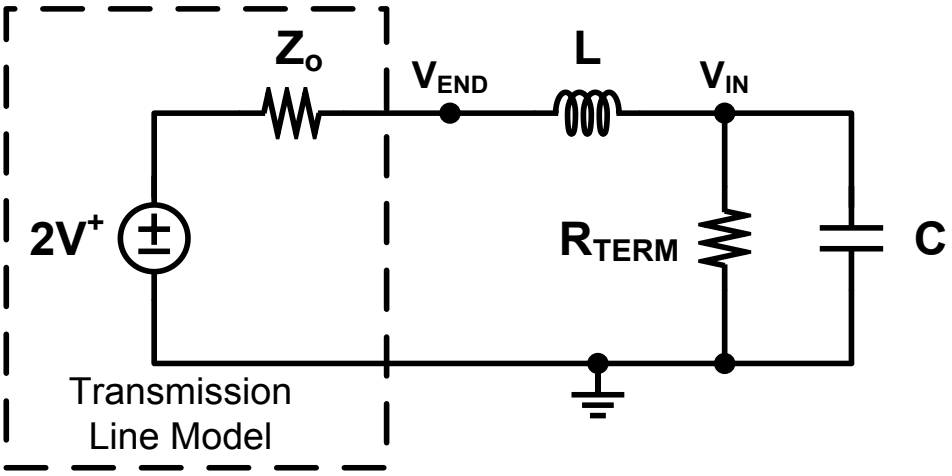


Figure 3.4: Receiver input lumped element model for the dual-diode ESD protection scheme of Figure 3.3. For this circuit, $V_{PAD} = V_{IN}$ and $C = C_{BONDPAD} + C_{ESD} + C_{RX}$.

If a CDM protection level is specified, additional ESD circuitry may be implemented on-chip. This is shown in Figure 3.5, where a series CDM resistor R_{CDM} and secondary ESD protection circuit have been added between the input pad and the receiver input. Here, the secondary ESD protection circuit is depicted as another dual-diode circuit, presumably with much smaller devices than in the primary protection circuit. A new lumped element model is generated for this CDM-protected receiver, as shown in Figure 3.6. In this model, the parasitic capacitances are partitioned on either side of R_{CDM} into C_{PAD} and C_{IN}, where $C_{PAD} = C_{BONDPAD} + C_{ESD}$ and $C_{IN} = C_{ESD2} + C_{RX}$. The parasitic capacitance of the secondary protection device(s) is denoted by C_{ESD2}.

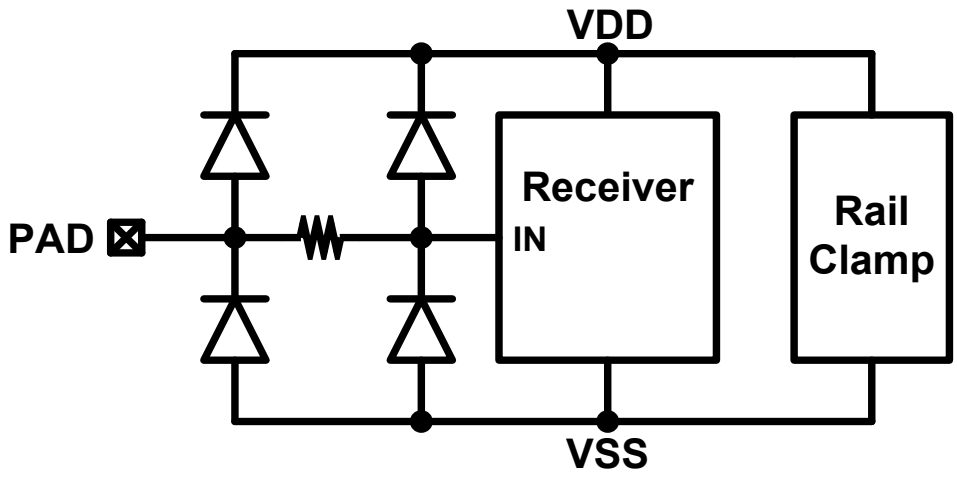


Figure 3.5: Dual-diode primary ESD protection scheme which has been augmented with a series resistor and secondary dual-diode devices to improve CDM resiliency.

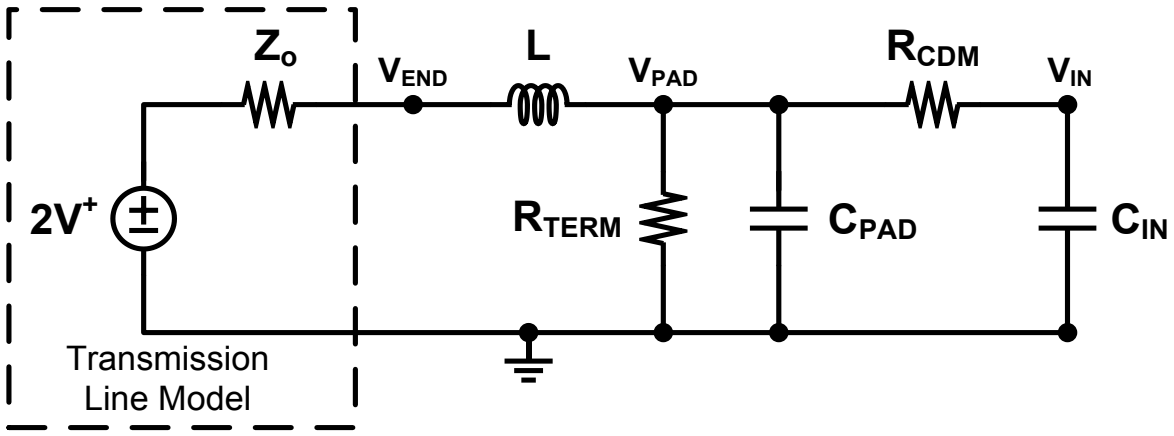


Figure 3.6: Receiver input lumped element model for the ESD protection scheme of Figure 3.5. $C_{PAD} = C_{BONDPAD} + C_{ESD}$ and $C_{IN} = C_{ESD2} + C_{RX}$.

3.4 Bandwidth Estimation Method

Using the transmitter model of Figure 3.2 or receiver models of Figures 3.4 and 3.6, one can derive a transfer function to relate $V_{BEGIN}(f)$ to $I_{IN}(f)$ for the transmitter or $V_{IN}(f)$ *to* $V^{+}(f)$ for the receiver. The transfer function of a linear system with lumped elements, $H(s)$ [20], is

$$H(s) = \frac{a_0 + a_1 s + a_2 s^2 + \ldots + a_m s^m}{1 + b_1 s + b_2 s^2 + \ldots + b_n s^n}. \tag{3.4}$$

Deriving the transfer function of a multiple element RLC circuit can be tedious, and thus it is preferable to use convenient techniques to obtain the coefficients of the first one (b_1) or two (b_1

and b_2) terms in the denominator [20]. If the system has dominant poles, then the overall bandwidth can be well approximated using these few coefficients.

The procedure for calculating the bandwidth is an application of the circuit analysis in [20] and expanded to calculate the bandwidth of the second-order system. The bandwidth of the first-order system, $\hat{f}_{-3dB}$, is equal to

$$\hat{f}_{-3dB} = \frac{1}{2\pi b_1}, \tag{3.5}$$

which requires the calculation of b_1. The bandwidth of the second-order system, $\bar{f}_{-3dB}$, is equal to

$$\bar{f}_{-3dB} = f_n \sqrt{1 - \frac{1}{2Q^2} + \sqrt{\frac{1}{4Q^4} - \frac{1}{Q^2} + 2}}, \tag{3.6}$$

which requires the calculation of f_n, the natural frequency of the second-order system, and Q, the quality factor of the second-order system. The natural frequency is equal to

$$f_n = \frac{1}{2\pi\sqrt{b_2}}, \tag{3.7}$$

which requires the calculation of b_2. The second-order quality factor can be calculated from b_1 and b_2 as

$$Q = \frac{\sqrt{b_2}}{b_1} \tag{3.8}$$

and rewritten in terms of $\hat{f}_{-3dB}$ and $\bar{f}_{-3dB}$:

$$Q = \frac{\hat{f}_{-3dB}}{f_n} \tag{3.9}$$

The calculation of b_1 and b_2 for the receiver lumped element models is shown in bandwidth calculations in Section 3.5.

The analysis of [20] notes the impact of zeros on the bandwidth estimation and derives a modified bandwidth estimation in the presence of significant zero(s). Inspecting the circuits in Figure 3.2, 3.4, and 3.6, it is obvious that all transfer constants, $H^{ijk\cdots}$, defined in [20] are equal to zero since infinitely valuing any reactive element results in isolation between the transmission line and the transmitter or receiver [20]. Therefore,

$$a_i = 0, \text{for } i \geq 1. \tag{3.10}$$

Since the numerator of the transfer function is a constant, no finite zeros exist. Therefore, the bandwidth equations (3.5) and (3.6) do not need to be modified to improve their accuracy in the presence of significant zeros since no finite zeros exist.

3.5 Receiver Bandwidth Estimation

3.5.1 First-Order Bandwidth Derivation

To calculate the first-order system bandwidth for Figures 3.4 and 3.6, b_1 must be derived. From [20], b_1 is equal to the sum of all zero-value time constants. Therefore, for Figure 3.4

$$b_1 = \frac{L}{Z_o+R_{TERM}} + (Z_o||R_{TERM})C \tag{3.11}$$

and

$$b_1 = \frac{L}{Z_o+R_{TERM}} + (Z_o||R_{TERM})C_{PAD} + [R_{CDM} + (Z_o||R_{TERM})]C_{IN} \tag{3.12}$$

for Figure 3.6. For Figure 3.4, the first-order bandwidth can be calculated by inserting (3.11) into (3.5):

$$f_{-3dB} = \frac{Z_o+R_{TERM}}{2\pi(L+R_{TERM}Z_oC)}. \tag{3.13}$$

If the termination is matched to the characteristic impedance of the channel, then (3.13) simplifies to

$$f_{-3dB} = \frac{1}{\pi(\frac{L}{Z_o}+Z_oC)}. \tag{3.14}$$

Similarly for Figure 3.6, the first-order bandwidth can be calculated by inserting (3.12) into (3.5):

$$f_{-3dB} = \frac{Z_o+R_{TERM}}{2\pi[L+R_{TERM}Z_o(C_{PAD}+C_{IN})+(R_{TERM}+Z_o)R_{CDM}C_{IN}]}. \tag{3.15}$$

If the termination is matched to the characteristic impedance of the channel, then (3.15) simplifies to

$$f_{-3dB} = \frac{1}{\pi\left[\frac{L}{Z_o}+Z_o(C_{PAD}+C_{IN})+2R_{CDM}C_{IN}\right]}. \tag{3.16}$$

3.5.2 Second-Order Bandwidth Derivation

To calculate the second-order system bandwidth for Figures 3.4 and 3.6, b_1 and b_2 must be derived. The derivation of the first-order bandwidth in Section 3.5.1 includes the derivation of b_1. From [20], b_2 can be calculated from a set of time constants. Therefore, for Figure 3.4

$$b_2 = \frac{L}{Z_o+R_{TERM}}R_{TERM}C \tag{3.17}$$

and

$$b_2 = \frac{L}{Z_o+R_{TERM}} R_{TERM} C_{PAD} + \frac{L}{Z_o+R_{TERM}} (R_{TERM} + R_{CDM}) C_{IN} + (Z_o || R_{TERM}) R_{CDM} C_{PAD} C_{IN} \quad (3.18)$$

for Figure 3.6. For Figure 3.4, the second-order natural frequency can be calculated by inserting (3.17) into (3.7):

$$f_n = \frac{1}{2\pi} \sqrt{\frac{R_{TERM}+Z_o}{R_{TERM} LC}}. \quad (3.19)$$

If the termination is matched to the characteristic impedance of the channel, then (3.19) simplifies to

$$f_n = \frac{1}{\pi} \sqrt{\frac{1}{2LC}}. \quad (3.20)$$

Similarly for Figure 3.6, the second-order natural frequency can be calculated by inserting (3.18) into (3.7):

$$f_n = \frac{1}{2\pi} \sqrt{\frac{R_{TERM}+Z_o}{R_{TERM} L(C_{PAD}+C_{IN}) + (L + R_{TERM} Z_o C_{PAD}) R_{CDM} C_{IN}}}. \quad (3.21)$$

If the termination is matched to the characteristic impedance of the channel, then (3.21) simplifies to

$$f_n = \frac{1}{\sqrt{2}\pi} \sqrt{\frac{1}{L(C_{PAD}+C_{IN}) + (\frac{L}{Z_o} + Z_o C_{PAD}) R_{CDM} C_{IN}}}. \quad (3.22)$$

The second-order system quality factor can then be calculated from first-order bandwidth and second-order natural frequencies:

$$Q = \frac{\hat{f}_{-3dB}}{f_n}. \quad (3.23)$$

Then, the natural frequency and quality factor can be utilized to calculate the second-order bandwidth

$$\bar{f}_{-3dB} = f_n \sqrt{1 - \frac{1}{2Q^2} + \sqrt{\frac{1}{4Q^4} - \frac{1}{Q^2} + 2}}. \quad (3.24)$$

3.5.3 Bandwidth Estimation Method

The first- and second-order bandwidth derivations from Sections 3.5.1 and 3.5.2 can be utilized to generate a bandwidth estimation method. The first-order bandwidth estimation does not determine the location of the poles on the real-imaginary plane [20]. If the dominant poles of the system are complex, then the bandwidth can be severely underestimated by the first-order bandwidth estimation [20]. In the second-order system, the quality factor, Q, provides information about the location of the two poles. If $Q < 0.5$, then the second-order system consists of two distinct real poles. If $Q = 0.5$, then the second-order system consists of two identical real poles (a double pole). If $Q > 0.5$, then the second-order system consists of a pair of complex conjugates. While the second-order system is a better approximation for the system since fewer higher-order terms are ignored, the first-order system is desirable for small Q values. When Q is < 0.5 and small, changes in the element values may quickly change Q. When Q quickly changes, it is difficult to understand how the parasitic elements interact to limit the bandwidth of the system. On the other hand, for small Q, the system is well approximated by the first-order system since the second-order system has a dominant pole. Therefore, the first-order bandwidth approximation can be used without a significant loss in accuracy while providing an equation which allows for analysis of how the parasitic elements interact to limit the bandwidth of the system. Therefore, the first-order bandwidth is utilized to approximate the bandwidth when $Q < 0.5$ and the second-order bandwidth is utilized to approximate the bandwidth when Q is ≥ 0.5.

The procedure for bandwidth estimation is detailed below, for the case that $R_{TERM} = Z_O$, i.e., the termination has been matched to the characteristic impedance of the transmission line. If the termination is not matched to the transmission line, then the corresponding equations from Sections 3.5.1 and 3.5.2 should be used instead.

Step 1: Calculate the estimator $\hat{f}_{-3dB}$ using

$$\hat{f}_{-3dB} = \frac{1}{\pi(\frac{L}{Z_o}+Z_oC)} \tag{3.25}$$

for the receiver model of Figure 3.4, or

$$\hat{f}_{-3dB} = \frac{1}{\pi\left[\frac{L}{Z_o}+Z_o(C_{PAD}+C_{IN})+2R_{CDM}C_{IN}\right]} \tag{3.26}$$

for that shown in Figure 3.6.

Step 2: Calculate f_n as

$$f_n = \frac{1}{\pi\sqrt{2LC}} \tag{3.27}$$

for the model of Figure 3.4, or

$$f_n = \frac{1}{\pi\sqrt{2[L(C_{PAD}+C_{IN})+\left(\frac{L}{Z_o}+Z_oC_{PAD}\right)R_{CDM}C_{IN}]}} \tag{3.28}$$

for the model of Figure 3.6.

Step 3: Calculate Q as

$$Q = \frac{\hat{f}_{-3dB}}{f_n}\ . \tag{3.29}$$

Step 4: Calculate the estimator $\bar{f}_{-3dB}$ as

$$\bar{f}_{-3dB} = f_n\sqrt{1-\frac{1}{2Q^2}+\sqrt{\frac{1}{4Q^4}-\frac{1}{Q^2}+2}}\ . \tag{3.30}$$

Step5: If $Q \geq 0.5$, then the estimated −3 dB bandwidth f_{-3dB} is

$$f_{-3dB} = \bar{f}_{-3dB}\ , \tag{3.31}$$

else the estimated −3 dB bandwidth f_{-3dB} is

$$f_{-3dB} = \hat{f}_{-3dB}\ . \tag{3.32}$$

In step 1, the estimator $\hat{f}_{-3dB}$, the approximate bandwidth of the first-order system, is calculated. In step 2, f_n, the natural frequency of the second-order system, is calculated. In step 3, the quality factor, Q, of the second-order system is calculated, and in step 4, the −3 dB

bandwidth of the second-order system is calculated. In step 5, the Q value is used to determine if the second-order system contains complex poles. If $Q \geq 0.5$, then there are complex poles (or double pole in the event of $Q = 0.5$), and the second-order system bandwidth is the bandwidth estimate. Otherwise, the first-order system bandwidth is the bandwidth estimate.

3.5.4 Bandwidth Simulations

To evaluate the receiver bandwidth estimation formulas presented above, the circuits of Figures 3.4 and 3.6 were simulated using Spectre and the results are compared to (3.31) and (3.32). In all simulations, the characteristic impedance of the transmission line and the termination is 50 Ω.

The simulated bandwidth of the circuit in Figure 3.4 is shown in Figure 3.7, as a function of L and C. The results are equivalent to the results for the circuit in Figure 3.6 with $R_{CDM} = 0$ Ω and $C = C_{PAD} + C_{IN}$. For the investigated range of L and C values, $Q \geq 0.5$, and, according to (3.31), $\bar{f}_{-3dB}$ should be used to estimate f_{-3dB}. The results shown in Figure 3.7 confirm this assertion and, in fact, $\bar{f}_{-3dB}$ is exactly equal to the simulated bandwidth, because the circuit of Figure 3.4 is precisely a second-order system.

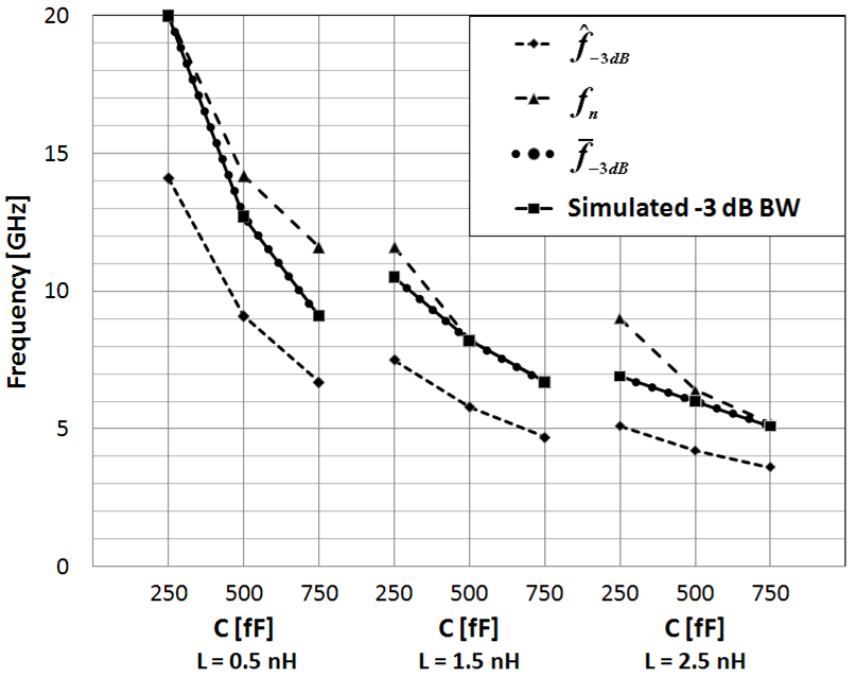


Figure 3.7. Simulated bandwidth of the circuit in Figure 3.4, as a function of C for L = 0.5 nH, 1.5 nH, and 2.5 nH. $Z_o = R_{TERM} = 50$ Ω. Equations (3.22), (3.24), and (3.27) are also plotted.

The simulated bandwidth of the circuit in Figure 3.6 is shown in Figure 3.8. Here, C_{PAD} and C_{IN} are fixed, but L and R_{CDM} are varied. As R_{CDM} increases, the bandwidth is deteriorated. In the considered range of L and C values, the quality factor is greater than 0.5, and $\bar{f}_{-3dB}$ should be used to estimate f_{-3dB}. Due to the presence of a third pole in the complete transfer function, $\bar{f}_{-3dB}$ provides a conservative estimate of the bandwidth, as shown in Figure 3.8.

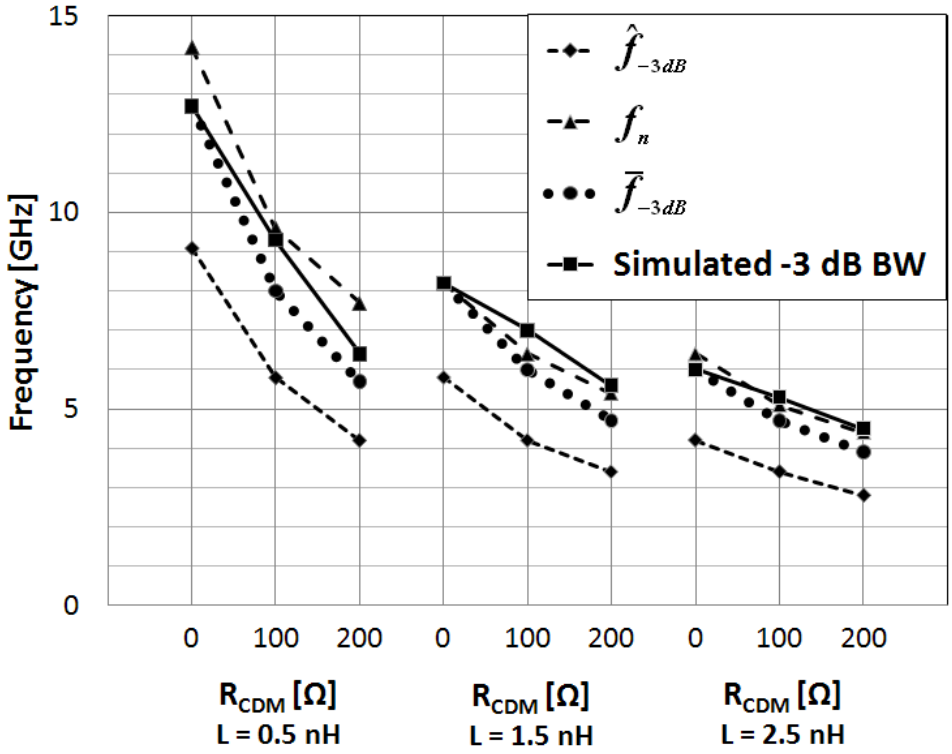


Figure 3.8. Simulated bandwidth of the circuit in Figure 3.6 as a function of R_{CDM} for L = 0.5 nH, 1.5 nH, 2.5 nH. C_{PAD} = 400 fF, C_{IN} = 100 fF, and $Z_o = R_{TERM}$ = 50 Ω. Equations (3.23), (3.25), and (3.27) are also plotted.

The results of Figures 3.7 and 3.8 show that, for the investigated range of parasitic elements, a two pole model of the receiver input is more accurate than a single pole model. For wirebond packages with inductances on the order of a nanohenry, such as QFNs, the improved accuracy will allow for the formulation of a more precise capacitance budget for the ESD protection circuit [21]. In contrast, for very low inductance packages, such as flip chip BGAs, a single pole model of the receiver may provide an accurate estimate of the frequency response.

While the bandwidth estimation method can produce an excellent bandwidth estimate, as shown in Figures 3.7 and 3.8, the bandwidth estimation method may not be sufficient for the optimization of ESD protection. Whenever a high-order system is approximated with a lower

order system, the approximation will result in a loss in precision, potentially changing the sensitivity of the bandwidth to an element. For some systems, the loss in precision may be small as the dominant system dynamics can be represented by the lower order system, while for other systems, the loss in precision may be significant as the dominant system dynamics cannot be represented by the lower order system. In Figures 3.7 and 3.8, the impact of reducing the lower order system approximation from second-order to first-order is shown to be significant. In Chapter 4, it will be shown that both poles of the circuit in Figure 3.4 are significant for a wide range of element values, resulting in an imprecise first-order system approximation. To determine the impact of approximating a third-order system, such as the circuit in Figure 3.6 or the receiver input with negative capacitance presented in Chapter 5, as a second-order system in order to estimate the bandwidth, the dynamics of the third-order system must be ascertained. In Chapter 4 and 5, a root locus analysis method is presented to determine how an element within a circuit impacts the location of the poles. If the root locus analysis demonstrates, for the magnitude range of interest for each element, that two poles dominate the response, then the bandwidth estimation method should provide excellent accuracy. In this scenario, the bandwidth estimation method can be utilized in order to optimize the design of the ESD protection circuitry. Utilizing the bandwidth estimation method is desirable as the relationship between the elements and the −3 dB bandwidth can be determined algebraically. If the root locus analysis demonstrates that three or more poles dominate the response, then the bandwidth estimation method will underestimate the −3 dB bandwidth and the sensitivity of the bandwidth to individual elements may be distorted. In this scenario, the root locus analysis method, instead of the bandwidth estimation method, should be utilized to optimize the design of the ESD protection circuitry.

3.5.5 Analysis of the Estimated Bandwidth

The simulation results indicate that both the parasitic series inductance and parasitic capacitance limit the input bandwidth of the receiver. Therefore, the amount of parasitic capacitance that a high-speed receiver can tolerate for a specific signal integrity requirement is strongly dependent on the package inductance. An engineering rule of thumb can be formulated for receiver inputs with $Q < 0.5$, based on (3.25): if $Z_O = 50\ \Omega$, the signal integrity effect of 0.25 nH of package inductance (L) is equivalent to 100 fF of parasitic capacitance from the ESD

protection circuit (C_{ESD}). A separate rule of thumb can be formulated for receiver inputs with $Q \geq 0.5$; this rule is based on f_n of (3.24) which, as shown in Figure 3.7 and 3.8, provides an upper-bound on f_{-3dB}. For receiver inputs with $Q \geq 0.5$ and for a fixed bandwidth specification, a 20% decrease in the parasitic inductance allows for a 25% increase in the total parasitic capacitance at the input of the receiver.

The derivation in Section 3.5.3 shows that the addition of a series resistor R_{CDM} results in an additional term in the denominator of the bandwidth estimator, reducing the bandwidth. By reorganizing (3.26) to

$$\hat{f}_{-3dB} = \frac{1}{\pi[\frac{L}{Z_o}+Z_oC_{PAD}+(1+2\frac{R_{CDM}}{Z_o})Z_oC_{IN}]} \tag{3.33}$$

and (3.28) to

$$f_n = \frac{1}{\pi\sqrt{2[LC_{PAD}+(1+\frac{R_{CDM}}{Z_o}+\frac{Z_oC_{PAD}}{L}R_{CDM})LC_{IN}]}} \tag{3.34}$$

for Figure 3.6 and comparing to (3.25) and (3.27) of Figure 3.4, the addition of the series resistor can be seen as applying a magnification factor to the effect of C_{IN}. For a receiver with $Q < 0.5$, R_{CDM} magnifies the bandwidth limiting effect of C_{IN} parasitic capacitance by a factor of $1 + 2\frac{R_{CDM}}{Z_o}$. For a receiver with $Q \geq 0.5$, R_{CDM} magnifies the effect of C_{IN} by a factor of $1 + \frac{R_{CDM}}{Z_o} + \frac{Z_oC_{PAD}}{L}R_{CDM}$. Since R_{CDM} is often greater than Z_o, the magnification factor can be significant. For example, if R_{CDM} is 100 Ω and Z_o is 50 Ω, then C_{RX} and C_{ESD2} are magnified by at least a factor of 3.

3.6 Optimization of Receiver Secondary Protection Circuits

A common CDM design rule is to set the product of the series CDM resistor and width of the secondary protection device(s) to a constant. Reasonably assuming that C_{ESD2} is a linear function of the width of the secondary protection device(s), this design rule requires that τ, defined below, also be a constant.

$$\tau = R_{CDM}C_{ESD2}\,. \tag{3.35}$$

Utilizing (3.35) as a design constraint to the bandwidth estimation method, a plot of the bandwidth as a function of R_{CDM} can be generated as in Figure 3.9. From the plot, an optimal CDM resistor value exists which optimizes the bandwidth. Utilizing (3.26) and (3.28)

constrained by (3.35), the CDM resistor value which maximizes the bandwidth, $R_{CDM,OPT}$, while achieving the targeted CDM resiliency level can then be calculated from

$$R_{CDM,OPT} = \sqrt{\frac{Z_O \tau}{2C_{RX}}} \tag{3.36}$$

for $Q < 0.5$ and

$$R_{CDM,OPT} = \sqrt{\frac{\tau}{(\frac{1}{Z_O} + \frac{Z_O C_{PAD}}{L})C_{RX}}} \tag{3.37}$$

for $Q \geq 0.5$. For example, if $\tau = 100\ \Omega \cdot 25$ fF, $C_{PAD} = 400$ fF, $L = 1.5$ nH, $C_{RX} = 75$ fF, and $Z_O = 50\ \Omega$, then $R_{CDM,OPT}$ is 31.6 Ω.

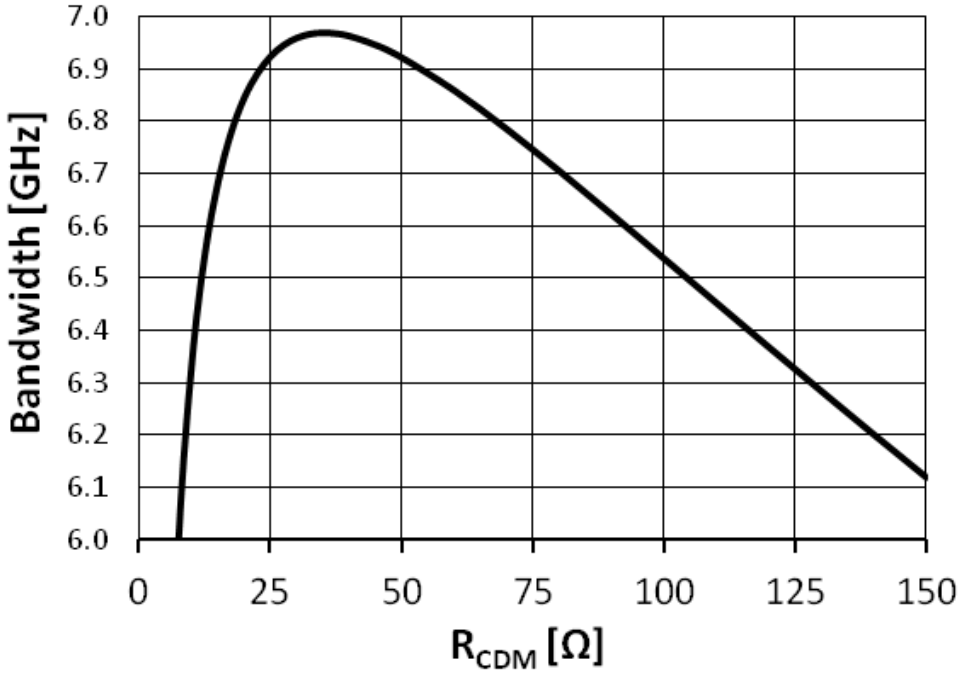


Figure 3.9: Bandwidth as a function of R_{CDM}, with a design constraint $\tau = R_{CDM}C_{ESD2} = 2500\ \Omega$·fF. $L = 1.5$ nH, $C_{PAD} = 400$ fF, and $C_{RX} = 75$ fF.

3.7 Transmitter Output Bandwidth Estimation

In Section 3.2, a lumped element model for a transmitter, shown in Figure 3.2, was generated. Similar to the receiver input poles, the transmitter output poles can degrading the signal integrity. By observation, the sourceless lumped element model of Figure 3.2 is identical to that of the receiver input, Figure 3.4, with relabeled nodes. Since the poles of a circuit are independent of the choice of input or output [20], the poles of Figure 3.2 are identical to the poles of Figure 3.4. Additionally, by inspection, the transfer function for the network of

Figure 3.2 does not contain any finite zeroes [20], so the entire analysis of the receiver of Figure 3.4 can be applied directly to the lumped element model of the transmitter in Figure 3.2.

3.8 Generalized Bandwidth Estimation Method

The bandwidth estimation method presented in Section 3.5 can be utilized as a general bandwidth estimation method. In the lumped element models investigated, Figures 3.2, 3.4, and 3.6, second- or third-order models of the transmitter output and receiver input ports are utilized. If a higher-order lumped element model of the transmitter and receiver is generated (or for any other input or output network), then the method described in Section 3.4 can be utilized to generate a first- and second-order bandwidth estimate by re-deriving b_1 and b_2 utilizing the methods in [20].

3.9 Estimating Bandwidth Extension of Negative Capacitance Circuits

In Chapter 5, circuits which present an impedance to negate parasitic capacitance will be presented as a method to enhance bandwidth. When evaluating whether it is worthwhile to include a negative capacitance circuit in the receiver front-end, one estimates the maximum possible bandwidth that can be obtained by ideally cancelling some or all of the pad capacitance. Utilizing the appropriate receiver model – Figure 3.4 or Figure 3.6 – one can evaluate the corresponding equations in Section 3.5.3 to find the maximum values of $\hat{f}_{-3dB}$ and $\bar{f}_{-3dB}$ as a function of C [C_{PAD}]; these, in turn, are used to obtain a good estimate of the maximum possible bandwidth. Due to the presence of packaging inductance, the maximum possible bandwidth does not necessarily occur for C or C_{PAD} equal to zero. For some applications, other parasitics will limit the achievable bandwidth extension. For the packages with significant series inductance; such as the test chip presented in Chapter 5, the bondwire inductance will limit the effectiveness of the negative capacitance circuits and place an upper-bound on the achievable bandwidth extension. If the optimal negative capacitance value is not achievable due to performance limitations on the negative capacitance circuits, the bandwidth estimation method can also be

used to estimate the achievable bandwidth extension as a function of the magnitude of achievable negative capacitances.

When estimating the bandwidth extension utilizing the method presented above, the capacitance mitigation is assumed to be ideal. In Chapter 5, the performance of negative capacitance circuits is shown to be limited by the negative resistance component of the impedance. The negative resistance of the circuit drastically impacts the receiver input pole locations and invalidates the ideal capacitance mitigation assumption. Therefore, estimating the bandwidth extension is a useful tool when determining the initial viability of a negative capacitance circuit, while the root locus method presented in Chapter 5 should be utilized to determine how much impact the addition of a negative capacitance circuit can have on the receiver input pole locations and bandwidth.

CHAPTER 4

RECEIVER INPUT ROOT LOCUS ANALYSIS

4.1 Introduction

In Chapter 3, a bandwidth estimation method was generated to analyze the bandwidth limitation at the receiver input. In this chapter, a scalable root locus plot of the receiver input poles will be generated in order to analyze how the poles change as a function of the parasitic capacitance. Additionally, a scalable plot of the corresponding −3 dB bandwidth will be generated with the significant points on the root locus identified. The results will lead to a suggested optimal design point for the magnitude of the parasitic capacitance. If the parasitic capacitance at the input of the receiver is under the optimal design point, then the addition of parasitic capacitance to the input is trivial. If the parasitic capacitance at the input of the receiver is over the optimal design point, the utilization of negative capacitance circuits for bandwidth extension will be investigated in Chapter 5. At the end of this chapter, the scalable −3 dB bandwidth plot will be used to generate parasitic capacitance and inductance budgets for the input of high-speed receivers.

4.2 Receiver Lumped Element Model

In Chapter 3, lumped element models were generated for the signal path consisting of a wireline receiver and its feed, i.e., the transmission line, package, and all on-chip elements up to the input transistor. For a receiver utilizing a primary-only ESD protection circuit, the corresponding lumped element model is shown in Figure 4.1, where V^+ is the signal incident at the end of the trace, V_{END} is voltage at the end of the transmission line where it abuts the package, and V_{IN} is the voltage at the input of the receiver. The trace is defined to be the transmission line with characteristic impedance Z_O, and is terminated with an on-chip termination resistance R_{TERM}. While an acceptable lumped element model for a package will vary, in this dissertation, the effect of packaging-related series inductance, L, located between the

trace and the receiver will be investigated. On-chip, there are several sources of parasitic capacitance at the input of the receiver which combine to generate a shunt capacitance *C*: the bondpad, $C_{BONDPAD}$, the ESD protection devices, C_{ESD}, and receiver input, C_{RX}.

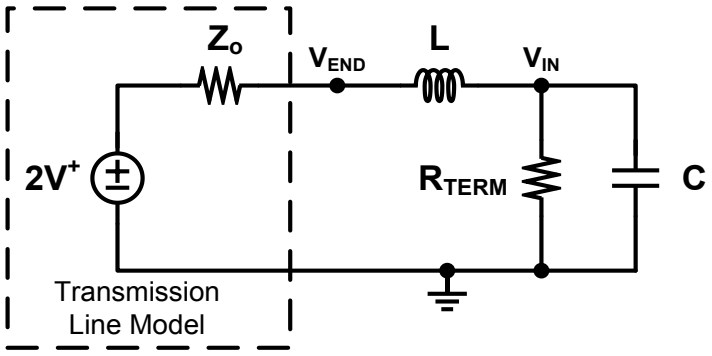


Figure 4.1: Receiver input lumped element model with a primary-only ESD protected circuit. $C = C_{BONDPAD} + C_{ESD} + C_{RX}$.

4.3 Receiver Input Transfer Function

4.3.1 Transfer Function Derivation

Using a lumped element model of the receiver permits the derivation of a transfer function, *H(s)*, in the Laplace domain to relate V_{IN} to V^+.

$$H(s) = \frac{V_{IN}}{V^+} = \frac{a_0 + a_1 s + a_2 s^2 + \ldots + a_m s^m}{1 + b_1 s + b_2 s^2 + \ldots + b_n s^n}. \tag{4.1}$$

Deriving the transfer function of a multiple element RLC circuit can be tedious, and thus it is preferable to use short-cuts to obtain the coefficients, see, for example, [20]. Once the coefficients are obtained, the transfer function can be utilized to perform a root locus analysis for the elements of the model.

In this dissertation, it is assumed that the termination has been matched to the characteristic impedance of the trace, i.e., $R_{TERM} = Z_O$. The effect of mismatch between the termination and characteristic impedance can also be investigated through the methods presented, but this dissertation will focus on the effect of the parasitic reactive elements. The transfer function for the circuit of Figure 4.1 is

$$H(s) = \frac{V_{IN}}{V^+} = \frac{\frac{1}{2}}{1 + \frac{1}{2}\left(\frac{L}{Z_O} + Z_O C\right)s + \frac{LC}{2}s^2}. \tag{4.2}$$

4.3.2 Transfer Function Coefficient Derivations

To derive the coefficients of *H(s)*, the methods presented in [20] are utilized. The circuit in Figure 4.1 has no significant zero since no reactive element can be infinite valued without making the resultant transfer constants, $H^{ijk\cdots}$, zero [20]. Therefore,

$$a_i = 0, \text{for } i \geq 1. \tag{4.3}$$

Since the numerator of the transfer function is a constant, no finite zeros exist. Solving for the DC gain,

$$a_0 = \frac{R_{TERM}}{Z_o + R_{TERM}} \tag{4.4}$$

$$= \frac{1}{2}. \tag{4.5}$$

The circuit of Figure 4.1 has two reactive elements; therefore,

$$b_i = 0, \text{for i} \geq 3. \tag{4.6}$$

Solving for the two denominator coefficients,

$$b_1 = \frac{L}{Z_o + R_{TERM}} + (Z_o || R_{TERM})C \tag{4.7}$$

$$= \frac{1}{2}\left(\frac{L}{Z_o} + Z_o C\right) \tag{4.8}$$

$$b_2 = \frac{L}{Z_o + R_{TERM}} R_{TERM} C \tag{4.9}$$

$$b_2 = \frac{LC}{2} \tag{4.10}$$

4.4 Receiver Input Root Locus

To determine the pole locations of *H(s)*, let us define, and set to zero, the denominator, *D(s)*, of (4.2)

$$D(s) = 1 + \frac{1}{2}\left(\frac{L}{Z_o} + Z_o C\right)s + \frac{LC}{2}s^2 = 0\,. \tag{4.11}$$

To demonstrate how the magnitude of the parasitic capacitance, *C*, impacts the poles of the receiver input, a root locus analysis will be performed. Traditionally, root locus analysis is performed on closed-loop systems with feedback, and the forward gain is utilized as a variable. In this dissertation, the root locus analysis will instead demonstrate how the receiver input poles change as a function of the parasitic capacitance. Rearranging (4.11) in an attempt to generate an equation suitable for root locus analysis produces

$$1 + C\frac{\frac{Z_O}{2}s + \frac{L}{2}s^2}{1 + \frac{L}{2Z_O}s} = 0\,. \tag{4.12}$$

Since the order of the numerator is higher than the order of the denominator, a result of applying the root locus method to a receiver input, (4.12) is improperly constructed for root locus analysis. The equivalent system

$$1 + \frac{1}{C}\frac{1 + \frac{L}{2Z_O}s}{\left(\frac{Z_O}{2}s + \frac{L}{2}s^2\right)} = 0 \tag{4.13}$$

can be used to generate the standard root locus plots [22]. In the following analysis, the root locus will be presented as if C, instead of $1/C$, is the independent variable which is changing from zero to infinity along the root locus branches.

The root locus plot for parasitic capacitance is shown in Figure 4.2. The root locus diagram is normalized by a factor of Z_o/L, as the shape of the root locus is independent of the characteristic impedance and the parasitic inductance. As the characteristic impedance or inductance is modulated, the root locus plot scales accordingly. Additionally, the characteristic impedance and parasitic inductance impact where on the root locus branches the poles will be located.

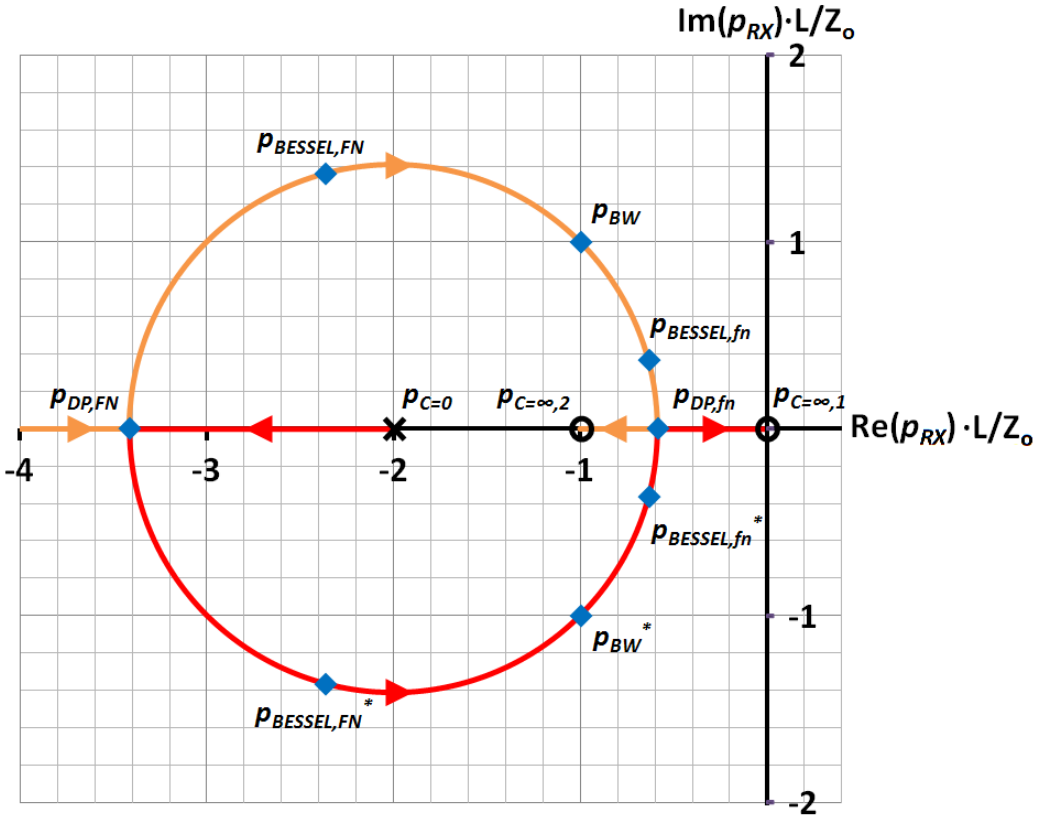


Figure 4.2: Root locus plot, as a function of parasitic capacitance, for the circuit of Figure 4.1.

For a capacitance of zero, there is one finite (and dominant) real pole at

$$p_{C=0} = -2\frac{Z_o}{L}. \tag{4.14}$$

As *C* is increased above zero, the dominant pole increases in magnitude while the second pole becomes finite with a monotonically decreasing magnitude until

$$C_{DP,FN} = 0.172\frac{L}{{Z_O}^2}, \tag{4.15}$$

at which point the system has a double real pole at

$$p_{DP,FN} = -(2+\sqrt{2})\frac{Z_o}{L}. \tag{4.16}$$

As *C* is increased above $C_{DP,FN}$, the two poles become complex conjugates along a circle with center on the real axis at

$$p_{C=\infty,2} = -\frac{Z_o}{L} \tag{4.17}$$

and radius $\sqrt{2}\,Z_o/L$. For any point on the circle with angle φ with respect to the real axis,

$$C = \frac{\sqrt{2}L}{{Z_o}^2[3\sqrt{2}-4\cos\varphi]}. \tag{4.18}$$

As *C* is further increased to

$$C_{DP,fn} = 5.282\frac{L}{{Z_O}^2}, \tag{4.19}$$

the complex conjugate poles meet on the real axis to form a double real pole at

$$p_{DP,fn} = -(2-\sqrt{2})\frac{Z_o}{L}. \tag{4.20}$$

As *C* is further increased above $C_{DP,fn}$, the poles split and move along the negative real axis in opposite directions, one approaching the origin,

$$p_{C=\infty,1} = 0\,, \tag{4.21}$$

and the other pole approaching $p_{C=\infty,2}$ for infinite *C*.

Between $C_{DP,FN}$ and $C_{DP,fn}$, the poles are complex conjugates and can generate a Butterworth filter or one of two Bessel filters. The pole locations which generate a Butterworth filter, p_{BW} and ${p_{BW}}^*$, occur for the damping ratio

$$\zeta_{BW} = \frac{\sqrt{2}}{2}. \tag{4.22}$$

The damping ratio, ζ, is equal to sin(θ), where θ is the angle of the pole with respect to the imaginary axis [22]. Additionally, ζ_{BW} is the minimum damping ratio achievable along the root locus branches. The sets of poles which generate the Bessel filter occur for the damping ratio

$$\zeta_{BESSEL} = \frac{\sqrt{3}}{2}. \tag{4.23}$$

The two possible Bessel filters will be distinguished by their relative natural frequency, f_n, which is the distance from the pole to the origin. The poles which generate the high f_n Bessel filter are $p_{BESSEL,FN}$ and $p_{BESSEL,FN}{}^*$, while the poles which generate the low f_n Bessel filter are $p_{BESSEL,fn}$ and $p_{BESSEL,fn}{}^*$.

Table 4.1 presents a list of capacitances which generate the specific pole discussed above and highlighted on the root locus plot in Figure 4.2. The capacitances are normalized by a factor $L/{Z_o}^2$, resulting in a proportional relationship to inductance and inverse relationship with characteristic impedance for specific pole locations on the root locus branches. For each pole location, Table 4.1 also lists the normalized natural frequency of the poles, f_n, the ratio of natural frequency to −3 dB bandwidth for the poles, f_{-3dB}/f_n, and the −3 dB bandwidth, f_{-3dB}.

Table 4.1: Root Locus Pole Locations

Normalized Capacitance $C \cdot \frac{{Z_O}^2}{L}$	Pole(s)	Normalized Natural Frequency $f_n \cdot \frac{L}{Z_O}$	$\frac{f_{-3dB}}{f_n}$	Normalized −3 dB Bandwidth $f_{-3dB} \cdot \frac{L}{Z_O}$
0	$p_{C=0}$	0.318	1	0.318
0.172	$p_{DP,FN}$	0.543	0.644	0.350
0.268	$p_{BESSEL,FN}$, $p_{BESSEL,FN}{}^*$	0.435	0.786	0.342
1	p_{BW}, $p_{BW}{}^*$	0.225	1	0.225
3.732	$p_{BESSEL,fn}$, $p_{BESSEL,fn}{}^*$	0.117	0.786	0.092
5.828	$p_{DP,fn}$	0.093	0.644	0.060
∞	$p_{C=\infty,1}$, $p_{C=\infty,2}$	0	1	0

4.4.1 Root Locus Double Root Proof

Multiple root locations on the root locus can be determined by solving

$$0 = (b\frac{da}{ds} - a\frac{db}{ds}) \tag{4.24}$$

where *b(s)* and *a(s)* are the numerator and denominator, respectively, of the root locus transfer function *L(s)* [22], where *L(s)* is determined from the root locus characteristic equation (4.13):

$$L(s) = \frac{1+\frac{L}{2Z_o}s}{\left(\frac{Z_o}{2}s+\frac{L}{2}s^2\right)[F]}. \tag{4.25}$$

Solving (4.24) utilizing (4.25),

$$0 = \left(1+\frac{L}{2Z_o}s\right)\left(\frac{Z_o}{2}+Ls\right) - \left(\frac{Z_o}{2}s+\frac{L}{2}s^2\right)\frac{L}{2Z_o})$$

$$0 = \frac{L^2}{4Z_o}s^2 + Ls + \frac{Z_o}{2}$$

$$s = -(2 \pm \sqrt{2})\frac{Z_o}{L}. \tag{4.26}$$

Therefore, double real poles occur at locations defined by (4.26). Additionally, the magnitude of parasitic capacitance corresponding to the double real pole locations can be determined by replacing (4.26) in (4.13):

$$C = \frac{L}{{Z_o}^2[3\mp 2\sqrt{2}]}. \tag{4.27}$$

4.4.2 Root Locus Circle Proof

To demonstrate that the circle

$$(s - p_{C=\infty,2}) = \frac{\sqrt{2}Z_o}{L}e^{j\varphi} \tag{4.28}$$

lies on the root locus branches, $L(s = p_{C=\infty,2} + \frac{\sqrt{2}Z_o}{L}e^{j\varphi})$ must have phase equal to –π with the corresponding *C* greater than or equal to 0, where *φ* is the angle of the point on the circle with respect to the positive real axis. Plugging (4.28) into (4.25) produces

$$L(s = p_{C=\infty,2} + \frac{\sqrt{2}Z_o}{L}e^{j\varphi}) = \frac{\frac{\sqrt{2}}{2}e^{j\varphi}}{\left(-\frac{2Z_o}{L}+\frac{\sqrt{2}Z_o}{L}e^{j\varphi}\right)\left(-\frac{Z_o}{2}+\frac{\sqrt{2}Z_o}{2}e^{j\varphi}\right)}$$

$$= \frac{\frac{\sqrt{2}}{2}}{\left(-\frac{2Z_o}{L}+\frac{\sqrt{2}Z_o}{L}e^{j\varphi}\right)\left(-\frac{Z_o}{2}e^{-j\varphi}+\frac{\sqrt{2}Z_o}{2}\right)}$$

$$= \frac{\frac{\sqrt{2}}{2}}{\frac{{Z_o}^2}{L}e^{-j\varphi}-\frac{\sqrt{2}{Z_o}^2}{L}-\frac{{Z_o}^2}{\sqrt{2}L}+\frac{{Z_o}^2}{L}e^{j\varphi}}$$

$$= \frac{\sqrt{2}L}{{Z_o}^2[4\cos\varphi-3\sqrt{2}]}, \tag{4.29}$$

which has phase –π for all *φ*. Resolving (4.13) with the simplified *L(s)* from (4.19) assigns

$$C = \frac{\sqrt{2}L}{{Z_o}^2\left[3\sqrt{2}-4\cos\varphi\right]} \tag{4.30}$$

for any point on the circle, which is indeed positive for all φ. Therefore, all points on the circle given by (4.28) lie on the root locus branches.

4.5 Maximum −3 dB Bandwidth

Since random NRZ data has a wideband spectrum, the bandwidth of the receiver input can significantly impact the signal integrity. In general, the higher the receiver input bandwidth, the higher the data rate the receiver can support. As such, maximizing the −3 dB bandwidth at the input of the receiver is important for multi-Gb/s serial links.

When the magnitude of parasitic capacitance, *C*, is increased above zero in the root locus plot in Figure 4.2, the root locus branch departs $p_{C=0}$ along the real axis, away from the origin. Therefore, as the parasitic capacitance is increased from a value of zero, the dominant pole moves to higher frequencies. In other words, due to the parasitic inductance, bandwidth is not a monotonically decreasing function of *C*; for the circuit in Figure 4.1, the optimum value of parasitic capacitance is greater than zero. In Figure 4.3, the exact −3 dB bandwidth of the circuit in Figure 4.1 is plotted as a function of parasitic capacitance. Both the capacitance and bandwidth are normalized by the same factors as the root locus plot in Figure 4.2. As noted above, increasing *C* above zero actually increases the bandwidth until a maximum is achieved, followed by an asymptotical decrease toward zero. The derivation of Section 4.5.1 proves how at the double real pole, $p_{DP,FN}$, where

$$C_{DP,FN} \approx 0.172\frac{L}{{Z_o}^2}, \tag{4.31}$$

the −3 dB bandwidth, $f_{-3dB,MAX}$, reaches a maximum of

$$f_{DP,FN} = f_{-3dB,MAX} \approx 0.35\frac{Z_o}{L}. \tag{4.32}$$

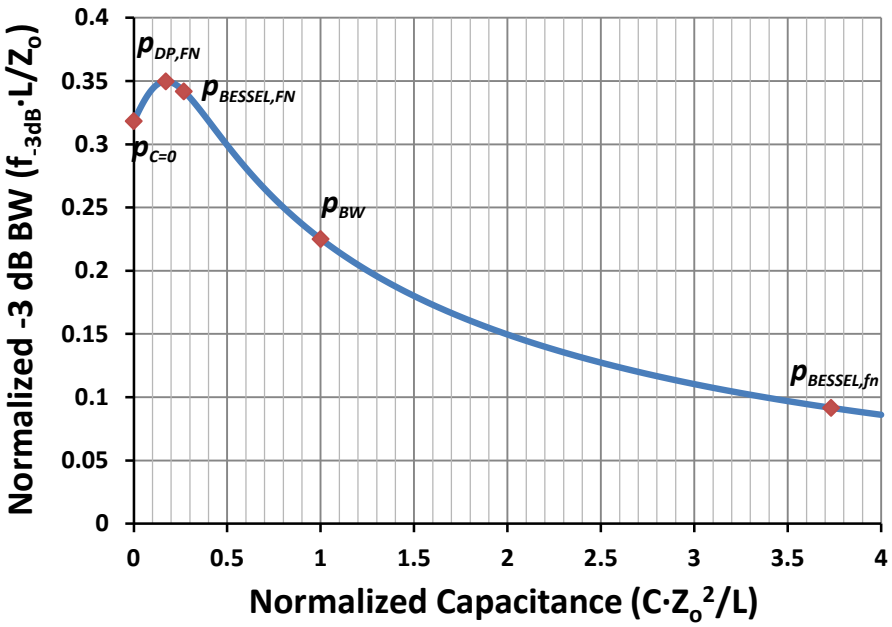


Figure 4.3: Normalized −3 dB bandwidth ($f_{-3dB} \cdot L/Z_o$) as a function of normalized parasitic capacitance (C·$Z_o{}^2/L$).

4.5.1 Exact and Maximum −3 dB Bandwidth Derivation

To determine an equation for the exact bandwidth of the second order system (4.2) of Figure 4.1, it is easiest to normalize the capacitance by

$$C_o = \frac{Z_o{}^2}{L} C \tag{4.33}$$

and frequency by

$$s_o = \frac{L}{Z_o} s \tag{4.34}$$

to produce

$$H_o(s_o) = \frac{V^+}{V_{IN}} = \frac{1}{1+\frac{1}{2}(1+C_o)s_o+\frac{C_o}{2}s_o{}^2}. \tag{4.35}$$

The −3 dB bandwidth occurs when the magnitude of the denominator of (4.35) is $\sqrt{2}$; therefore, the equation which defines the −3 dB bandwidth, $\omega_{o,-3dB}$, in angular frequency is

$$\sqrt{2} = \left\| 1 + \frac{1}{2}(1 + C_o)s_{o,-3dB} + \frac{C_o}{2}s_{o,-3dB}{}^2 \right\|$$

$$2 = (1 - \frac{C_o}{2}\omega_{o,-3dB}{}^2)^2 + \frac{1}{4}(1 + C_o)^2\omega_{o,-3dB}{}^2$$

$$2 = 1 + (\frac{1}{4}(1 + C_o)^2 - C_o{}^2)\omega_{o,-3dB}{}^2 + \frac{C_o{}^2}{4}\omega_{o,-3dB}{}^4$$

$$0 = -4 + ((1 + C_o)^2 - 4{C_o}^2){\omega_{o,-3dB}}^2 + {C_o}^2{\omega_{o,-3dB}}^4$$

$$0 = -4 + (C_o - 1)^2{\omega_{o,-3dB}}^2 + {C_o}^2{\omega_{o,-3dB}}^4 . \quad (4.36)$$

Therefore, the −3 dB bandwidth, $f_{o,-3dB}$, is equal to

$$f_{o,-3dB} = \frac{1}{2\pi}\sqrt{\frac{-(C_o-1)^2+\sqrt{(C_o-1)^4+16{C_o}^2}}{2C_o}} . \quad (4.37)$$

Additionally, it can be shown that the maximum bandwidth occurs at $p_{DP,FN}$, (4.16), with a normalized capacitance of

$$C_o = \frac{1}{3+2\sqrt{2}}, \quad (4.38)$$

since

$$\frac{f_{o,-3dB}}{\partial C_o}\Big|_{C_o=\frac{1}{3+2\sqrt{2}}} = 0 . \quad (4.39)$$

While proving (4.39) is complicated, it is much simpler to compare $f_{o,-3dB}$, (4.37), for $C_o = \frac{1}{3+2\sqrt{2}}$ and $C_o = \frac{1}{3+2\sqrt{2}} \pm \delta$ to demonstrate a decrease in −3 dB bandwidth for any non-zero δ.

4.5.2 Evaluation of Bandwidth Estimation Method

In Figure 4.4, the exact −3 dB bandwidth, Figure 4.3, is compared to the −3 dB bandwidth estimate generated in Chapter 3. For the receiver input lumped element model of Figure 4.1, the dominant single pole approximation does not reproduce the relationship between increasing capacitance and the −3 dB bandwidth for capacitances below the maximum −3 dB bandwidth, $f_{-3dB,MAX}$. Therefore, the single pole estimate is insufficient for analyzing the effect of parasitic capacitance at very low magnitudes. The maximum bandwidth estimation error occurs for capacitances slightly less than $C_{DP,FN}$, resulting in a maximum underestimate of 22.3%. For capacitance greater than or equal to $C_{DP,FN}$, the second-order bandwidth estimate is utilized and exactly calculates the −3 dB bandwidth, since the circuit of Figure 4.1 is a second-order system.

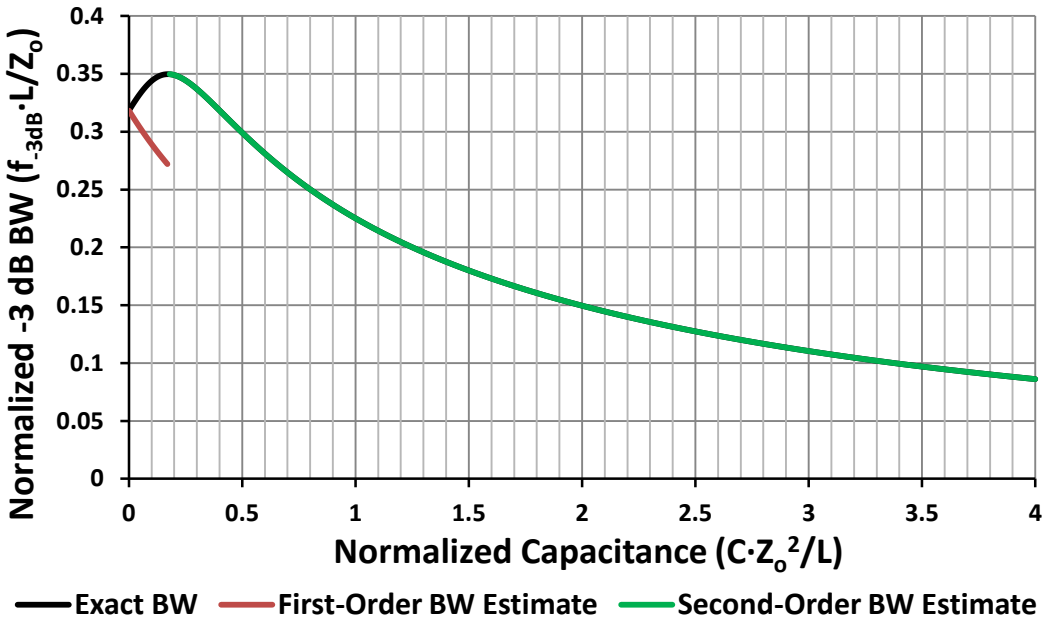


Figure 4.4: Normalized −3 dB bandwidth ($f_{-3dB} \cdot L/Z_o$), first-order bandwidth estimate, and second-order bandwidth estimate as a function of normalized parasitic capacitance ($C{\cdot}{Z_o}^2/L$).

4.6 Optimal Capacitance for High-Speed Serial I/O

While maximizing the −3 dB bandwidth attempts to minimize distortion due to attenuation, distortion of the received signal due to phase can also degrade signal integrity. In Section 4.4, pole locations which generate Bessel and Butterworth filter responses were highlighted. The Bessel filter response is advantageous for producing maximally linear phase while Butterworth filter response is advantageous for producing maximally flat magnitude [23]. Both Table 4.1 and Figure 4.3 demonstrate a significant reduction in the −3 dB bandwidth, with respect to $f_{-3dB,MAX}$, for poles which generate the Butterworth and low f_n Bessel filter responses. The poles which generate the high f_n Bessel filter response, $p_{BESSEL,FN}$ and ${p_{BESSEL,FN}}^*$, result in a −3 dB bandwidth of

$$f_{-3dB,BESSEL} \approx 0.342 \frac{Z_o}{L}, \tag{4.40}$$

an insignificant 2.3% reduction from $f_{-3dB,MAX}$. Considering the maximally linear phase property of the Bessel filter response, designing the receiver input poles such that they occur at $p_{BESSEL,FN}$ and ${p_{BESSEL,FN}}^*$ optimizes the phase response with an insignificant reduction in −3 dB bandwidth. An additional benefit of the high f_n Bessel filter is a 58% increase in the parasitic capacitance with respect to $C_{DP,FN}$ to

$$C_{BESSEL} = C_{OPT} \approx 0.268 \frac{L}{{Z_o}^2}, \tag{4.41}$$

permitting a parasitic capacitance 56% larger than $C_{DP,FN}$. These factors contribute to a recommended optimal design target of $p_{BESSEL,FN}$ and ${p_{BESSEL,FN}}^*$ receiver input poles. Unfortunately, in receiver applications with low parasitic inductance or high ESD resiliency requirements, achieving C_{BESSEL} may not be feasible. For these receivers, bandwidth extension circuits, such as negative capacitance circuits, can be utilized to partially mitigate the capacitance at the cost of power [9], [10], [24].

A secondary approach to optimizing the capacitance for high-speed I/O receivers is to utilize the receiver input poles as an inherent noise filter, designing the parasitics to achieve a desired response. The range of achievable filters for the receiver input circuit model of Figure 4.1 is given by the root locus plot of Figure 4.2.

4.6.1 Filter Capacitance Derivation

It can be shown that a Butterworth filter response can be generated with an angle $\varphi = 45°$ on the circle defined by (4.28), generating the poles p_{BW} and ${p_{BW}}^*$. The capacitance required to generate the Butterworth filter response, (4.18), is

$$C = \frac{L}{{Z_o}^2}. \tag{4.42}$$

It can be shown that two Bessel filter responses can be generated with angles $\varphi = 15°$ and $\varphi = 105°$ on the circle defined by (4.28). $\varphi = 15°$ generates the poles $p_{BESSEL,fn}$ and ${p_{BESSEL,fn}}^*$ for the low f_n Bessel response, while $\varphi = 105°$ generates poles and $p_{BESSEL,FN}$ and ${p_{BESSEL,FN}}^*$ for the high f_n Bessel response. The capacitance required to generate the poles for the low f_n Bessel response is

$$C = \frac{L}{(2-\sqrt{3}){Z_o}^2} = 3.732, \tag{4.43}$$

while the capacitance required to generate the poles for the high f_n Bessel response is

$$C = \frac{L}{(2+\sqrt{3}){Z_o}^2} = 0.268 \frac{L}{{Z_o}^2}. \tag{4.44}$$

4.7 Parasitic Capacitance Budget for High-Speed I/O

High-speed serial links are often produced through the efforts of multiple designers. To guide designers, budgets for parasitic elements are often generated. One such budget would limit the magnitude of parasitic capacitance at the receiver input. The parasitic capacitance budget can then be utilized to guide the ESD protection network, bondpad, and receive amplifier design.

Utilizing a bandwidth limit based on the data rate of a link and the −3 dB bandwidth, an appropriate first-order capacitance budget can be formulated. A signal integrity rule of thumb is to maintain

$$f_{-3dB} \geq 0.75 \cdot Data\ Rate\ , \tag{4.45}$$

as signal integrity degrades rapidly below 0.75 [25]. In this dissertation, the utilization of only binary NRZ data is assumed, so data rate and baud (and their units) will be used interchangeably. Utilizing (4.45) and scaled data from Figure 4.3, capacitance budgets for various magnitudes of parasitic inductance for a 50 Ω characteristic impedance link can be generated, as shown in Figure 4.5.

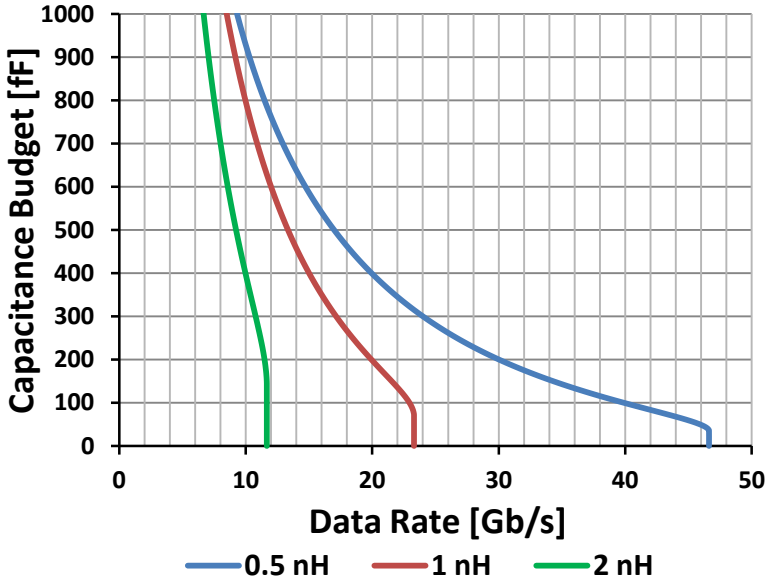


Figure 4.5: Parasitic capacitance budgets for parasitic inductances of 0.5 nH, 1 nH, and 2 nH. Budget assumes the receiver input bandwidth is the primary bandwidth limitation.

Figure 4.5 demonstrates how drastically parasitic inductance can impact the parasitic capacitance budget. At data rates below 5 Gb/s, parasitic inductance and capacitance

combinations up to 2 nH and 1 pF do not generate a receiver input bandwidth that will dramatically degrade the signal integrity. For data rates between 5 Gb/s and 10 Gb/s, inductance will significantly impact the ESD capacitance budget. For example, at 10 Gb/s, a reduction in parasitic inductance from 1 nH to 0.5 nH doubles the parasitic capacitance budget from 400 fF to 800 fF. For data rates above 10 Gb/s, parasitic inductance on the order of 1 nH can limit the maximum achievable data rate. For these high data rates, the sensitivity to parasitic capacitance has also increased to the point that the addition of 100 fF of capacitance can drastically impact the achievable data rate.

In the capacitance budget generation, multiple assumptions were utilized, so the engineering challenge occurs when the true complexity of the link is considered. Other signal integrity degradation factors, such as channel induced ISI and limited transmitter bandwidth, may require an increase in the bandwidth-to-data-rate ratio requirement given in (4.45), decreasing the parasitic capacitance budget. However, permitting a degree of signal integrity degradation or the utilization of equalization [21] may decrease the bandwidth-to-data-rate ratio requirement, (4.45), allowing for an increase in the parasitic capacitance budget.

4.8 Parasitic Inductance Budget for High-Speed I/O

Another possible budget would limit the magnitude of packaging-related parasitic inductance. This budget could then be utilized to select appropriate packaging, guide package substrate design, and determine package pin-out. Utilizing the maximum achievable bandwidth, $f_{-3dB,MAX}$, and the bandwidth limit based on the data rate of a link, (4.45), a first-order inductance budget can be formulated as shown in Figure 4.6. For a Bessel filter response, the inductance budget is reduced by 2.3%.

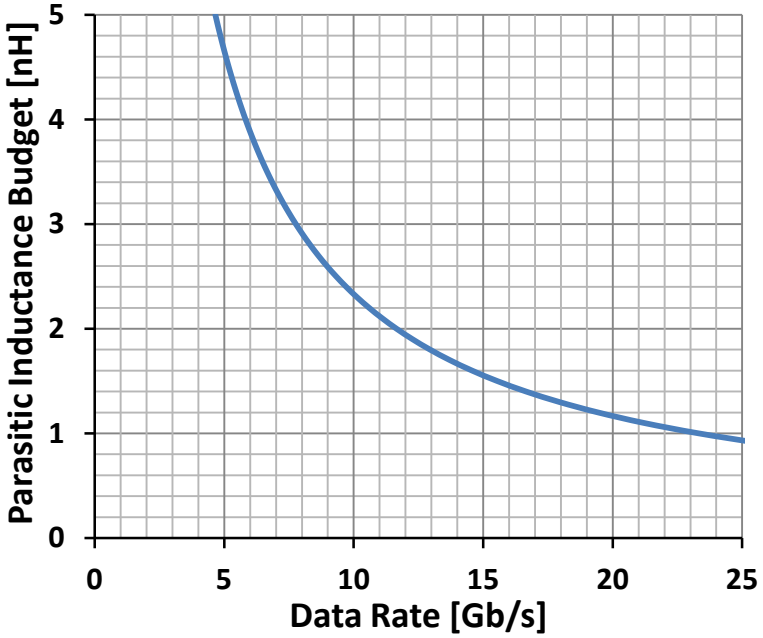


Figure 4.6: Parasitic inductance budget as a function of data rate. Budget assumes the parasitic capacitance can be set to $C_{DP,FN}$, to maximize the −3 dB bandwidth. For low magnitudes of parasitic inductance, it is important to factor in bandwidth reduction due to capacitance from Figure 4.3.

For many applications, including high data rates, it may not be feasible to achieve the parasitic capacitance for the maximum achievable bandwidth, $C_{DP,FN}$, or Bessel filter, $C_{BESSEL,FN}$. For these applications, an estimation of the percentage reduction of −3 dB bandwidth due to capacitance (estimated via Figure 4.3) should be used to reduce the inductance budget.

4.9 Transmitter Output Root Locus

In Section 3.2, a lumped element model for a transmitter, shown in Figure 3.2, was generated. Similar to the receiver input poles, the transmitter output poles can degrade the signal integrity. By observation, the sourceless lumped element model of Figure 3.2 is identical to that of the receiver input, Figure 4.1, with relabeled nodes. Since the poles of a circuit are independent of the choice of input or output [20], the poles of Figure 3.2 are identical to the poles of Figure 4.1. Additionally, by inspection, the transfer function for the network of Figure 3.2 does not contain any finite zeroes [20], so the entire root locus and parasitic element budget analysis of the receiver of Figure 4.1 can be applied directly to the lumped element model of the transmitter in Figure 3.2.

CHAPTER 5

NEGATIVE CAPACITANCE CIRCUITS

5.1 Introduction

In Chapters 3 and 4, the effect of ESD protection on the signal integrity of multi-Gb/s wireline transmitters and receivers was investigated. In this chapter, test chip measurement data are provided to quantify the effect of ESD protection on the signal integrity at the receiver input as a function of the data rate, and also to validate the lumped model analysis. It will be shown that, given the wideband nature of wireline signaling, excellent signal integrity can obtained at data rates in excess of 5 Gb/s with high levels of ESD resiliency. At higher data rates, e.g., 10 Gb/s, reduced signal integrity is detected. If this has a detrimental effect on the bit error rate (BER), the designer can partially mitigate the deleterious loading by the ESD protection devices by using a negative capacitance circuit. Placed at a receiver input, a negative capacitance circuit has an effect similar to that of a T-coil circuit [8] – it increases the input pole frequency, thereby improving the receiver bandwidth. A negative capacitance circuit is an active circuit and thus is far more compact than the inductor-based T-coil, but as in all active circuits, it consumes power. Additionally, a novel negative capacitance circuit, the *g_m-boosted negative capacitance circuit*, is implemented and measured for the first time. The effect of the negative capacitance circuits on the signal integrity is quantified, and the ESD resiliency of receivers with negative capacitance circuits is evaluated.

5.2 Test Chip Overview and Measurement Methodology

A 2 mm x 2 mm receiver test chip was designed and manufactured in a low-power 90 nm CMOS technology and packaged within a 6 mm x 6 mm open-cavity QFN. The package was attached to a custom high-speed test board using a high bandwidth elastomer socket for signal integrity measurements. The socket is attached to a custom 4-inch-square, two-layer test board manufactured using Rodgers 4003 dielectric. A photograph of the high-speed test board is shown

in Figure 5.1. Grounded coplanar transmission lines are used to implement the high-speed 50 Ω traces used for the receiver inputs and outputs.

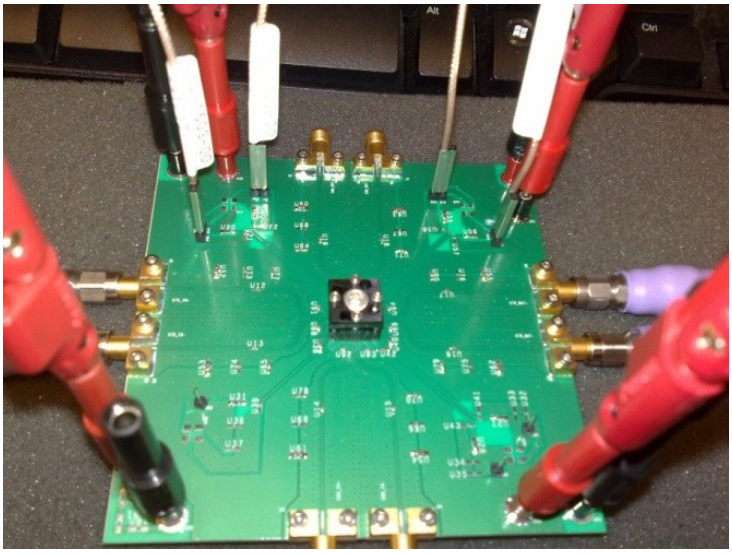


Figure 5.1: Test board photograph. A high bandwidth elastomer socket is used to attach the QFN packaged test chip to the test board.

On the receiver test chip, the same receiver design was instantiated six times, each with a different front-end. In this dissertation, the front-end is defined as all of the circuitry at the input of the receiver up to the receive amplifier; that is, it consists of the input ESD protection circuit and the transmission line termination circuit. The die micrograph is shown in Figure 5.2, with the six receiver instantiations labeled as 1A, 1B, 2A, 2B, 3A, and 3B. Each receiver consists of a fully-differential two-stage input amplifier and differential buffers leading off-chip for measurement purposes. Each amplifier stage and each buffer is implemented as a resistively loaded source-coupled differential amplifier with resistive source degeneration, as shown in Figure 5.3, and the amplifier bandwidth exceeds the receiver input bandwidth. Each die is offset within the package to center one of the receiver instantiations for testing and to lengthen the input bondwires to approximately 2.5 mm while shortening the output bondwires. The long, input bondwires are expected to produce an inductance of approximately 2.5 nH [26]. This series inductance, in combination with the other parasitics at the input of the receiver, ensures that the overall receiver bandwidth will be limited by the poles at the circuit input.

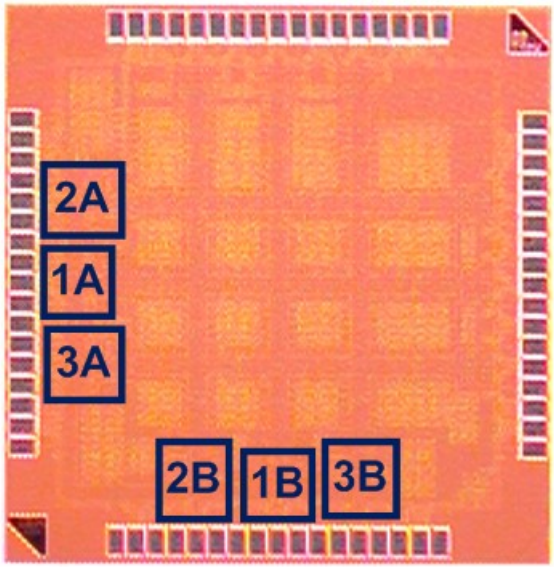


Figure 5.2: Test chip micrograph with receivers 1A, 2A, 3A, 1B, 2B, and 3B highlighted.

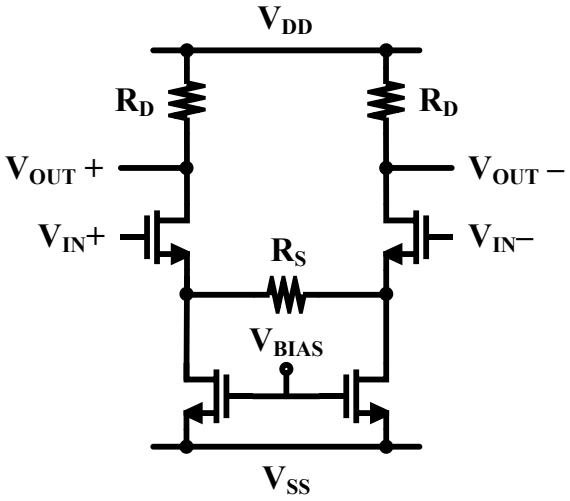


Figure 5.3: Resistively loaded source-coupled differential amplifier with resistive source degeneration. Used for both stages of the receive amplifier and for the differential buffers leading off-chip.

As noted previously, each of the six receiver instantiations has a unique front-end. Two ESD protection schemes, type A and type B, were implemented. The protection scheme of Figure 3.3 is type A, while the protection scheme of Figure 3.5 is type B. The type A circuit is intended to provide a 3 kV HBM protection level, and type B is intended to provide a 3 kV HBM protection level and a 500 V CDM protection level. Both schemes use dual-diode primary ESD protection circuits at the input pads, and active clamps between the power and ground rails [27], as shown in Figure 3.3. A rail clamp is instantiated underneath every power and ground pad on the die; this along with the wide power and ground bus rails virtually ensures that any ESD failures can be

attributed to failures at the receiver input, i.e., in the dual-diodes or the receiver input circuitry. In the circuits with type B protection, each of the receive amplifier inputs is decoupled from the pad by a 100 Ω series resistor and secondary dual-diodes are placed at the amplifier inputs. Both protection schemes followed the foundry design kit instructions and used the ESD library cells. The naming convention is such that receiver 2A contains type A input protection, whereas receiver 3B contains type B protection.

Receivers 1A and 1B do not contain a bandwidth extension circuit at the receiver input. These circuits allow one to study the impact of secondary protection on signal integrity. Receivers 2A and 2B contain a traditional negative capacitance circuit at the input, while receivers 3A and 3B contain a g_m-boosted negative capacitance circuit. The negative capacitance circuits are presented in Section 5.4.

The most effective demonstration of signal integrity is to generate an eye diagram by applying a pseudo-random bit sequence (PRBS) to the input of the receiver. In this dissertation, only binary NRZ data is used, so the data rate is equal to the baud rate and the two terms will be used interchangeably. The PRBS input to the receiver is reasonably ideal, such that any degradation of the signal integrity can be attributed to the receiver non-idealities. The differential input amplitude was attenuated to 125 mV_{pp}, to prevent clipping within the amplifier and consequent distortion in the eye diagram. The receiver output is then connected to a high-frequency oscilloscope which generates the eye diagram.

Since the available network analyzers do not perform differential S-parameter measurements with differential signaling, S-parameters could not be measured in order to evaluate the bandwidth extension circuits. To evaluate the bandwidth extension circuits, differential time-domain transmission (TDT) was measured instead. The differential TDT tester sends a differential step into the receiver which can be observed at the circuit output. The receiver input bandwidth will limit the minimum observable risetime. To quantify the improvement from the negative capacitance circuits, the risetime is measured and used to calculate a normalized bandwidth. This dissertation uses the 20% to 80% risetime obtained after filtering the raw data for noise. Since the frequency response is dominated by the input poles, a normalized bandwidth is calculated by taking the reciprocal of the risetime and normalizing it to that of the basic receiver. The TDT measurements are generated by connecting the receiver input and output to an oscilloscope with differential TDT capabilities.

The ESD resiliency of each of the receiver instantiations is evaluated using commercial HBM and FICDM testers. The HBM resiliency was determined by stressing between its IN+ pin and IN− pin, which is expected to be the worst-case HBM stress. Three samples were tested until failure, defined as a 10% change in the measured TDT risetime or amplitude for the receiver. The CDM resiliency was determined by stressing the input pins of the receivers until failure, once again defined as a 10% change in the TDT measurement results.

5.3 Basic Receiver Front-End

5.3.1 Circuit

Receivers 1A and 1B are the reference cases against which the bandwidth extended receivers will be compared. Receivers 1A and 1B use a basic transmission line termination circuit, shown in Figure 5.4. The basic input termination circuit incorporates the transmission line termination, input bias, and receiver offset cancellation. To generate the receiver input DC bias, a common-mode bias network [21] with DC offset cancellation is utilized.

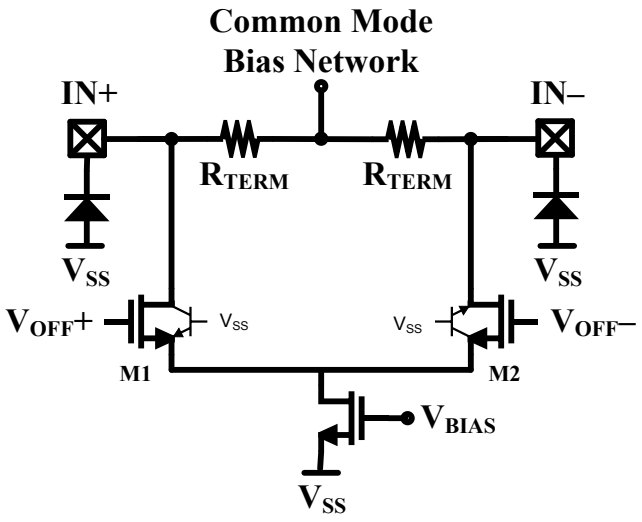


Figure 5.4: Basic transmission line termination circuit used in receiver instantiations 1A and 1B. IN+ and IN− are the input ports of the high-speed receiver. The circuit also incorporates the input bias and offset cancellation. The parasitic NPNs associated with M1 and M2 are indicated in the figure. Only the bottom part of dual-diode circuit is shown.

5.3.2 Signal Integrity Results

The 5, 7.5, and 10 Gb/s single-ended eye diagrams for receiver 1A are shown in Figure 5.5. At 5 Gb/s (Figure 5.5(a)), excellent signal integrity is obtained with the basic receiver front-end, as the eye is very open and the zero-crossing jitter is small. As the data rate increases to 7.5 Gb/s (Figure 5.5(b)), the eye begins to close and the zero-crossing jitter begins to increase. At a data rate of 10 Gb/s (Figure 5.5(c)), the signal integrity is further degraded, significantly impacting the eye opening and zero-crossing jitter.

The estimated bandwidth of circuit 1A is 5.7 GHz, and its Q value is 0.68. The eye diagrams demonstrate that when the ratio of the receiver input bandwidth to the data rate is greater than 0.75, the signal integrity is not significantly degraded [25]. For ratios below 0.75, the signal integrity degrades rapidly when the ratio is reduced [25].

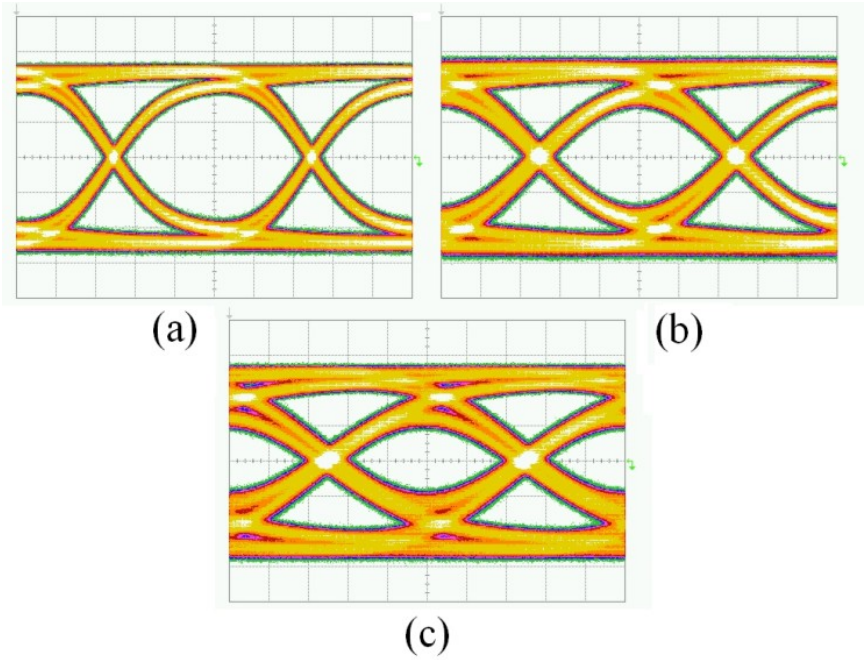
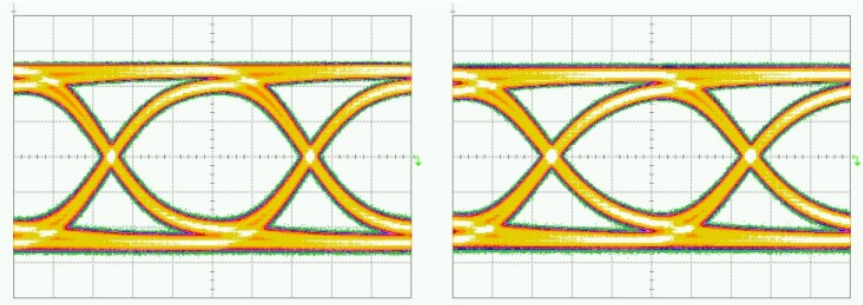


Figure 5.5: PRBS single-ended eye diagrams for receiver 1A at (a) 5, (b) 7.5, and (c) 10 Gb/s. Differential input amplitude is 125 mV_{pp}. [Horizontal scale: 40 ps/div. for (a), 26.7 ps/div. for (b) and 20 ps/div. for (c); vertical scale: 20 mV/div.].

The 5, 7.5, and 10 Gb/s single-ended eye diagrams for receivers 1A and 1B are shown in Figures 5.6, 5.7, and 5.8, respectively. At 5 Gb/s (Figure 5.6), both receivers have similar eye diagrams; then at 7.5 (Figure 5.7) and 10 Gb/s (Figure 5.8), the 1B receiver, with CDM protection, has a smaller eye opening. A quantitative comparison between the 10 Gb/s eye

diagrams of receivers 1A and 1B may be made using the data of Table 5.1. The estimated bandwidth of receiver 1B is 4.7 GHz, and its *Q* value is 0.68. At 5 Gb/s, for both receivers, the ratio of estimated bandwidth to data rate is greater than 0.75, resulting in little difference between the signal integrity of the two receivers, 1A and 1B. At 7.5 Gb/s, this ratio is 0.76 for receiver 1A and 0.63 for receiver 1B, explaining why there is noticeably more signal degradation at receiver 1B. At 10 Gb/s, this ratio is 0.57 for receiver 1A and 0.47 for receiver 1B, explaining why there is noticeably more signal degradation at receiver 1B. Notably, however, both eyes are still open at 10 Gb/s.



Figures 5.6: 5 Gb/s PRBS single-ended eye diagrams for (a) receiver 1A, (b) receiver 1B. Differential input amplitude is 125 mV_{pp}. [Horizontal scale: 40 ps/div.; vertical scale: 20 mV/div.].

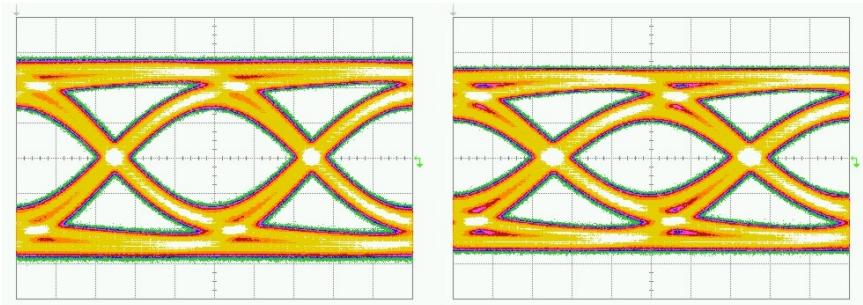
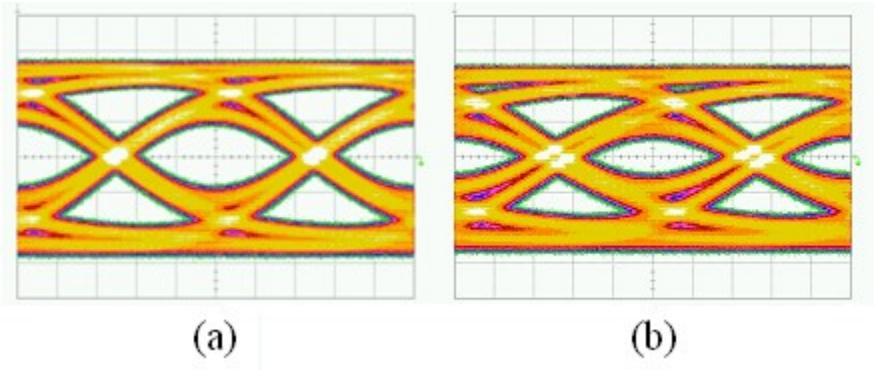


Figure 5.7: 7.5 Gb/s PRBS single-ended eye diagrams for (a) receiver 1A, (b) receiver 1B. Differential input amplitude is 125 mV_{pp}. [Horizontal scale: 26.7 ps/div.; vertical scale: 20 mV/div.].



Figures 5.8: 10 Gb/s PRBS single-ended eye diagrams for (a) receiver 1A, (b) receiver 1B. Differential input amplitude is 125 mV_{pp}. [Horizontal scale: 20 ps/div.; vertical scale: 20 mV/div.].

Table 5.1: 10 Gb/s PRBS Single-ended Eye Diagram Measurement Results for Receivers 1A and 1B

	Receiver 1A	Receiver 1B
Eye Amplitude [mV_{pp}]	98.7	93.1
Eye Height [mV_{pp}]	32.7	16.5
Jitter [ps_{pp}]	31.6	44.2

5.3.3 ESD Resiliency Results

Receiver 1A failed for HBM stresses above 4 kV; the amplitude of the eye was significantly degraded. Post-stress electrical analysis revealed that the offset cancellation transistors (M1 and M2 in Figure 5.4) were damaged. The two offset cancellation transistors were implemented using low-supply voltage core transistors and are in parallel with the termination resistors, thereby experiencing the entire pin voltage. Post-stress electrical measurements showed a drop in the differential input resistance of the receiver, matching the reduction of the eye amplitude. There was no change to the resistance measured between either receiver input pin and the common mode bias network, indicating that the termination resistors, R_{TERM}, were undamaged. Most importantly, manual modification of the offset feedback loop outputs (V_{OFF}+ and V_{OFF}–) failed to change the input bias point, which is consistent with there being drain-source damage to M1 and M2. To improve the ESD resiliency, the offset cancellation transistors should be implemented using I/O transistors; this would not impact performance as the offset cancellation has a low loop bandwidth.

During IN+ to IN− stress, V_{SS} is raised at least one diode V_{ON} above ground, due to the dual-diode input protection scheme. For positive HBM stresses, the lateral NPN associated with M2 will be on, due to the elevated VSS potential. However, the stress current will be shunted through the protection network until the IN+ pin voltage exceeds the drain breakdown voltage of M1, approximately 4 V. Once M1 turns on, ESD current will flow through M1 and M2, resulting in damage to both devices, judging from the post-stress electrical measurements.

All of the receiver 1A and 1B samples passed CDM testing on the input pins of the receiver, up to the maximum precharge voltage of the tester (>1500 V), i.e., no performance degradation occurred. At first glance, the results might be surprising, since the input amplifier circuit uses low-voltage thin-oxide transistors and half of the receivers do not contain an explicit CDM protection circuit. The high CDM resiliency is attributed to the limited charge storage of the small QFN package and die and the long bondwires at the input of the receiver, as the maximum peak current measured during any single CDM test was only 5.6 A This result highlights the danger of using small test chips to project CDM robustness of large product ICs, unless peak current rather than precharge voltage is used to quantify the stress.

5.4 Bandwidth Extension Circuits

5.4.1 Negative Capacitance Circuits

Using the methodology presented in Chapter 3, the −3 dB bandwidth limitation of a receiver input due to parasitic inductance and capacitance can be estimated. If the −3 dB bandwidth of the receiver limits the performance of the link, then the results of Chapter 3 can also be utilized to estimate the maximum achievable bandwidth given the ability to ideally mitigate capacitance. If the extended bandwidth improves the performance of the link, then a negative capacitance circuit can be utilized to partially mitigate capacitance [9], [10]. In previous works, the negative capacitance circuit is essentially regarded as ideal during analysis and the negative resistance component of the impedance is essentially disregarded. In this dissertation, the impacts of both negative capacitance and negative resistance components on the receiver input poles are included to demonstrate limitation of the circuit so that an optimal design point can be determined.

5.4.2 Traditional Negative Capacitance Circuit

The traditional negative capacitance circuit [9] may be integrated into the transmission line termination circuit, as shown in Figure 5.9. In this configuration, all of the typical functions of the termination circuit, including the transmission line termination, input bias, and offset cancellation, are preserved, while also implementing bandwidth extension. In Figure 5.9, C_C is the source degeneration capacitor, C_P is the parasitic capacitance at the drain of each current source transistor, C_{PAD} is the parasitic capacitance at the input pad, and R_{TERM} is an on-chip matching resistor.

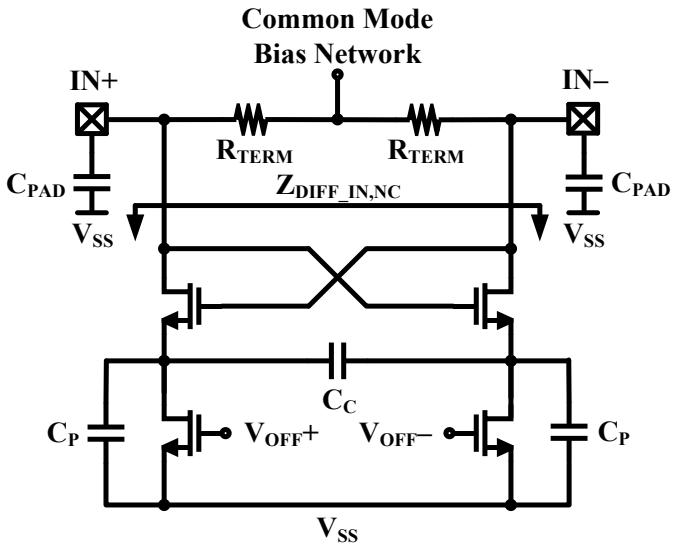


Figure 5.9: Traditional negative capacitance circuit. Utilized in receiver instantiations 2A and 2B. The circuit also incorporates the transmission line termination, input bias, and offset cancellation.

Assuming the output impedance of the cross-coupled transistors is large, the differential impedance looking into the drain terminals of the cross-coupled pair, $Z_{DIFF\_IN,NC}$, may be written as

$$Z_{DIFF\_IN,NC} = -\frac{2}{g_m} - \left(\frac{1}{j\omega C_C} \parallel \frac{2}{j\omega C_P} \parallel 2r_o\right), \tag{5.1}$$

where g_m is the transconductance of the cross-coupled transistors, and r_o is the output impedance of the current source transistors [10]. The negative capacitance circuit reactance has two regions: a positive inductance region and a negative capacitance region, divided by a peak in the reactance of

$$MAX\left(Im\{Z_{DIFF\_IN,NC}\}\right) = -\mathrm{r_o}\,, \tag{5.2}$$

at a frequency of

$$f_{PEAK} = \frac{1}{4\pi r_o(C_C+\frac{1}{2}C_P)}\,[10]. \tag{5.3}$$

In the negative capacitance region, the impedance $Z_{DIFF\_IN,NC}$ is well approximated as

$$Z_{DIFF\_IN,NC} \approx -\frac{2}{g_m} - \frac{1}{j\omega(C_C+\frac{C_P}{2})}\,. \tag{5.4}$$

Assuming the receiver input is driven differentially, the differential impedance can be converted into a single-ended impedance for each input pin,

$$Z_{IN,NC} \approx -\frac{1}{g_m} - \frac{1}{j\omega(2C_C+C_P)}\,, \tag{5.5}$$

to an AC ground. The impedance consists of a negative resistance and an effective negative capacitance of

$$C_{NEG} = 2C_C + C_P\,. \tag{5.6}$$

In Figures 5.10 and 5.11, the normalized magnitude and phase of (5.5) is plotted against the single-ended impedance of an ideal negative capacitance circuit with magnitude C_{NEG}:

$$Z_{IN,IDEAL\_NC} \approx -\frac{1}{j\omega C_{NEG}}\,. \tag{5.7}$$

At frequencies below

$$f_{CEILING} = \frac{\mathrm{g_m}}{2\pi C_{NEG}} \tag{5.8}$$

in the negative capacitance region, the traditional negative capacitance circuit presents an acceptable approximation to the ideal negative capacitance circuit, while at frequencies greater than $f_{CEILING}$, the traditional negative capacitance circuit presents a differential impedance approximately equal to that of a negative resistance circuit

$$Z_{DIFF\_IN,NR} \approx -\frac{2}{g_m}\,. \tag{5.9}$$

This high-frequency impedance occurs when the impedance of the degeneration capacitor becomes relatively small, transforming the circuit into a standard cross-coupled pair. Therefore, in the proximity of $f_{CEILING}$, the traditional negative capacitance circuit effectively transitions from a negative capacitance circuit to a negative resistance circuit as frequency increases. As such, a bound on the effective frequency range of the negative capacitance circuit can be determined from f_{PEAK} and $f_{CEILING}$:

$$\frac{1}{2\pi r_o C_{NEG}} < f < \frac{g_m}{2\pi C_{NEG}}\,. \tag{5.10}$$

In this frequency range, the traditional negative capacitance circuit provides capacitance mitigation approximated by C_{NEG}.

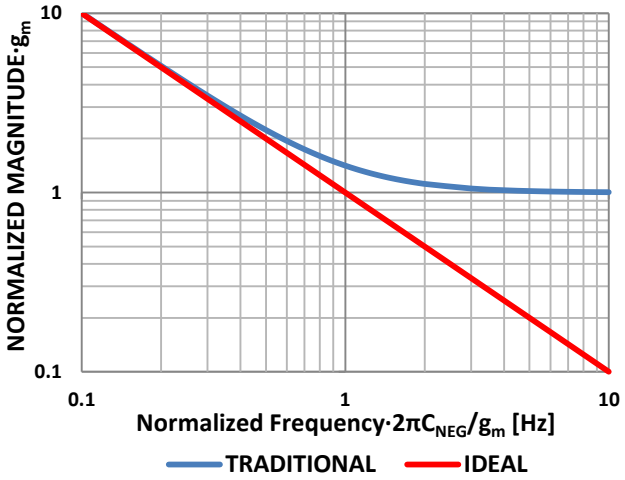


Figure 5.10: Normalized magnitude of the traditional negative capacitance circuit and ideal negative capacitance circuit with magnitude C_{NEG}.

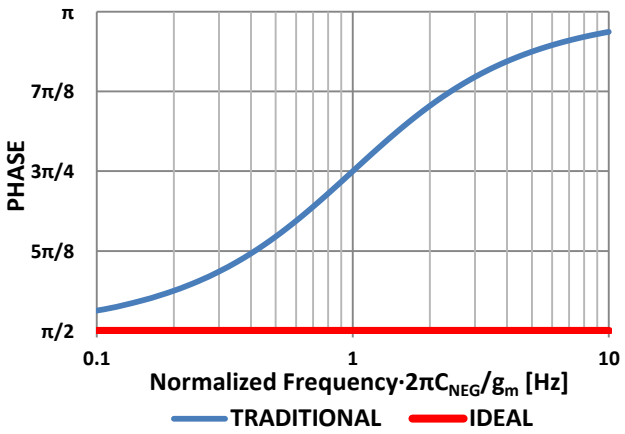


Figure 5.11: Normalized phase of the traditional negative capacitance circuit and ideal negative capacitance circuit with magnitude C_{NEG}.

In general, to extend the receiver bandwidth, the poles of the receiver input must lie well within the effective frequency range of the negative capacitance circuit as defined in (5.10). Attempting to impact poles in the vicinity of $f_{CEILING}$ utilizing negative capacitance circuits will be less effective as the negative capacitance circuit is in the process of transitioning into a negative resistance circuit. To insure the frequency range is appropriate, a suitably large g_m must be realized. To maximize g_m, thin-gate oxide transistors are utilized in conjunction with additional bias current, increasing power consumption. Attempting to increase g_m by increasing

the cross-coupled transistor width may be counter-productive due to the corresponding increase in parasitic capacitance at the receiver input. Even so, achieving a suitable g_m is difficult when, utilizing (5.10), a

$$g_m = 6.28 \frac{f_{CEILING} \cdot C}{1\ \text{GHz} \cdot 1\ \text{pF}} \left[\frac{\text{mA}}{\text{V}}\right] \tag{5.11}$$

is required. For an $f_{CEILING}$ of 10 GHz with 250 fF of negative capacitance, g_m is required to be a substantial 15.7 mA/V.

At DC and low frequencies, $Z_{DIFF\_IN,NC}$ is well approximated as

$$Z_{IN,NC} \approx -\frac{2}{g_m} - 2r_o \ . \tag{5.12}$$

Equation (5.12) indicates that the resistance $2 \cdot R_{TERM}$ required to match the receiver to the differential channel impedance Z_d is strictly less than Z_d and is given by

$$R_{TERM} \approx \frac{1}{2}\left(\frac{1}{Z_d} + \frac{g_m}{2+2g_m r_o}\right)^{-1} . \tag{5.13}$$

The real part of the receiver input impedance is forced to be positive in order to match the channel; consequently, the cross-coupled pair is prevented from oscillating at low frequencies.

5.4.3 G_m-Boosted Negative Capacitance Circuit

In low-power multi-Gb/s links, achieving a g_m large enough to efficiently operate the traditional negative capacitance circuit may be infeasible. This limitation can be partially mitigated by leveraging the inherent gain of the receive amplifier in a g_m-boosted architecture [10]. G_m-boosting is a common circuit technique used to enhance the performance of an analog circuit [28]. A schematic of the g_m-boosted negative capacitance circuit is shown in Figure 5.12, where A is the gain of the amplifier, and all other symbols remain as defined for Figure 5.9.

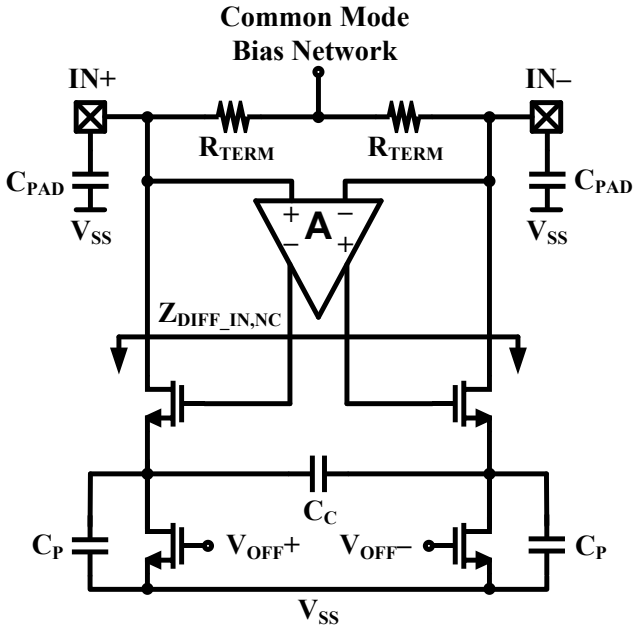


Figure 5.12: G_m-boosted negative capacitance circuit integrated with a basic transmission line termination circuit. Utilized in receiver instantiations 3A and 3B. The feedback amplifier is implemented by utilizing the receiver amplifier.

Using the same assumptions as Section 5.4.2 and further assuming that the amplifier gain A is constant over the frequency band of interest, the differential input impedance looking into the drain terminals of the cross-coupled pair may be written as

$$Z_{DIFF\_IN,NC} = -\frac{2}{Ag_m} - \left(\frac{1}{j\omega AC_C} \parallel \frac{2}{j\omega AC_P} \parallel \frac{2r_o}{A}\right). \tag{5.14}$$

The addition of g_m-boosting does not impact f_{PEAK}, $f_{CEILING}$, or the effective frequency range (5.10) of the negative capacitance circuit. The addition of g_m-boosting does impact the peak reactance,

$$MAX\big(Im\{Z_{IN,NC}\}\big) = -\frac{r_o}{2A}, \tag{5.15}$$

the impedance in the negative capacitance region,

$$Z_{IN,NC} \approx -\frac{2}{Ag_m} - \frac{1}{j\omega A\left(C_C+\frac{C_P}{2}\right)}, \tag{5.16}$$

the single-ended impedance,

$$Z_{IN,NC} \approx -\frac{1}{Ag_m} - \frac{1}{j\omega A(2C_C+C_P)}, \tag{5.17}$$

the effective negative capacitance,

$$C_{NEG} = A(2C_C + C_P), \tag{5.18}$$

and the impedance of the negative capacitance circuit above $f_{CEILING}$,

$$Z_{IN,NR} \approx -\frac{2}{Ag_m}. \quad (5.19)$$

The advantage of the g_m-boosted topology lies in the multiplication of the effective negative capacitance by a factor of A without impacting $f_{CEILING}$. Therefore, to produce the same effective magnitude of negative capacitance as the traditional negative capacitance circuit, the magnitude $C_C + C_P/2$ can be reduced by a factor of A. The reduction increases $f_{CEILING}$ by a factor of A. The increase in $f_{CEILING}$ can be utilized improve the efficiency of the negative capacitance and/or reduce g_m by a factor up to A to save power.

The gain-bandwidth trade-off inherent to amplifier design limits the obtainable g_m-boost. For proper operation, the g_m-boosted circuit requires that the feedback amplifier have a bandwidth higher than that of the desired negative capacitance region specified in (5.10). In high performance CMOS processes, where high bandwidth amplifiers can be realized, the g_m-boosted negative capacitance circuit is expected to outperform the traditional negative capacitance circuit. If the bandwidth is not significantly higher than $f_{CEILING}$, then the impedance in the negative capacitance region will degraded. Examining (5.17), the gain A can be factored from the denominator of all the terms. When the magnitude of A is reduced due to limited amplifier bandwidth, the effective g_m-boost is reduced, diminishing the benefits of the g_m-boosted architecture. Even worse, the phase shift of frequency components in the vicinity of the limited amplifier bandwidth will increase the phase of (5.17). This phase shift adds to the inherent phase shift in the vicinity of $f_{CEILING}$ (see Figure 5.12), hastening the transition from negative capacitance circuit to negative resistance circuit.

5.4.4 Test Chip Negative Capacitance Circuits

Receiver instantiations 2A and 2B were augmented with identical, traditional negative capacitance circuits, and instantiations 3A and 3B were augmented with identical, g_m-boosted negative capacitance circuits. The negative capacitance circuits were designed to be active in a range centered around 5.7 GHz, the estimated bandwidth of receiver 1A. The output of just the first stage of the amplifier is utilized for g_m-boosting, due to the feedback phase constraint. Due to the gain-bandwidth limitation in the first stage, the optimal bandwidth extension occurred for an effective feedback gain, A, of two. In both negative capacitance circuits, the thin-gate oxide cross-coupled transistors were sized to optimize the bandwidth extension and are held constant in

both implementations. As such, the cross-coupled transistors have the same g_m in both circuits. However, to achieve the maximum possible bandwidth of approximately 7.0 GHz, the degeneration capacitance, C_C, was independently optimized for each implementation.

The power and area results for each receiver are listed in Table 5.2. Receivers 2 and 3 dissipate a few additional mW of power compared to receiver 1 in order to maximize the upper bound of the bandwidth extension frequency range, (5.10).

Table 5.2: Measured Power and Area of Test Chip Receiver Input Circuitry Including Two-Stage Amplifier

Receiver	Power [mW]	Area [mm^2]
1	11.0	0.042
2	15.6	0.048
3	15.6	0.047

5.5 Negative Capacitance Circuit ESD Analysis

In addition to reducing the g_m requirement or improving the circuit efficiency, the g_m-boosted architecture offers a significant advantage in terms of its ESD reliability. The traditional negative capacitance circuit contains an ESD hazard; during an IN+ to IN− HBM stress, the entire pin voltage appears across the gate oxide of one of the cross-coupled transistors. Dielectric breakdown in this transistor can be expected to limit the ESD resilience of the part, especially if thin-gate oxides are utilized, as would generally be needed to have suitable high-frequency performance in the cross-coupled pair. In contrast, for the g_m-boosted architecture, the gate of the cross-coupled pair is connected to the output of the receive amplifier. Therefore, the stress voltage applied across the gate oxide of the cross-coupled transistor is only a fraction of the pin voltage, which should result in a higher ESD resiliency.

On the test chip, the g_m-boosted negative capacitance circuit is driven by the amplifier in Figure 5.3. The output of the amplifier, and therefore the gate of the cross-couple transistor, settles to V_{DD} during a HBM event due to the resistor load of the amplifier and the bias voltage of the amplifier being coupled to V_{SS}, which results in a high-impedance current source. Since V_{DD} settles to a voltage one diode drop below the pin voltage due to the primary dual-diode protection scheme as shown in Figure 3.3, the maximum voltage across the gate oxide for the

cross-coupled transistors is reduced by a diode drop as a result of the implementation of the g_m-boosting architecture.

5.6 Receiver Input with Negative Capacitance Root Locus

5.6.1 Receiver Input with Negative Capacitance Model

In Chapter 3, lumped-element models were generated for the signal path consisting of a wireline receiver and its feed, i.e., the transmission line, package, and all on-chip elements up to the input transistor. To generate a lumped-element model of a receiver input with negative capacitance circuit, the lumped-element model for the receiver with primary-only ESD protection circuit is augmented with a circuit model of the negative capacitance circuit, as shown in Figure 5.13. For the receiver model, V^+ is the signal incident at the end of the trace, V_{END} is voltage at the end of the transmission line where it abuts the package, and V_{IN} is the voltage at the input of the receiver. The trace is defined to be transmission line with characteristic impedance Z_O, and is terminated with an on-chip termination resistance R_{TERM}. While an acceptable lumped element model for a package will vary, in this dissertation, the effect of packaging-related series inductance, L, located between the trace and the receiver will be investigated. On-chip, there are several sources of parasitic capacitance at the input of the receiver which combine to generate a shunt capacitance C: the bondpad, $C_{BONDPAD}$, the ESD protection devices, C_{ESD}, and receiver input, C_{RX}. The deviation of the magnitude and phase of the negative capacitance circuit impedance, shown in Figures 5.10 and 5.11, requires the inclusion of both the negative capacitance and negative resistance components of the impedance, (5.5), to be included in the lumped-element model of the circuit. Therefore, the negative capacitance circuit model consists of a negative capacitance, $-C_{NEG}$, and negative resistance, $-1/g_m$, in series between the receiver input, V_{IN}, and AC ground.

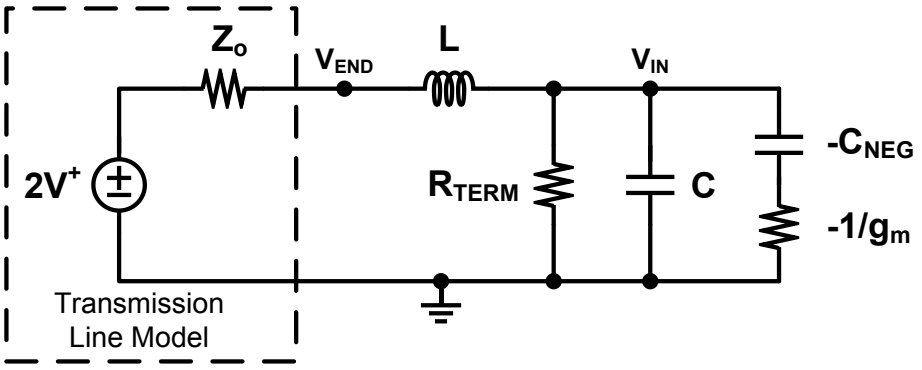


Figure 5.13: Lumped element model for a receiver input with a traditional negative capacitance circuit. The receiver contains a primary-only ESD protected circuit. $C = C_{BONDPAD} + C_{ESD} + C_{RX}$. To generate the lumped-element model for a receiver with g_m-boosted negative capacitance circuit, g_m is multiplied by a factor of A and the proper equation for C_{NEG} is utilized.

To generate the equivalent lumped-element model for the g_m-boosted negative capacitance circuit, the lumped-element model in Figure 5.13 is utilized with g_m multiplied by a factor of A and (5.18) is utilized to determine the magnitude of C_{NEG} instead of (5.6).

5.6.2 Receiver Input with Negative Capacitance Transfer Function

Utilizing the same process as Section 4.3.1, one can derive the transfer function, *H(s)* with form (4.1), for Figure 5.13 in the Laplace domain to relate V_{IN} to V^+. Utilizing [20], the coefficients of the third-order denominator and first-order numerator can be obtained, producing

$$H(s) = \frac{\frac{1}{2}\left(1+\frac{C_{NEG}}{g_m}s\right)}{1+\left[\frac{L}{2Z_o}+\frac{Z_o}{2}(C-C_{NEG})+\frac{C_{NEG}}{g_m}\right]s+\left[\frac{L(C-C_{NEG})}{2}+\frac{C_{NEG}}{g_m}\left(\frac{L}{2Z_o}+\frac{Z_oC}{2}\right)\right]s^2+\frac{LC}{2}\frac{C_{NEG}}{g_m}s^3}. \quad (5.20)$$

The transfer function contains a zero on the real axis at

$$z_{RX} = -\frac{g_m}{C_{NEG}}, \quad (5.21)$$

which equals $f_{CEILING}$, as both represent the frequency at which the magnitude of the negative resistance equals the reactance in (5.5).

To simplify the transfer function, let the first- and second-order denominator coefficients of the receiver without negative capacitance, (4.2), be defined as $\overline{b_1}$ and $\overline{b_2}$, respectively:

$$\overline{b_1} = \frac{1}{2}\left[\frac{L}{Z_o} + Z_oC\right] \quad (5.22)$$

and

$$\overline{b_2} = \frac{LC}{2}. \quad (5.23)$$

Additionally, let the first- and second-order denominator coefficients of a receiver input with ideal negative capacitance be defined as

$$\overline{\overline{b_1}} = \frac{1}{2}\left[\frac{L}{Z_o} + Z_o(C - C_{NEG})\right] \tag{5.24}$$

and

$$\overline{\overline{b_2}} = \frac{L(C-C_{NEG})}{2}. \tag{5.25}$$

Simplifying the transfer function (5.20) utilizing the newly defined coefficients results in

$$H(s) = \frac{\frac{1}{2}\left(1+\frac{C_{NEG}}{g_m}s\right)}{1+\left[\overline{\overline{b_1}}+\frac{C_{NEG}}{g_m}\right]s+\left[\overline{\overline{b_2}}+\overline{b_1}\frac{C_{NEG}}{g_m}\right]s^2+\overline{b_2}\frac{C_{NEG}}{g_m}s^3}. \tag{5.26}$$

To modify the equation for a g_m-boosted circuit, multiply g_m by the gain of the feedback, A, and utilize (5.18) to determine C_{NEG}.

In (5.26), $f_{CEILING}$, which is equal to $g_m/(2\pi C_{NEG})$, can be utilized to indicate the effectiveness of the negative capacitance circuit. If $f_{CEILING}$ is extremely large, then the transfer function will approach the transfer function of a receiver input with an ideal negative capacitance circuit:

$$H(s) \approx \frac{\frac{1}{2}}{1+\overline{\overline{b_1}}s+\overline{\overline{b_2}}s^2}. \tag{5.27}$$

If $f_{CEILING}$, $g_m/(2\pi C_{NEG})$, is relatively small, then the transfer function approaches that of the receiver input without negative capacitance, (4.2):

$$H(s) \approx \frac{\frac{1}{2}}{1+\overline{b_1}s+\overline{b_2}s^2}. \tag{5.28}$$

Therefore as $f_{CEILING}$ is increased, the response of the receiver input in the vicinity of $f_{CEILING}$ transitions from one without a negative capacitance circuit to one with a negative capacitance circuit. Unfortunately, to approximate the transfer function with (5.27) – a receiver input with ideal negative capacitance circuit – $f_{CEILING}$ must dominate both $1/2\pi\overline{\overline{b_1}}$ and $\overline{b_1}/2\pi\overline{\overline{b_2}}$, which is often infeasible. Therefore, the receiver high-frequency response will lie within the complicated transition region, making further analytical analysis difficult.

In the above analysis, the dependence of $\overline{\overline{b_1}}$, $\overline{\overline{b_2}}$, and $f_{CEILING}$ on the magnitude of C_{NEG} should be noted. While increasing C_{NEG} could lead to more capacitance mitigation, the increase reduces $f_{CEILING}$, resulting in a decrease in the negative capacitance circuit effectiveness. Likewise, reducing C_{NEG} improves the efficiency of the negative capacitance circuit, but reduces the mitigation of capacitance.

5.6.2.1 Transfer Function Coefficient Derivations

To derive the coefficients of *H(s)*, the methods derived in [20] are utilized. The circuit in Figure 5.13 has one significant zero since exactly one reactive element can be infinite valued without making the resultant transfer constants, $H^{ijk\ldots}$ zero [20]. As such,

$$a_i = 0, for\ i \geq 2\ . \tag{5.29}$$

Solving for the DC gain,

$$\begin{aligned} a_0 &= \frac{R_{TERM}}{Z_o + R_{TERM}} \\ &= \frac{1}{2} \end{aligned} \tag{5.30}$$

and

$$\begin{aligned} a_1 &= -C_{NEG}((Z_o||R_{TERM}) - \frac{1}{g_m})\frac{1}{2 - g_m Z_o} \\ &= \frac{C_{NEG}}{2g_m}. \end{aligned} \tag{5.31}$$

The circuit of Figure 5.13 has three reactive elements, so

$$b_i = 0, \text{for } i \geq 4. \tag{5.32}$$

Solving for the denominator coefficients,

$$\begin{aligned} b_1 &= \frac{L}{Z_o + R_{TERM}} + (Z_o||R_{TERM})C - (Z_o||R_{TERM} - \frac{1}{g_m})C_{NEG} \\ &= \frac{L}{2Z_o} + \frac{Z_o}{2}(C - C_{NEG}) + \frac{C_{NEG}}{g_m} \end{aligned} \tag{5.33}$$

$$b_2 = \frac{L}{Z_o + R_{TERM}} R_{TERM} C - \frac{L}{Z_o + R_{TERM}} \left(R_{TERM} - \frac{1}{g_m}\right) C_{NEG} + (Z_o||R_{TERM}) \frac{1}{g_m} C_{NEG}$$

$$b_2 = \frac{L(C - C_{NEG})}{2} + \frac{C_{NEG}}{g_m}\left(\frac{L}{2Z_o} + \frac{Z_o C}{2}\right) \tag{5.34}$$

$$b_3 = \frac{L}{Z_o + R_{TERM}} R_{TERM} C(-C_{NEG}) \frac{-1}{g_m}$$

$$b_3 = \frac{LC}{2} \frac{C_{NEG}}{g_m} \tag{5.35}$$

5.6.3 Receiver Input with Negative Capacitance Root Locus

Since further analysis of (5.20) is difficult, most circuit designers would naturally resort to SPICE simulations to determine the effect of C_{NEG} and g_m on the receiver input poles. In this dissertation, the application of control theory, specifically root locus analysis, is utilized to efficiently generate plots demonstrating how the receiver input poles move across the complex plane as a function of C_{NEG} for a given g_m. Traditionally, root locus analysis is utilized to

determine the impact of forward gain on the closed-loop poles of a system with feedback [22]. In this dissertation, an appropriate equation for root locus analysis will be generated for the receiver input with C_{NEG} as the independent variable.

To generate the appropriate equation for root locus analysis, the denominator of (5.20) is set to zero:

$$0 = D(s) = 1 + \left[\frac{L}{2Z_o} + \frac{Z_o}{2}(C - C_{NEG}) + \frac{C_{NEG}}{g_m}\right]s + \left[\frac{L(C-C_{NEG})}{2} + \frac{C_{NEG}}{g_m}\left(\frac{L}{2Z_o} + \frac{Z_oC}{2}\right)\right]s^2 + \frac{LC}{2}\frac{C_{NEG}}{g_m}s^3 \tag{5.36}$$

Rearranging (5.36), with the usage of $\overline{b_1}$ and $\overline{b_2}$, in an attempt to generate an equation appropriate for root locus analysis with C_{NEG} as the independent variable produces

$$1 + C_{NEG}\frac{(\frac{1}{g_m} - \frac{Z_o}{2})s + (\frac{\overline{b_1}}{g_m} - \frac{L}{2})s^2 + \frac{\overline{b_2}}{g_m}s^3}{1 + \overline{b_1}s + \overline{b_2}s^2} = 0\,. \tag{5.37}$$

Since (5.37) is improperly constructed for root locus analysis, the equivalent system with $1/C_{NEG}$ as the independent variable,

$$1 + \frac{1}{C_{NEG}}\frac{1 + \overline{b_1}s + \overline{b_2}s^2}{(\frac{1}{g_m} - \frac{Z_o}{2})s + (\frac{\overline{b_1}}{g_m} - \frac{L}{2})s^2 + \frac{\overline{b_2}}{g_m}s^3} = 0\,, \tag{5.38}$$

is proper, allowing for standard root locus analysis [22]. For the remainder of the chapter, the plots and analysis will be transformed such that C_{NEG} is the parameter which changes along the root locus branches.

The addition of negative capacitance circuits has the ability to move poles into the right-half plane, which should be avoided to insure unconditional stability at the receiver input. While not a conclusive test for instability, a transfer function is guaranteed to contain a pole in the right-half plane if a denominator coefficient is negative [22]. In either root locus form, (5.37) or (5.38), the numerator and denominator correspond to the denominator of the transfer function if the independent variable is set to infinity or zero respectively. For a C_{NEG} of infinity, g_m fulfilling

$$g_m > \frac{2}{Z_o} \tag{5.39}$$

or

$$g_m > (\frac{1}{Z_o} + \frac{Z_oC}{\mathrm{L}}) \tag{5.40}$$

produce a denominator coefficient which is negative, placing a pole in the right-half plane. As such, the possibility for right-half plane poles is a design consideration, especially for large values of g_m or C_{NEG}.

Equation (5.39) is not surprising, considering how the negative capacitance circuit can operate as a negative resistance circuit. At high frequencies, the negative resistance is effectively in parallel with the termination resistance. If the magnitude of the negative resistance is smaller than the parallel combination of the characteristic impedance and termination resistance, then the input impedance of the receiver may have a negative real component, leading to instability. In high-speed links, the characteristic impedance and termination impedance are typically small, preventing instability resulting from the addition of a negative capacitance circuit for a reasonable range of g_m.

5.6.4 Impact of C_{NEG} on Receiver Input Poles

To investigate the impact C_{NEG} on the receiver input poles, the root locus plot shown in Figure 5.14 was generated for a 50 Ω characteristic impedance system with L = 0.5 nH, g_m = 25 mA/V, and two magnitudes of parasitic capacitance: C = 200 fF and C = 400 fF. To generate the root locus data, a single MATLAB command, rlocus, was executed utilizing (5.38). In the root locus plots, only the two complex conjugate poles are shown; the third pole is analyzed later. The plot includes the root locus branches for ideal capacitance mitigation, C_{NEG} = 0 fF with C varied along the branch, for reference. By observation, as the magnitude of C_{NEG} increases, the input poles of the receiver with negative capacitance deviate from the pole locations for ideal capacitance mitigation, curling toward the imaginary axis. Therefore, the negative resistance component is significantly changing the impact of the negative capacitance circuit and is critical to the analysis and optimization of negative capacitance circuits.

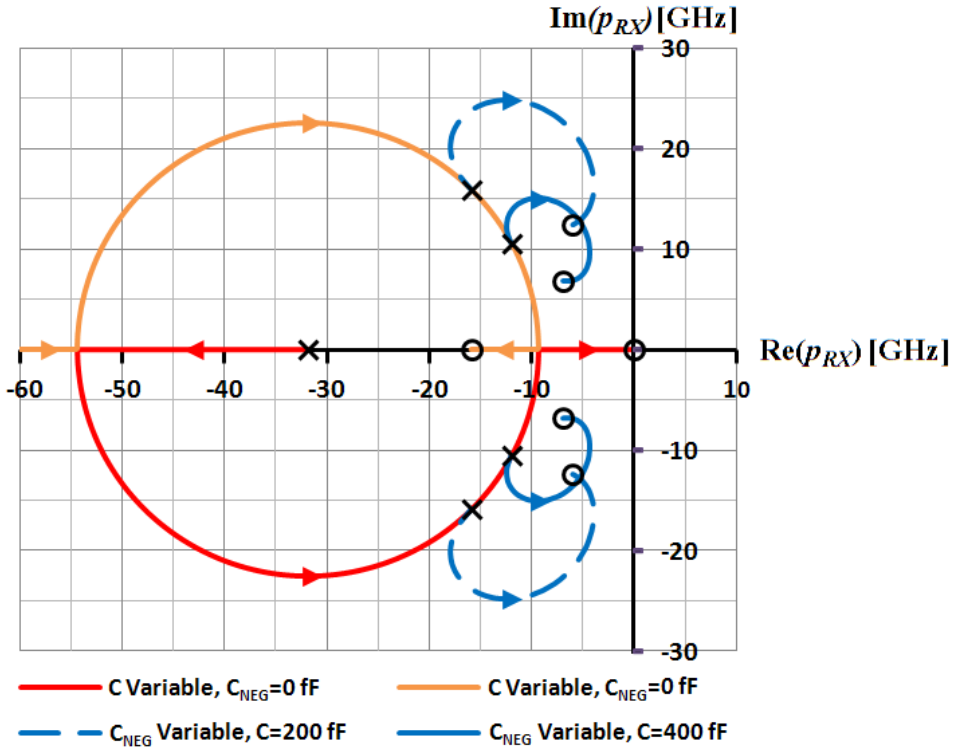


Figure 5.14: Root locus plot of the complex poles of the receiver input with negative capacitance circuit for $L = 0.5$ nH, $g_m = 25$ mA/V, and $Z_o = R_{TERM} = 50$ Ω. Along the root locus branches with $C_{NEG} = 0$, C varies from 0 to ∞. Along the root locus with $C = 200$ fF or $C = 400$ fF, C_{NEG} varies from 0 to ∞.

For relatively small magnitudes of C_{NEG}, increasing C_{NEG} increases the natural frequency, f_n, the distance from the pole to the origin [22]. An increase in f_n results in an extension of the −3 dB bandwidth for a given damping ratio, ζ. ζ is equal to $\sin(\theta)$, where θ is the angle of the pole with respect to the imaginary axis [22]. As C_{NEG} is increased further, the f_n of the receiver poles eventually achieves a maximum. As C_{NEG} is increased above this magnitude, f_n decreases.

In addition to the change in f_n, increasing C_{NEG} changes ζ. The root locus branches of Figure 5.14 demonstrate that receivers with negative capacitance circuits can achieve smaller ζ than circuits with ideal capacitance mitigation. A reduction in ζ increases the f_{-3dB}/f_n ratio, resulting in an extension of the bandwidth for a given f_n. A combination of increased f_n and f_{-3dB}/f_n can result in a −3 dB bandwidth which exceeds the maximum bandwidth extension using ideal capacitance mitigation.

The optimal receiver input pole locations may not be the pole locations with the maximum −3 dB bandwidth. As ζ is reduced below 0.7, the transient response to the complex poles at the receiver input will show an increase in overshoot and ringing (5% overshoot for an ideal step

response when $\zeta = 0.7$), while demonstrating peaking in the frequency response [22]. To cap the undesirable overshoot and ringing, a minimum ζ for the receiver input poles can be set. In general, to optimize the negative capacitance circuit, the relationship between ζ and f_n along the root locus branches can be analyzed to optimize the signal integrity. Utilizing MATLAB for system level simulations is suggested, allowing the MATLAB generated root locus pole locations to be utilized during the optimization of the link.

The receiver input with a negative capacitance circuit, i.e. the circuit of Fig. 5.13, contains a non-complex conjugate (third) pole and a zero; their locations on the real axis are shown in Figure 5.15 for a 50 Ω characteristic impedance system with $L = 0.5$ nH, $g_m = 25$ mA/V, and two values of parasitic capacitance: $C = 200$ fF and $C = 400$ fF. The plot demonstrates how the pole and zero are within close proximity of each other for a wide range of C_{NEG}, minimizing their effect on the receiver input response. Therefore, the inclusion of third pole and zero is optional, but may be included to increase the simulation accuracy for small values of C_{NEG} since the response at the receiver input will be dominated by the complex poles. In general, C_{NEG} values large enough that the third pole and zero dominate the response at the receiver input will not improve the signal integrity at the receiver input.

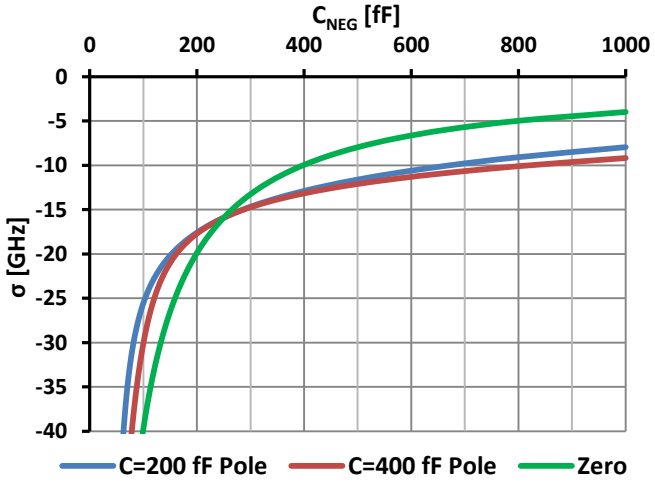


Figure 5.15: Non-complex conjugate (third) pole and zero locations on the real axis as a function of C_{NEG} for $L = 0.5$ nH, $Z_o = R_{TERM} = 50$ Ω, and $C = 200$ fF or $C = 400$ fF. The zero location is independent of C, L, Z_o, and R_{TERM}.

5.6.5 Impact of Parasitic Inductance on Receiver Input Poles

To investigate the impact of L on the receiver input poles, the root locus plot shown in Figure 5.16 was generated for a 50 Ω characteristic impedance system with $L = 2$ nH, $g_m = 25$ mA/V, and two magnitudes of parasitic capacitance: $C = 400$ fF and $C = 800$ fF. The receiver input poles curl away from the pole locations for ideal capacitance mitigation for increasing C_{NEG} in the same manner as in Figure 5.14 with $L = 0.5$ nH. The significant differences between Figures 5.14 and 5.16 are a dramatic change in the frequency range and a change in the initial location of the root locus branch with respect to the ideal capacitance mitigation root locus for a given parasitic capacitance, C.

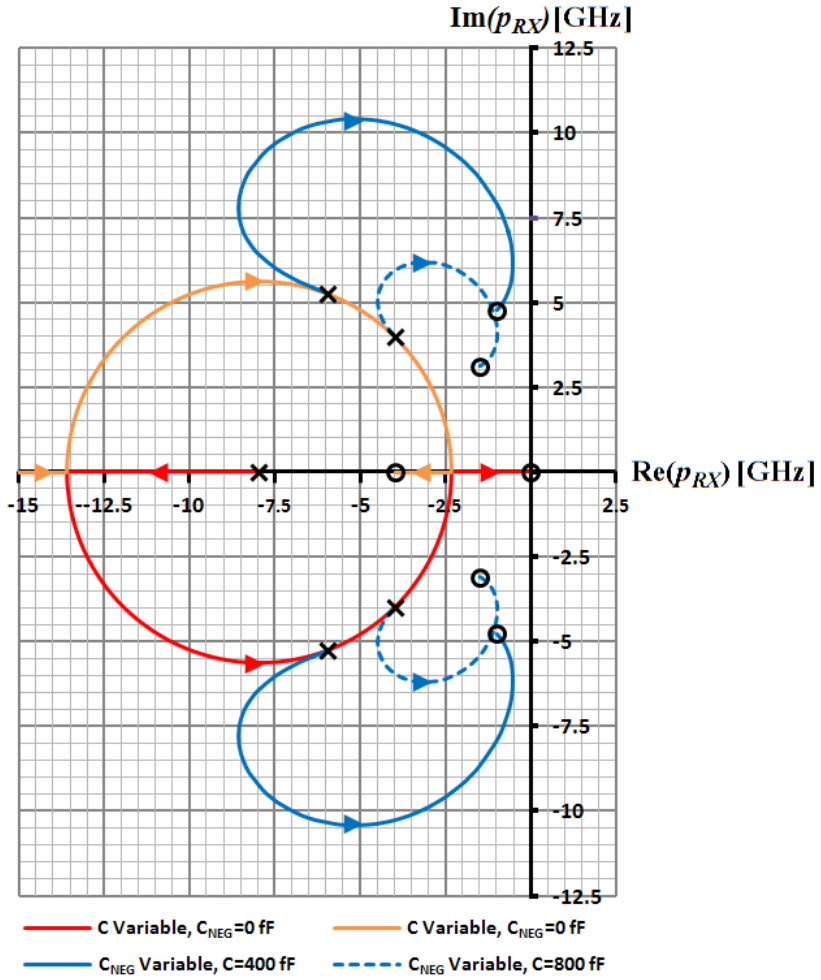


Figure 5.16: Root locus plot of the complex poles of the receiver input with negative capacitance circuit for $L = 2$ nH, $g_m = 25$ mA/V, and $Z_o = R_{TERM} = 50$ Ω. Along the root locus branches with $C_{NEG} = 0$ fF, C varies from 0 to ∞. Along the root locus branches with $C = 400$ fF or $C = 800$ fF, C_{NEG} varies from 0 to ∞.

An interesting observation can be made by comparing the root locus branches in Figures 5.14 and 5.16. The L = 0.5 nH with C = 200 fF and L = 2 nH with C = 800 fF root locus branches for variable C_{NEG} are identical when the succeeding branch is scaled larger by a factor of 4. In fact, while generating a scalable root locus plot for each variable is not possible, it is possible to scale the root locus if the parasitic inductance and capacitance are scaled by the same factor. To demonstrate this property, one can simply multiply the variables L, C, and C_{NEG} while dividing s of the transfer function (5.20) by the same factor to no overall effect. The scaling of C_{NEG} can then be ignored as it is the independent variable along the root locus branches.

The results of Figures 5.14 and 5.16, along with the scaling property, indicate that the response of the receiver input poles to a negative capacitance circuit is dependent on the ratio of parasitic capacitance to parasitic inductance. For the ratios shown, the achievable bandwidth extension is greater for smaller magnitudes of parasitic capacitance. The achievable bandwidth extension correlates to the sensitivity of the bandwidth to relatively low values of capacitance mitigation utilizing Figure 4.3 or the bandwidth extension method presented in Chapter 3. Therefore, negative capacitance circuits are most effective when the receiver input poles are sensitive to small magnitudes of capacitance mitigation.

5.6.6 Impact of g_m on Receiver Input Poles

To investigate the impact of g_m of the negative capacitance circuit on the receiver input poles, the root locus plot shown in Figure 5.17 was generated for a 50 Ω characteristic impedance system with L = 0.5 nH, C = 200 fF, and four magnitudes of g_m: 5 mA/V, 15 mA/V, 25 mA/V, and 35 mA/V. The results show how increasing g_m improves the efficiency of the negative capacitance circuit for low magnitudes of C_{NEG}, reducing the deviation of the root locus branch from the pole locations with ideal capacitance mitigation. The most significant advantage of increasing g_m is increasing the achievable f_n for a fixed ζ. As such, increasing g_m can increase the achievable −3 dB bandwidth of the receiver input while maintaining an acceptable level of overshoot and ringing.

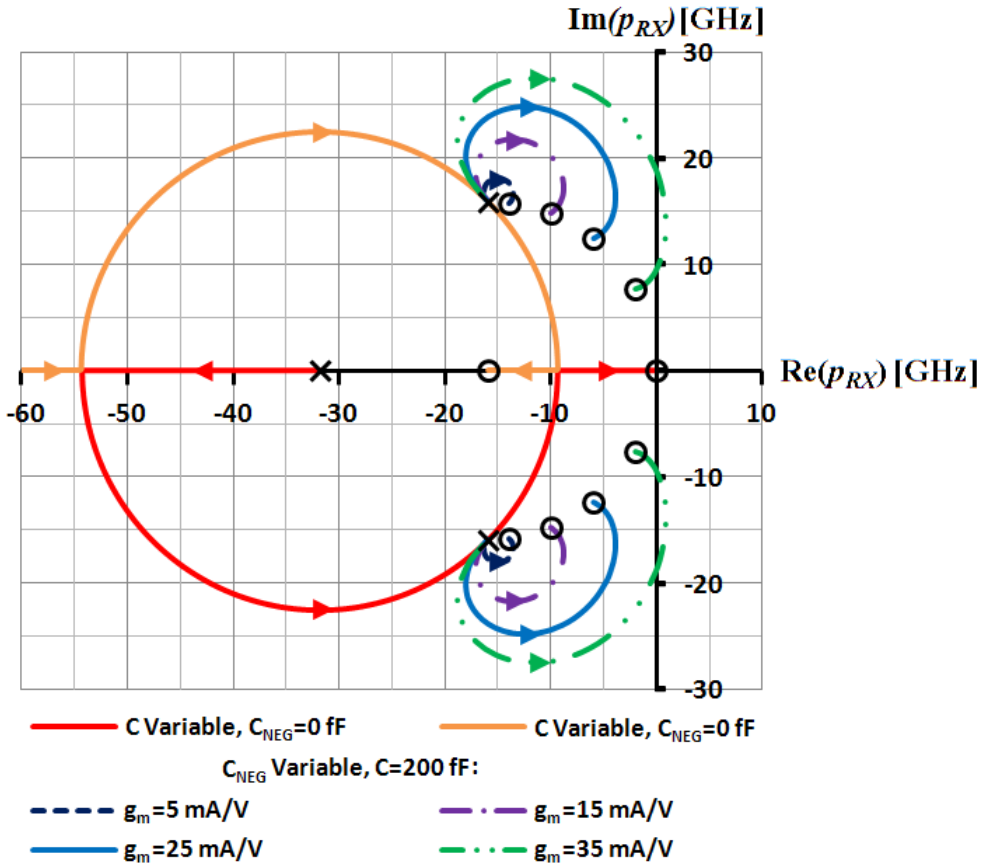


Figure 5.17: Root locus plot of the complex poles of the receiver input with negative capacitance circuit for L = 0.5 nH, $Z_o = R_{TERM}$ = 50 Ω, C = 200 fF, and g_m magnitudes of 5, 15, 25, and 35 mA/V.

The root locus branch in Figure 5.17 for g_m = 35 mA/V demonstrates how the addition of a negative capacitance circuit with large g_m and C_{NEG} can produce right-half plane poles. Since the equations (5.39) and (5.40) are not required for right-half plane poles, the root locus analysis method is useful in determining if a range of C_{NEG} will produce right-half plane poles for a given g_m.

5.6.7 G_m-Boosted Negative Capacitance Root Locus

The g_m-boosted circuit's advantage lies with the effective multiplication of g_m by the amplification factor A. To modify the root locus method presented for the g_m-boosted circuit, both g_m and C_{NEG} are multiplied by A. Since the C_{NEG} only determines the location of the poles along the root locus branches, the multiplication of C_{NEG} does not change the shape of the branches. On the other hand, a multiplication of g_m has the exact same effect as increasing g_m in

Section 5.6.6 without the power overhead: it increases the efficiency of the negative capacitance circuit, increasing the achievable −3 dB bandwidth.

If the receive amplifier is not sufficient for utilization in the feedback of the g_m-boosted architecture, the power overhead to modify the receive amplifier (or implement a stand-alone amplifier) can be compared to the power overhead to increase g_m in the traditional architecture. In many scenarios, even the g_m-boosted architecture with a stand-alone amplifier is expected to achieve power savings over the traditional negative capacitance architecture.

5.7 Bandwidth Extension Circuit Measurement Results

5.7.1 Signal Integrity Results

The 10 Gb/s single-ended eye diagrams for receivers 1A, 2A, and 3A are shown in Figure 5.18. It is evident that receivers 2A and 3A, with the bandwidth extension circuits, produce a more open eye and a substantial reduction in jitter. The quantitative performance metrics are listed in Table 5.3.

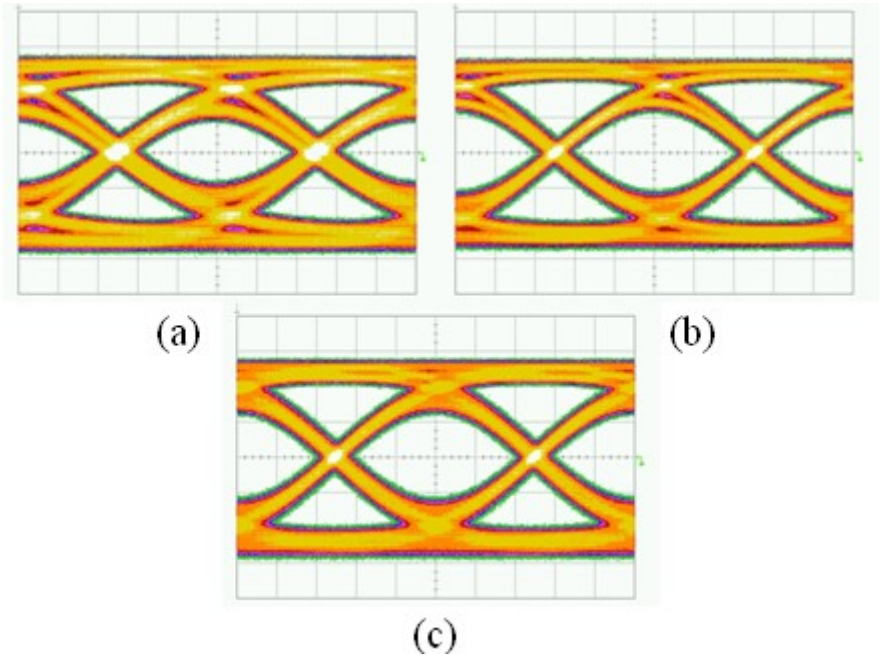


Figure 5.18: 10 Gb/s PRBS single-ended eye diagrams for (a) receiver 1A, (b) receiver 2A, and (c) receiver 3A. Differential input amplitude is 125 mV_{pp}. [Horizontal scale: 20 ps/div.; vertical scale: 20 mV/div.].

Table 5.3: 10 Gbps PRBS Single-ended Eye Diagram Measurement Results

	Receiver 1A	Receiver 2A	Receiver 3A
Eye Amplitude [mV_{pp}]	98.7	97.9	100.5
Eye Height [mV_{pp}]	32.7	42.7	45.9
Jitter [ps_{pp}]	31.6	24.2	23.6
	Receiver 1B	Receiver 2B	Receiver 3B
Eye Amplitude [mV_{pp}]	93.1	94.7	93.9
Eye Height [mV_{pp}]	16.5	24.5	24.1
Jitter [ps_{pp}]	44.2	31.8	33.8

The measurement results demonstrate that the negative capacitance circuits are also effective on receivers which utilize series CDM resistors. The addition of secondary ESD protection to the receiver input reduced the eye height by 50% and increased the jitter by 40%. The addition of negative capacitance circuits to the receiver with secondary protection reduced the jitter by approximately 25% and increased the eye height by approximately 50%. The addition of negative capacitance circuits to the receiver input effectively negated the jitter penalty and partially mitigated the eye height penalty of secondary ESD protection.

The differential TDT results for receivers 1A, 2A, and 3A are shown in Figure 5.19. The measurements indicate that both bandwidth extension circuits improve the −3 dB bandwidth, confirming that the input poles limit the circuit bandwidth. The variation in amplitude seen in the measurement results changes from chip to chip, and is likely due to variations in the termination resistors and the receive amplifiers. Both negative capacitance circuits exhibit a small amount of ringing. In order to maximize the bandwidth extension, the frequency responses of the negative capacitance circuits were allowed to exhibit a maximum of 1 dB of peaking at high frequencies. The risetime and normalized bandwidth are listed in Table 5.4. The results are in good agreement with the calculations. It had been estimated the original bandwidth of 5.7 GHz (receiver 1A) could be extended to 7.0 GHz, an extension factor of 1.23; the data of Table 5.4 indicate that the bandwidth of receivers 2A and 3A was indeed about 1.23 times larger.

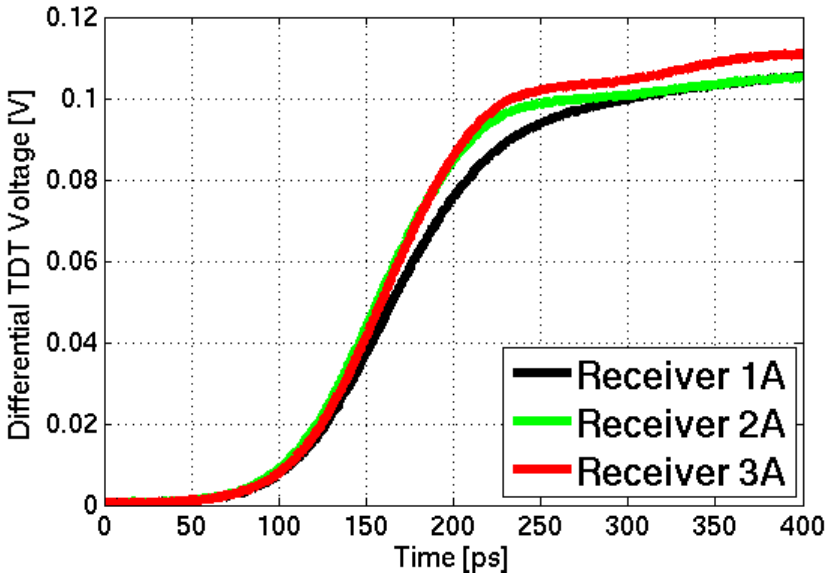


Figure 5.19: Differential TDT measurement results of the three receivers 1A, 2A, and 3A, for a differential step input amplitude of 65 mV.

Table 5.4: TDT Measurement Results

Receiver	20% to 80% Risetime [ps]	Normalized Bandwidth
1A	94.7	1.00
2A	76.7	1.24
3A	78.6	1.21

The signal integrity measurement results indicate that the difference between the two bandwidth extension circuits is minimal in the process technology used, and neither is preferable over the other on the basis of power or area (see Table 5.2). If the package inductance was reduced by shortening the bondwires, the achievable bandwidth extension would be increased for both negative capacitance circuits.

5.7.2 ESD Resiliency Results

IN+ to IN− HBM testing was performed on three samples of receivers 2A and 3A. Receiver 2A, with the traditional negative capacitance circuit, displayed a significant degradation in risetime (obtained from TDT) and output amplitude for HBM stresses above 5.5 kV. Receiver 3A, with the g_m-boosted negative capacitance circuit, functioned without degradation until HBM stresses above 7 kV were applied, at which point the receiver failed to produce an output.

Post-stress electrical analysis showed a drop in the differential input resistance of receiver 2A, corresponding to the drop in eye amplitude, while the input impedance of receiver 3A remained unchanged. These results strongly indicate that input amplifier was damaged in receiver 3A and that gate oxide failure occurred in a cross-coupled transistor of the negative capacitance circuit of receiver 2A. The results support the improved ESD resiliency claim for the g_m-boosted negative capacitance circuit.

All of the receiver 2A, 2B, 3A, and 3B samples passed CDM testing on the input pins of the receiver, up to the maximum precharge voltage of the tester (>1500 V), i.e., no performance degradation occurred. As was stated for receivers 1A and 1B, the resiliency is attributed to the limited charge storage of the package.

The ESD resilience results for all three receivers are summarized in Table 5.5.

Table 5.5: Receiver Input ESD Resilience

	Receiver 1	Receiver 2	Receiver 3
Input HBM Resilience [kV]	4	5.5	7
Input CDM Resilience [V]	>1500	>1500	>1500

5.8 Negative Capacitance Circuit Design Considerations

In this chapter, the traditional and g_m-boosted negative capacitance circuits were implemented to enhance the bandwidth of the input port of the receiver. This dissertation provides the first implementation and measurement results for the g_m-boosted negative capacitance circuit, enhancing the knowledge of the new architecture beyond the work presented in [10]. In this section, the main design considerations for negative capacitance circuit architectures are presented.

5.8.1 Utilizing Negative Capacitance Circuits

In measurement results of Section 5.7.1, it was shown that the negative capacitance circuits can be utilized to enhance the signal integrity at the input of the receiver. Since the circuits have been shown to improve signal integrity, the decision on whether to implement a negative capacitance circuit should be based on the following considerations.

Consideration 1: The link architecture must match the architecture of the negative capacitance circuit. For example, the link architecture must utilize differential signaling for effective negative capacitance circuit operation.

Consideration 2: The system level link analysis must show that the input bandwidth of the receiver; which can be estimated by the method developed in Chapter 3, limits the performance of the link. If the link can operate reliably without the implementation of a negative capacitance circuit, then implementation of a negative capacitance circuit is a waste of area, power, and design effort.

Consideration 3: The system level link analysis must show that implementation of a negative capacitance circuit is optimal, whether that is for reliability, area, power, design complexity or ESD protection metrics. This consideration is the most difficult, due to the large number of design trade-offs. The maximum bandwidth extension from negative capacitance circuits can be estimated from the methods developed in Chapters 3 and 4. The bandwidth extension can then be utilized within the system-level link analysis to obtain the resulting improvement in BER. If the power and area costs of negative capacitance circuits are greater than the power required to improve BER through other methods, such as equalization [7], then the optimal method should implemented. Other metrics, such as design effort and ESD resiliency, can be factored into the decision based on the individual application.

5.8.2 Utilizing the G_m-Boosted Negative Capacitance Circuit

If the considerations of Section 5.8.1 demonstrate that a negative capacitance circuit should be implemented, then the designer must choose between the traditional negative capacitance circuit and the g_m-boosted negative capacitance circuit. The decision on whether to implement the g_m-boosted negative capacitance circuit should be based on the following considerations.

Consideration 1: The bandwidth enhancement through the utilization of the g_m-boosted architecture can be determined through the root locus analysis presented in Section 5.6. The g_m-boosted architecture allows for an increase in the effective g_m without requiring an increase in

the cross-coupled transistor width or bias current. In an optimized circuit, the increase in effective g_m allows for a larger f_n for a given ζ, increasing the achievable −3 dB bandwidth extension.

Consideration 2: The ESD hazard inherent to the traditional negative capacitance circuit, as discussed in Section 5.5 and measured in Section 5.7, which may limit the ESD resiliency.

Consideration 3: If the receiver does not utilize an amplifier, then an independent feedback amplifier must be implemented for the g_m-boosted architecture. The power requirement due to the addition of an independent amplifier must then be included in the system-level link optimization.

If the receive amplifier does not have an appropriate architecture for use in the g_m-boosted architecture, then an independent feedback amplifier must also be implemented for the g_m-boosted architecture. For example, if the first stage of a receive amplifier includes a differential-to-single-ended conversion, then the amplifier cannot be utilized in the g_m-boosted architecture.

Consideration 4: If the design of the receive amplifier is significantly impacted by its usage in the g_m-boosted negative capacitance architecture and the redesign requires a significant increase in power, area, or design effort compared to the traditional negative capacitance architecture, then utilizing the g_m-boosted architecture is not advantageous. As discussed in Section 5.4.3, the g_m-boosted circuit requires that the feedback amplifier have a significantly higher bandwidth than that of the input. If the bandwidth is too low, then the effectiveness of the negative capacitance circuit can be limited by the feedback amplifier and the effects determined through simulation. An additional factor is the increased loading on the receive amplifier due to the cross-coupled transistors. Once again, an independent amplifier may be utilized to remove the additional burden on the receive amplifier.

In high-performance CMOS processes, the g_m-boosted negative capacitance circuit is expected to have an advantage over the traditional negative capacitance circuit. In lower performance (e.g., low-power) processes, achieving a sufficiently large, single-stage amplifier bandwidth might result in a gain factor not much larger than 1, and the g_m-boosted circuit may

not provide an improvement over the traditional negative capacitance circuit. In high-performance processes, a satisfactory gain factor is achievable, providing enhanced performance, as discussed in Section 5.6.

CHAPTER 6

SINGLE-CORE MULTIPLE-OUTPUT PASSIVE DAC

6.1 Introduction

In this chapter, a variable-reference ADC-based receiver test chip is designed. A block diagram of a transceiver utilizing an ADC for receive sampling is shown in Figure 6.1. To optimize the BER while lowering the power consumption of an ADC-based receiver, the requirement of uniform reference levels can be relaxed. With the reference levels set to BER-optimal thresholds, the number of required thresholds to achieve a reliable link can be reduced from 15 (4-bit) in a uniform-reference ADC to 7 (3-bit) in a non-uniform ADC [14]. The reduction in thresholds results in a significant power savings, especially when the ADC is implemented using a flash architecture [11]. To implement a variable-reference ADC, the standard core of the ADC is maintained, while the standard uniform-reference generating DAC is replaced with a variable-reference generating DAC.

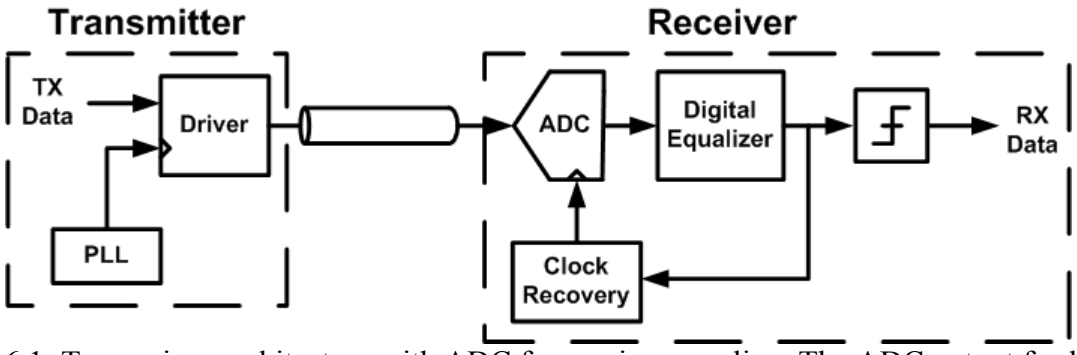


Figure 6.1: Transceiver architecture with ADC for receive sampling. The ADC output feeds a digital equalizer utilized to equalize the channel impairments.

6.2 Test Chip and Test Board

6.2.1 Test Chip and Test Board Overview

A 2 mm x 2 mm ADC receiver test chip was designed and manufactured in a low-power 90 nm CMOS technology. A layout die microphotograph is shown in Figure 6.2. The die is packaged within a 6 mm x 6 mm open-cavity QFN. The package is attached to a custom high-speed test board using a high-bandwidth elastomer socket for signal integrity measurements. The socket will be attached to a four-layer test board manufactured using Rodgers 4003 dielectric. A photograph of the test board is shown in Figure 6.3.

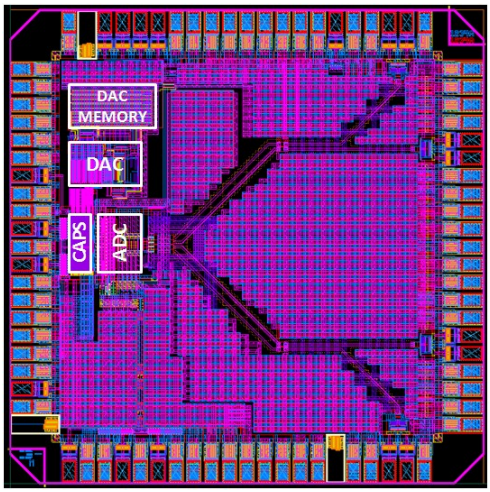


Figure 6.2: Test chip micrograph with ADC core, DAC, DAC memory, and reference voltage storage capacitors (CAPS) highlighted.

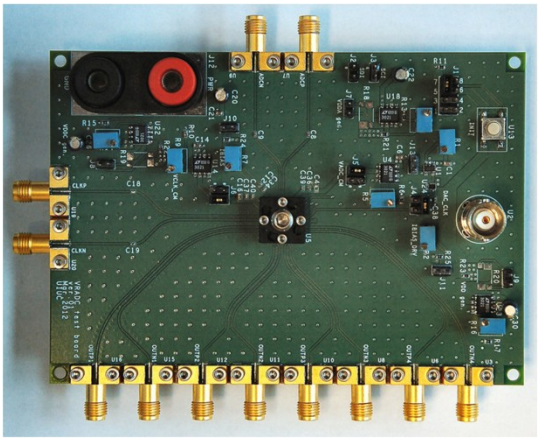


Figure 6.3: Test board photograph. A high-bandwidth elastomer socket (center) is used to attach the QFN packaged test chip to the test board.

6.2.2 Test Chip Architecture and Circuits

The test chip architecture is shown in Figure 6.4 and consists of a single variable-reference ADC-based receiver. The test chip has three power domains: analog, digital, and clock. The analog power domain contains all circuits which do not utilize a clock. The clock power domain contains all circuitry utilized to buffer the clock throughout the test chip. The digital power domain contains all the remaining circuitry.

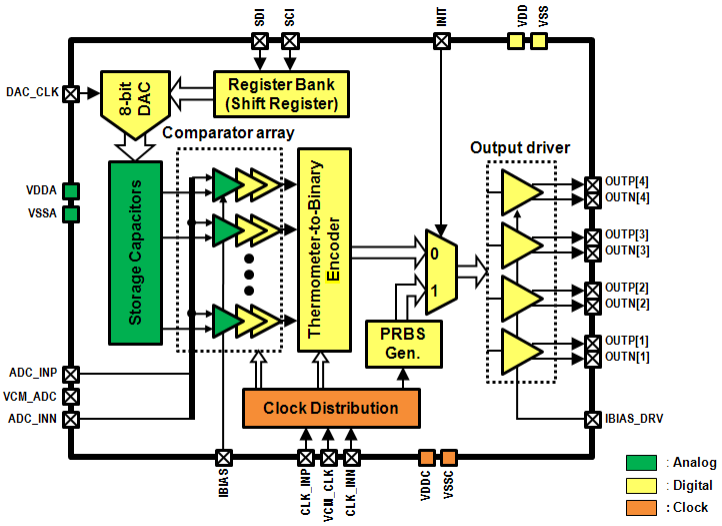


Figure 6.4: Test chip architecture with power supply domains highlighted.

The differential receiver input pins, ADC_INP and ADC_INN, are terminated utilizing 50 Ω resistors to match the channel impedance. The other terminal of the termination resistors is connected to the input VCM_ADC which is utilized to set the input common mode via a test board voltage regulator. The input is used as input to a 4-bit flash ADC core consisting of a comparator array and thermometer-to-binary encoder. The first stage of each comparator array path is a pre-amplifier, shown in Figure 6.5. The second and third stages of the comparator array are digital latches utilized to reduce the metastability region and generate a full-swing digital output. The thermometer-to-binary encoder encompasses a three-input NAND bubble error correction, thermometer-to-gray encoder, and a gray-to-binary encoder. The output of the encoder is buffered across the chip to high-speed 50 Ω CML drivers for transmitting the data off-chip.

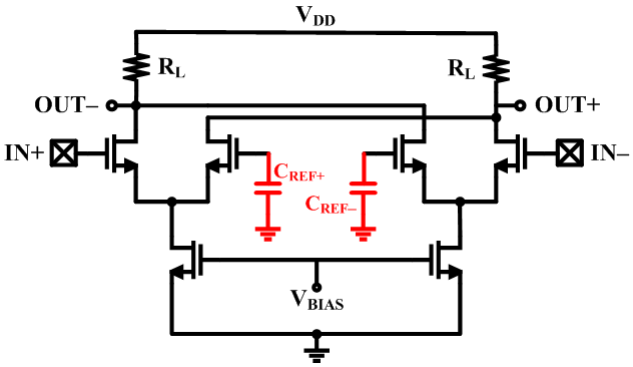


Figure 6.5: Pre-amplifier utilized as the first stage of the comparator array. The reference voltage storage capacitors are highlighted in red.

The system-level investigation showed that a 3-bit variable-reference ADC will have BER performance equal to or better than a 4-bit uniform ADC. To operate an ADC in either a 4-bit uniform-reference mode or a 3-bit variable reference mode, requires that the DAC which generates the voltage reference must be newly designed for flexibility. In this work, a 4-bit DAC which can generate both uniform and variable-references replaces the standard 4-bit uniform DAC.

To synchronize the test chip output with the FPGA, which accepts the ADC data driven off-chip, an initialization sequence is generated utilizing an on-chip PRBS generator. The initialization sequence is controlled by the user via push-button on the test board.

A user programmable shift register with 272 bits is utilized to set static test chip control signals and the DAC reference voltage codes. The shift register is controlled via clock, SCI, and data, SDI, pins. The shift register is designed utilizing positive edge-triggered D flip-flops which are risetime sensitive, so a series of inverters is utilized to improve the risetime of the clock signal.

6.3 DAC Design Requirements

To generate a variable-reference ADC, the DAC must be modified to generate the desired reference voltages. Since the DAC is utilized in a high-speed receive architecture, several design requirements must be fulfilled. First, the reference voltages must be valid at all times since the receiver is sampling the received signal at 4 GHz using a flash ADC core. Therefore the DAC architecture must not have an update phase, which results in temporarily invalid outputs. Second,

the system-level design requires the ADC be able to produce 3-bit uniform, 4-bit uniform, 3-bit non-uniform, and 4-bit non-uniform configurations. Therefore, the DAC is required to be able to generate 3- and 4- bit uniform and non-uniform reference levels. To meet this specification, the DAC is required to generate 30 non-uniform voltages for the 4-bit ADC comparators, including the set of reference levels for a 4-bit uniform DAC. Third, the minimum output voltage resolution is set based on a system-level BER simulation of a fixed, but non-uniform, reference level ADC [14], which generated a DAC precision requirement of 5-bits. The ADC is designed to receive signals with an input swing of 800 mV, so the single-ended range of the DAC was required to be 400 mV around a common-mode of 800 mV. A 5-bit DAC with a 400 mV single-ended range results in a least significant bit (LSB) step size of 12.5 mV. In order to achieve this requirement with an additional 3-bits of calibration (to cancel DAC and comparator variation) and without significantly increasing the design complexity, an 8-bit single-ended DAC was utilized. With 8-bit resolution, a 1.6 mV LSB step size is produced. During the design phase, circuit noise sources on the order of 1 LSB, 1.6 mV, were allowed on the reference voltage, as the LSB step size is an order of magnitude smaller than the design requirement.

The major design goal for the DAC was to maintain a minimum power overhead when compared to a 4-bit uniform-reference DAC. If the modified DAC requires a significant increase in power consumption, then the increase in power consumption would offset the power saved by reducing the number of thresholds in the ADC by utilizing a variable-reference architecture. The strict power requirement limited the number of viable DAC architectures. For example, an 8-bit switched capacitor DAC utilizes 2^8-1unit current sources. To implement such a DAC in a power budget of approximately 1 mW, the unit current sources in a 1.2 volt domain would be required to be 3.25 μA. A current mirror of 3.25 μA will result in a low current density, which is difficult to implement due to process variations. Advanced current source designs, such as cascode current sources, improve the sensitivity to variation but are troublesome to design in low-supply voltage applications.

A secondary DAC design goal was to implement a design that was reasonably area-efficient. The increase in DAC resolution, along with 30 outputs, requires additional area overhead when compared to the uniform-reference generation; so the architecture should minimize the area penalty of the variable-reference ADC. For example, implementing 30 single-ended 8-bit DACs,

each with 2^8-1 unit elements, to generate all of the references would require 7650 unit elements, requiring significant area consumption.

6.4 DAC Design Concept

6.4.1 Capacitor Storage of Voltage References

To generate the 30 reference voltages, a single-core multiple-output DAC architecture was chosen to reduce the power and area penalty of multiple DAC cores. To generate a DAC which can generate 30 non-related, always valid reference voltages, it was determined that the voltage references could be stored on an array of capacitors. In the design, two potential issues arise when utilizing capacitors to store the voltage reference: voltage droop due to leakage and noise.

To minimize leakage, thick-gate oxide, accumulation-mode NMOS in N-well devices were used to implement the reference voltage storage capacitors. The gate of the transistors is connected to the reference voltage while the N-well, source, and drain are connected to the analog supply ground (VSSA). Since the reference voltage is being held on the gate of a CMOS comparator, the only DC current is leakage current through the voltage storage capacitor and the gate of the CMOS comparator (see Figure 6.5). Utilizing data from the design manual, the CMOS comparator gate leakage is estimated to be 63 pA. The leakage of the thick-gate oxide NMOS is not given in the design kit documentation as it is considered negligible. Utilizing

$$\frac{\partial v}{\partial t} = \frac{i}{C} \tag{6.1}$$

to estimate the change in voltage as a function of time, the voltage changes at a rate of 63 volts per second for a capacitance of 1 pF. As such, the capacitor voltage refresh rate, for a 1 pF capacitor, is required to be 63 kHz to reduce the voltage droop due to leakage below 1 mV. To reduce the droop further, and to account for the small leakage of the NMOS in N-well devices, a higher refresh rate is utilized.

In high-fidelity ADC design, utilizing capacitors to store voltages is not common, as the high impedance at low frequencies can result in performance limiting variation on the reference voltages. For ADCs utilized in high-speed I/O receivers, noise on the order of a mV is acceptable, permitting the utilization of capacitors to store the references. Since the body, source, and drain of the reference voltage storage capacitor is referenced to ground, supply noise

rejection will be excellent given an acceptable level of decoupling capacitance. Therefore, the major noise source on the reference will be noise capacitively coupled from the pre-amplifier onto the reference through the capacitances of the transconductance transistors. To minimize the impact, the size of the capacitors can be increased until the noise is reduced to an acceptable magnitude. For the test chip pre-amplifier design, simulations were utilized to determine the capacitor size, 1.75 pF, which reduced noise level to a magnitude of one LSB step, 1.6 mV. The array of 32 1.75 pF reference voltage storage capacitors were placed between the ADC core and the input I/O pins as shown in Figure 6.3 to minimize the wire resistance between the capacitors and the comparators.

6.4.2 DAC Operation

To produce the voltage references, a DAC architecture based on the passive DAC architecture presented in [29] was utilized. In [29], the DAC is utilized to drive the varactor of a voltage-controlled oscillator within a digital phased-locked loop (DPLL) with an update rate of 50 MHz. The power consumption of the passive DAC is dominated by the resistor ladder with a small power overhead due the low frequency digital switching of the control logic. An additional benefit of the passive DAC is the two-stage segmentation, which reduces the complexity of the 8-bit DAC to two stages of 4-bit complexity. The reduction in complexity reduces the number of elements and relaxes element matching requirements.

To utilize the passive DAC for use in a variable-reference ADC application, a novel modification was necessary. The single-core single-output architecture of the passive DAC was expanded to provide single-core multiple-output operation. To do so, the transistor array between the unit capacitors and the reference voltage storage capacitor was modified to select one of the 32 1.75 pF storage capacitors. The 32 storage capacitors are updated in sequential order, which effectively reduces the update frequency by a factor of 32. The number of output capacitors was increased from 30 to 32 to reduce the digital control complexity. The nominal, post-extraction operating frequency of the DAC core is 12 MHz, resulting in an update frequency for each individual reference voltage storage capacitor of 375 kHz, a rate more than sufficient to counteract leakage within the circuit.

The DAC, with exception of the reference voltage storage capacitors, is placed in the 1.2 V digital power supply domain (see Figure 6.4) since it utilizes a clock. Due to the common mode and range requirements, a resistor divider is utilized to generate the maximum and minimum voltage references

$$REF\_L = 0.6 \text{ V} \tag{6.2}$$

and

$$REF\_H = 1.0 \text{ V}. \tag{6.3}$$

The updating of a single voltage reference by the DAC occurs over two non-overlapping phases: phase 1 and phase 2. The reference voltage of the 8-bit DAC is set by an 8-bit reference voltage code, which is prefetched from the DAC memory array unit during the update period before it is utilized. In the first phase, as shown in Figure 6.6, the 4 MSB bits, *M*, of the 8-bit reference voltage code are utilized to determine which voltage levels in the 4-bit resistor ladder (R_u = 68 Ω) are utilized to charge the 4-bit array of unit capacitors (C_u = 24.6 fF). The number of unit capacitors charged to the higher voltage resistor ladder reference is determined by the 4 LSB bits, *N*, of the 8-bit reference voltage code, with the other unit capacitors charged to the lower voltage resistor ladder reference. The period of phase 1 is determined by the settling time requirement as discussed in [30]. After phase 1 but before phase 2, the unit capacitors are isolated from the resistor ladder through the opening of the switches controlled by the phase 1 clock, DAC_CLK1. In phase 2, as shown in Figure 6.7, the unit capacitors are connected to the reference voltage storage capacitor by closing the switches controlled by the LOAD_SELECT bus. Instead of always connecting to the same storage capacitor during phase 2, the DAC updates the storage capacitor corresponding to the 8-bit voltage reference code. When the unit capacitors and storage capacitors are connected together, charge sharing occurs, resulting in a transition in the reference voltage toward the desired voltage. The period of phase 2 is determined by the settling time requirement as discussed in [30]. Effectively, the DAC is a first-order switch capacitor filter with the reference voltage settling according to a single-pole cut-off frequency

$$f_o = \frac{C_u}{4\pi C_{ref}} f_{DAC} \tag{6.4}$$

where f_{DAC} is the operating frequency of the DAC [29], e.g., the frequency at which DAC performs an update on any reference voltage. Each individual voltage reference is updated at a frequency of f_{DAC} /32. After phase 2 but before phase 1, the unit capacitors are isolated from the voltage storage capacitor through the opening of the switches controlled by the LOAD_SELECT

bus, completing a voltage reference update operation. The DAC updates each 32 of the reference voltage storage capacitors in sequential order over 32 DAC clock periods, continuing to update all reference voltages during normal operation (assuming the DAC is clocked).

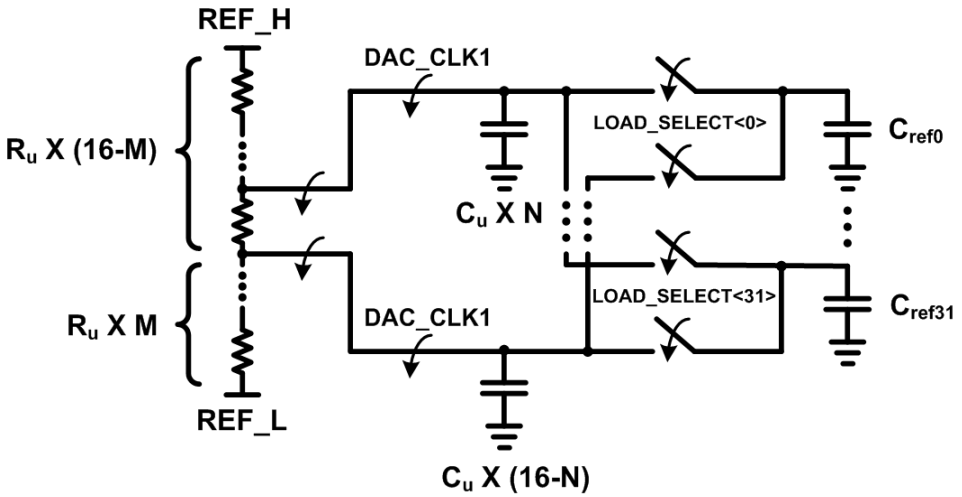


Figure 6.6: Passive single-core multiple-output DAC phase 1 operation. The 4-bit unit capacitor array is charged utilizing the voltage levels generated from a 4-bit resistor ladder.

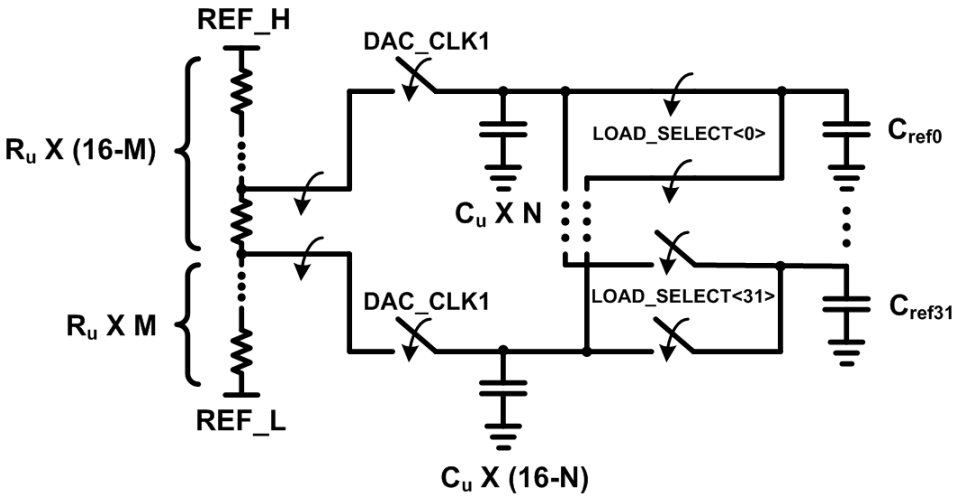


Figure 6.7: Passive single-core multiple-output DAC phase 2 operation. The unit capacitors are connected to the reference voltage storage capacitors to update the reference voltage.

6.4.3 Low-Area DAC Operation

The array of 512 switches required to connect the unit capacitor nodes to the storage capacitor nodes will require an increase in the area requirement of the DAC. To reduce the number of switches, an intermediate node can be introduced, as shown in Figure 6.8, modifying the operation of phase 2. During the modified phase 2, the DAC_CLK2 switches close,

connecting the unit capacitors to the intermediate node while a single LOAD switch is closed, connecting the target storage capacitor to the intermediate node. While charge sharing occurs between the unit capacitors and the target voltage storage capacitor, undesired charge sharing occurs with the parasitic capacitances of the intermediate node. The parasitic capacitances of the intermediate node are charged to the previously updated reference voltage, resulting in a deterministic offset in the voltage references. Due to the large number of switches connected to the intermediate node and wiring capacitance, the parasitic capacitance of the intermediate node is substantial enough to result in an offset comparable to 1 LSB. To combat the offset, an operational amplifier can be utilized to precharge the intermediate voltage to equal that of the target storage capacitor during phase 1. Unfortunately, the addition of an operational amplifier requires additional power consumption. For the variable reference ADC-based receiver, minimizing power consumption was a greater priority than minimizing area, and therefore the DAC operation described in Section 6.4.2 was utilized.

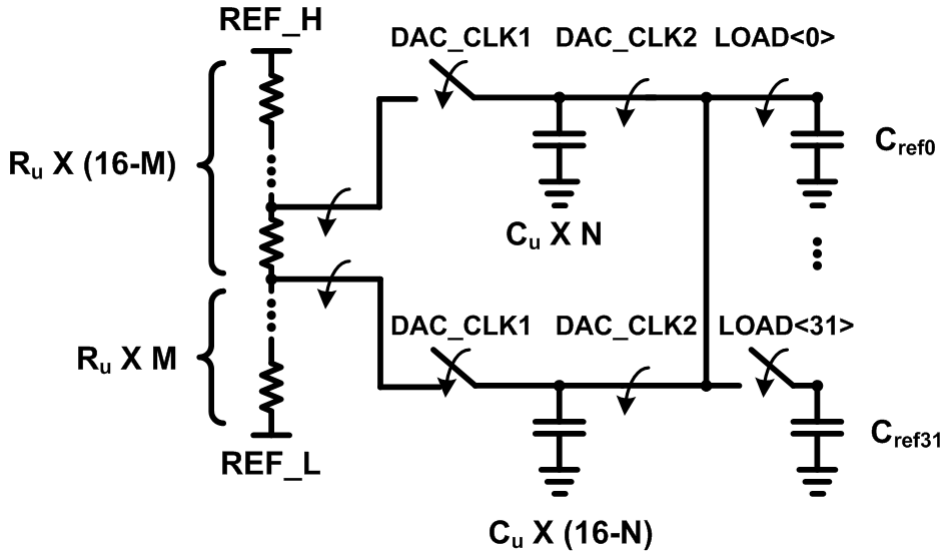


Figure 6.8: DAC phase 2 operation with reduced switch array area. The intermediate node between the DAC_CLK2 and LOAD switches produces a deterministic offset. The offset is produced when the parasitic capacitances of the intermediate node charge share with the unit capacitors and target voltage reference storage capacitor during phase 2. The parasitic capacitances of the intermediate node are charged to the previously updated reference voltage before the charge sharing occurs.

6.5 DAC Architecture and Circuits

6.5.1 DAC Architecture

A block diagram of the DAC is shown in Figure 6.9. The DAC consists of three main units: DAC memory, DAC digital control, and DAC core. The DAC memory stores the 32 8-bit reference voltage codes and prefetches the reference voltage code the update before it is utilized. The DAC digital control takes the 8-bit reference voltage code and generates the control signals for the DAC core. The DAC core contains all of the mixed-signal circuitry utilized to generate the desired analog voltage on the reference voltage storage capacitors.

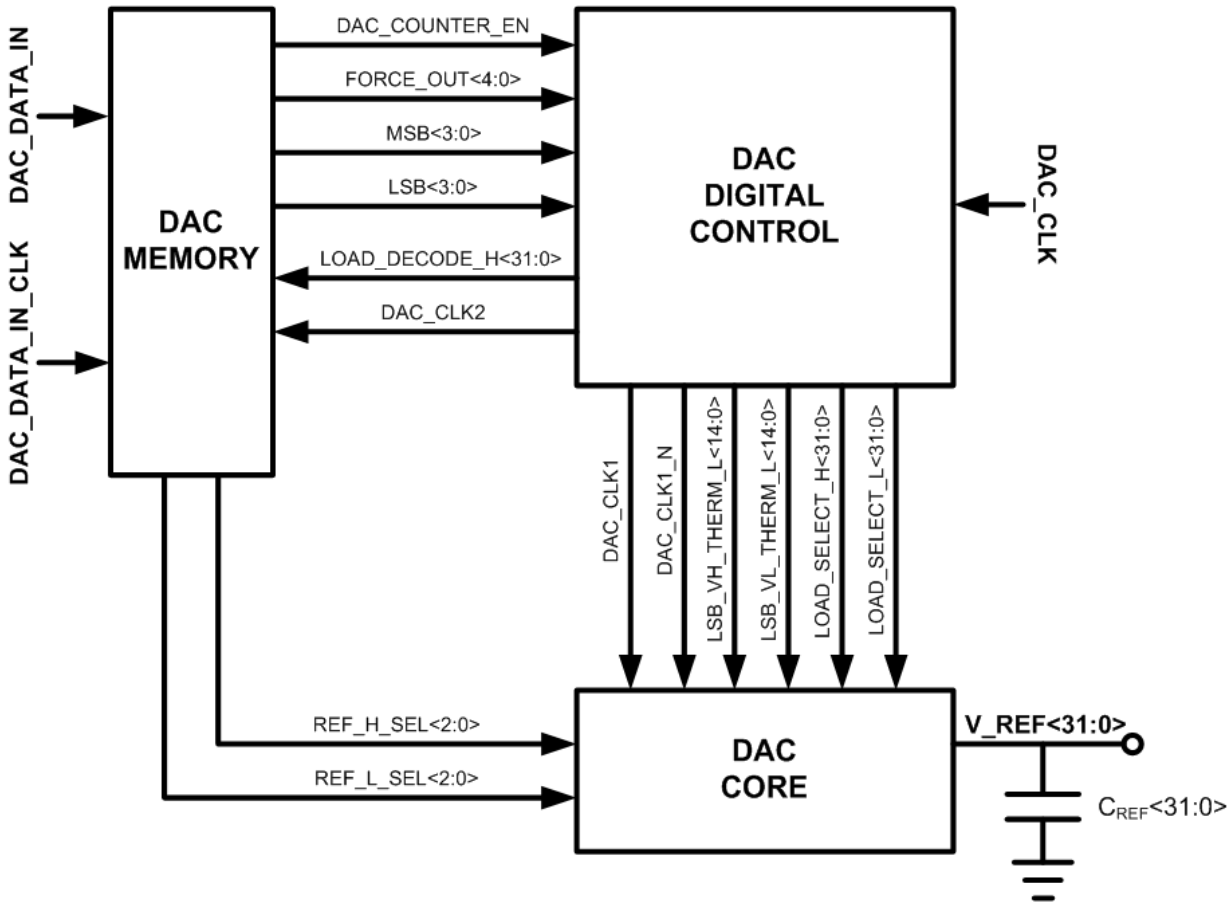


Figure 6.9: DAC block diagram. DAC input and output signals are highlighted in bold.

The DAC has 3 inputs; DAC_CLK, DAC_DATA_IN, and DAC_DATA_IN_CLK, and one output bus, V_REF. The DAC_CLK is utilized to control the frequency of operation, f_{DAC}. Since the DAC memory bits are a part of the register bank (see Figure 6.4), DAC_DATA_IN is

connected to the preceding register bank bit. DAC_DATA_IN_CLK is the register bank clock signal and is utilized when programming the DAC memory bits. The V_REF bus is the voltage references utilized by the ADC comparators.

6.5.2 DAC Memory

A block diagram of the DAC memory unit is shown in Figure 6.10. The memory unit is an array of 268 bits in a shift register alignment for programming. The first bit of the array, DAC_COUNTER_EN, is utilized to enable or disable the counter in the digital control unit for debugging purposes. Disabling the counter forces the DAC to only update a single reference rather than all 32. To determine which reference is updated, the bus FORCE_OUT, the next five bits of the memory array, is utilized. The next three bits which make up the bus REF_L_SEL are utilized to change the number of resistors in the resistor ladder between REF_L and VSS. The following three bits which make up the bus REF_H_SEL are utilized to change the number of resistors in the resistor ladder between REF_H and VDD. Combined, the ability to change the number of resistors in the ladder in both locations allows for the user to shift the DAC voltage range and either compress or expand the single-ended voltage range. The last 256 bits of the memory array are configured into 32 byte (8-bit voltage reference code) addressable units. The LOAD_DECODE_H bus is utilized to determine which byte (voltage reference code) is read from the memory. A single D flip-flop is utilized, clocked on the rising edge of phase 2, to store the voltage reference code for utilization during the next DAC phase 1 operation. Therefore, the LOAD_DECODE_H bus actually addresses the reference voltage code of the voltage reference which will be updated during the next update period. Therefore, the DAC is effectively prefetching the reference voltage codes. Since phase 1, when the LOAD_DECODE_H bus is updated, is approximately 40 ns long, there is more than sufficient timing slack for the read operation, and therefore, all of the circuits in the DAC memory unit utilize minimum-sized transistors to minimize area and power consumption. In this dissertation, a minimum-sized transistor is defined as having the minimum gate length and the minimum gate width for which the transistor p-cell generates two source and drain contacts. Two source and drain contacts are desired for reliability reasons.

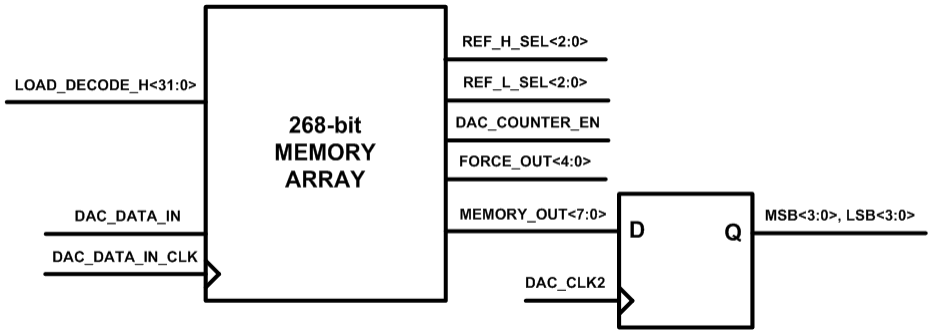


Figure 6.10: DAC memory unit block diagram.

6.5.3 DAC Digital Control

A block diagram of the DAC digital control unit is shown in Figure 6.11. The digital control unit generates the control signals utilized within the DAC core unit to update the reference voltages. Due to the low operational frequency of the DAC core, 12 MHz, all of the digital logic was implementable using minimum-sized transistors, as defined in Section 6.5.2, to minimize area and power consumption.

To generate the two non-overlapping phases of operation, DAC_CLK1 (phase 1) and DAC_CLK2 (phase 2), a two-phase non-overlapping clock generator with 14 inverter delays separating the clock phases is utilized. The non-overlapping clock generator also produces inverted polarity clock phases, DAC_CLK1_N and DAC_CLK2_N. The inverted polarity clock phase is designed to lead the non-inverted clock phase by one inverter delay to improve the charge injection cancellation performance, which will be presented in Section 6.5.5.

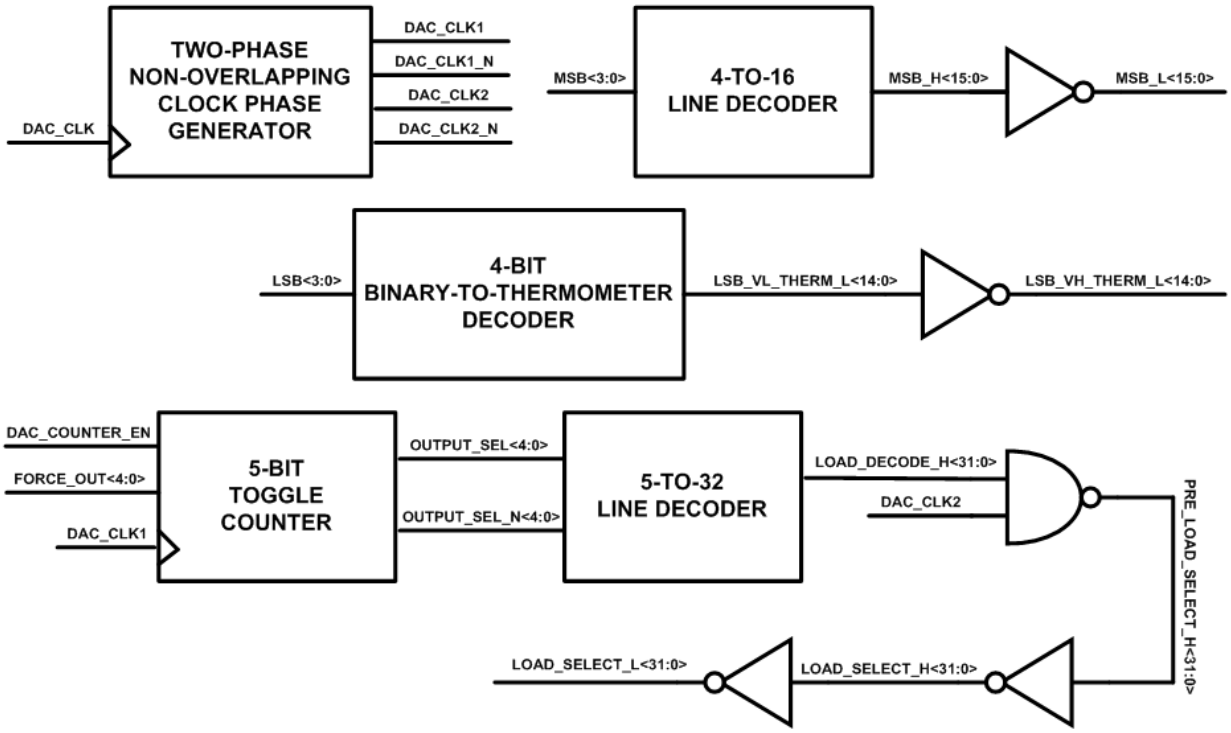


Figure 6.11: DAC digital control unit block diagram.

To determine which of the 32 reference voltages to update, a 5-bit toggle counter is utilized and increments at the rising edge of DAC_CLK1. To select the voltage reference capacitor to be updated, the counter output is decoded into a 32-bit bus, LOAD_DECODE_H, utilizing a 5-to-32 line decoder. The LOAD_DECODE_H bus is directly utilized to address the memory array holding the voltage reference codes, as discussed in Section 6.5.2. Since the unit capacitors and the voltage reference storage capacitors should only be connected during phase 2, the LOAD_DECODE_H bus and DAC_CLK2 are fed into a NAND gate. The NAND gate is followed by a series of inverters utilized to drive the large pass gate load and provide both inverted and non-inverted signal polarities. If only a single reference voltage is to be updated for debugging reasons, DAC_COUNTER_EN disables the 5-bit toggle counter and the FORCE_OUT bus is utilized as input to the 5-to-32 line decoder.

To generate the control signals that select which voltage levels in the 4-bit resistor ladder are utilized, a 4-to-16 line decoder is utilized. The line decoder takes the MSB bus as input and generates the bus MSB_H which is then inverted in order to provide the DAC core unit both signal polarities.

To generate the control signals that select the number of unit capacitors charged to the higher voltage ladder reference, a 4-bit binary-to-thermometer decoder is utilized. The decoder takes the LSB bus as input and generates the bus LSB_VL_THERM_L (an active low signal) which is then inverted to generate LSB_VH_THERM_L (an active low signal). LSB_VL_THERM_L is utilized to select the lower voltage ladder reference for the corresponding unit capacitor while LSB_VH_THERM_L is utilized to select the higher voltage ladder reference.

6.5.4 DAC Core

A block diagram of the DAC core is shown in Figure 6.12. The DAC core utilizes the control signals generated by the DAC memory unit and DAC digital control unit to update the reference levels. The DAC core is dominated by the DAC switch block which will be presented in detail in Section 6.5.5. The remainder of the DAC core is utilized to generate the resistor ladder reference voltages.

To generate the resistor ladder reference voltages, a series of resistors connect VDD to VSS with a nominal drop of 25 mV across each unit resistor, R_u (68 Ω). To generate the required 4-bit resolution, 18 resistors are implemented within the resistor reference ladder block. To meet the DAC range requirements specified in Section 6.5.2, static and selectable unit resistors are placed between VDD and REF_H in the REF_H generator block and between REF_L and VSS in the REF_L block.

In Section 6.5.2, the function of control busses REF_H_SEL and REF_L_SEL was presented in detail. In the DAC core, REF_H_SEL and REF_L_SEL are decoded utilizing a 3-to-8 line decoder to add that number (0 to 7) of unit resistors to the resistor ladder between VDD and REF_H or REF_L and VSS, respectively. The nominal setting is 4 for REF_H_SEL and 3 for REF_L_SEL. These nominal settings were chosen in order to be able to shift the common mode higher than it can be shifted lower since the test chip pre-amplifiers utilize NMOS transconductance transistors (see Figure 6.5). A PMOS switch, designed to produce a static 25 mV drop, is utilized to enable the string of resistors specified by REF_H_SEL in the REF_H generator block, while a NMOS switch, designed to produce a static 25 mV drop, is utilized to enable the string of resistors specified by REF_H_SEL in the REF_L generator block. With a nominal 25 mV of drop from each resistor, PMOS, and NMOS switch, two static unit resistors

are placed between VDD and REF_H and 19 static unit resistors are placed between REF_L and VSS to produce the required DAC range and common mode.

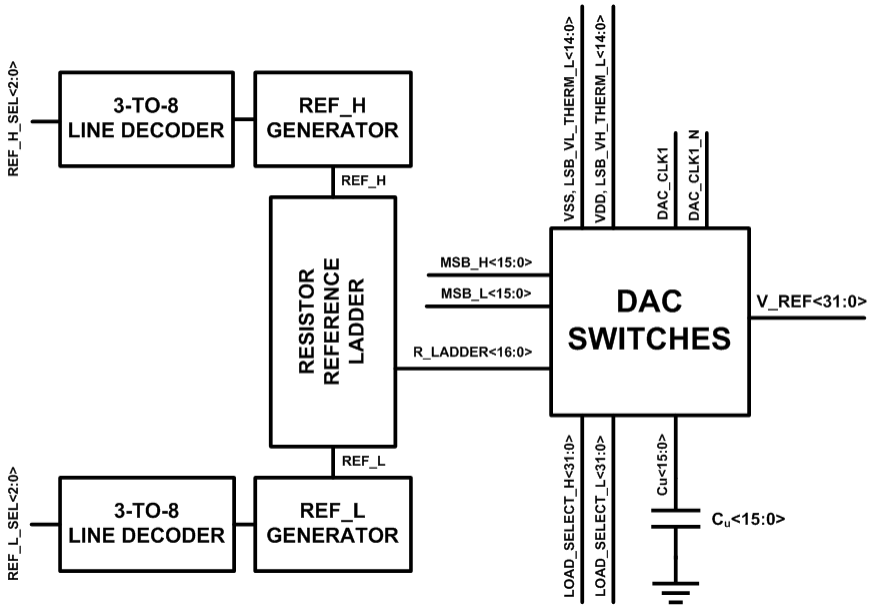


Figure 6.12: DAC core unit block diagram.

6.5.5 DAC Switches

A schematic of the DAC switches block is shown in Figure 6.13. The R_LADDER bus is connected to the 4-bit resistor, the Cu bus is connected to the 4-bit unit capacitor array, and the V_REF bus is connected to the 32 reference voltage storage.

The DAC switches block implements the operations outlined in Section 6.3. During phase 1, MSB_H and MSB_L are utilized to control pass gates that determine which voltage levels in the 4-bit resistor ladder will be utilized to charge the 4-bit array of unit capacitors. Then, NMOS transistors were found to be sufficient for selecting which unit capacitors are charged to the higher voltage resistor ladder reference (through the control bus LSB_VH_THERM_L) and which unit capacitors are charged to the lower resistor ladder reference (through the control bus LSB_VL_THERM_L). NMOS transistors were also sufficient for isolating the unit capacitors from the reference ladder when they are not being charged through the inverted phase 1 clock, DAC_CLK1_N. To combat charge injection resulting from the rising edge of the DAC_CLK1_N, NMOS capacitors half the size of the NMOS switch controlled by DAC_CLK1_N were added to unit capacitor bus, C_u. Since the NMOS capacitor is driven by

DAC_CLK1, which lags DAC_CLK1_N by one inverter delay (see Section 6.5.3), the deterministic offset on the unit capacitors due to charge injection is efficiently reduced.

In phase 2, the LOAD_SELECT and LOAD_SELECT_N buses are utilized to control a pass gate array that determines which reference voltage storage capacitor is updated. Due to charge injection, PMOS and NMOS capacitors half the combined size of the pass gate transistors are utilized to minimize the deterministic offset.

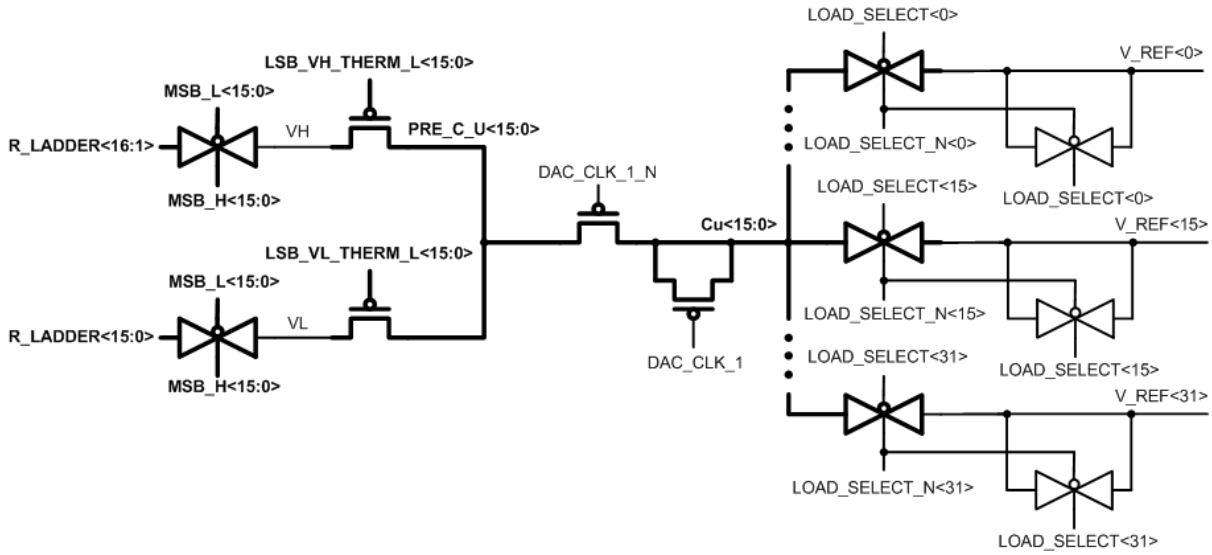


Figure 6.13: Schematic of the DAC switches. 16-way parallel pass gates and busses utilize thick lines. The R_LADDER bus is connected to the 4-bit reference ladder. The Cu bus is connected to the 4-bit unit capacitor array. The V_REF bus is connected to the reference voltage storage capacitor array.

6.6 DAC Simulation Results

To determine the performance of DAC, post-extraction simulation was utilized because the measured test chip performance will be based upon the entire ADC, not just the DAC. While additional pins could be utilized to measure the DAC performance, two issues arise. The first is the ability to measure the performance without impacting the DAC operation. If the reference voltages are connected to external measurement equipment, the capacitive load on the DAC and leakage current will increase, potentially changing the generated reference voltages. The second issue results from the limited number of package pins and the optimization of those pins for the system level investigation of a variable-reference ADC-based receiver.

6.6.1 DAC Initialization, Range, and Voltage Variations

In Figure 6.14, a post-extraction simulation of the DAC initialization is shown. The simulation utilizes a Verilog-A block to replace the DAC memory unit in order to reduce simulation time. The results show the settling of all 32 reference voltage storage capacitors to a set of reference voltages spanning the output range, 400 mV. Seventeen consecutive reference voltages are set to the 17 lowest references to demonstrate the LSB steps near the minimum reference and will be investigated in Section 6.6.2. The next 15 consecutive references are utilized to span the DAC range through incremental increase of the MSB bits. The simulation demonstrates how the reference voltages are updated sequentially with individual references updating at a rate of 375 kHz. The reference voltages approach a steady-state value after approximately 25 µs, equivalent to 10 updates. At steady-state, variation is observed on the reference voltages.

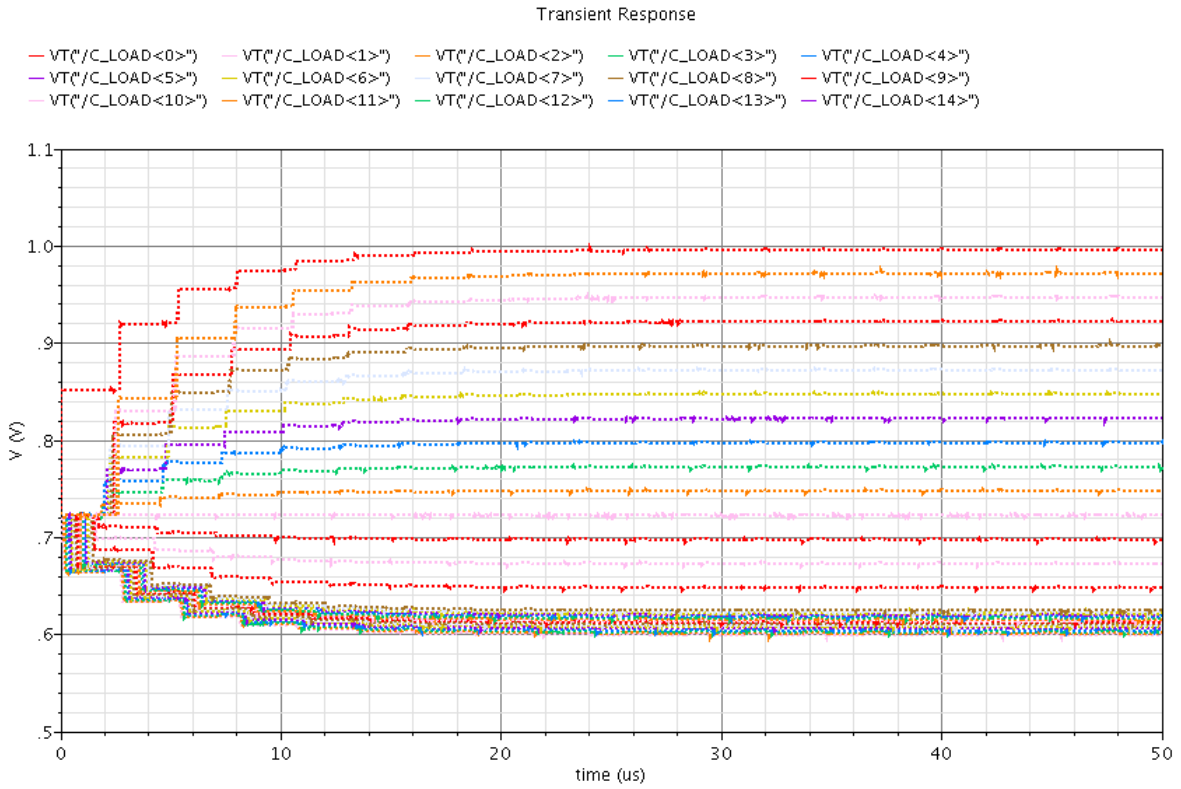


Figure 6.14: DAC initialization simulation for a set of 32 reference voltages spanning the output range.

A plot of a single reference voltage at steady-state from Figure 6.14 is shown in Figure 6.15. The variation observed in Figure 6.14 is shown to be deterministic in Figure 6.15 with a 2.7 μs period. The variation is correlated with the reference voltage which is being updated. The low variation region corresponds to updates that occur during the 17 consecutive updates of the minimum reference voltages. The triangular variation region, with significant change in the reference voltage on the order of 1.3 mV over 1.3 μs, occurs during the 15 consecutive updates that span the range of the DAC. Therefore, the post-extraction simulation results point toward capacitive coupling between the unit capacitor nodes and the storage capacitor nodes even when the switches isolating the nodes are open. The post-extraction netlist presents two sources of capacitance between the nodes: the drain-to-source capacitance in the pass gates implementing the switches and capacitance between the node wires. Since the capacitive coupling between the unit capacitor nodes and the reference voltage storage capacitors is effectively constant for each storage capacitor node, this deterministic noise is coupled onto all output nodes with approximately equal magnitude, and can therefore be described as a common mode variation on the ADC references. Since the ADC utilizes two reference voltages to generate a differential reference which is then compared to a differential input signal, this common mode variation can essentially be disregarded.

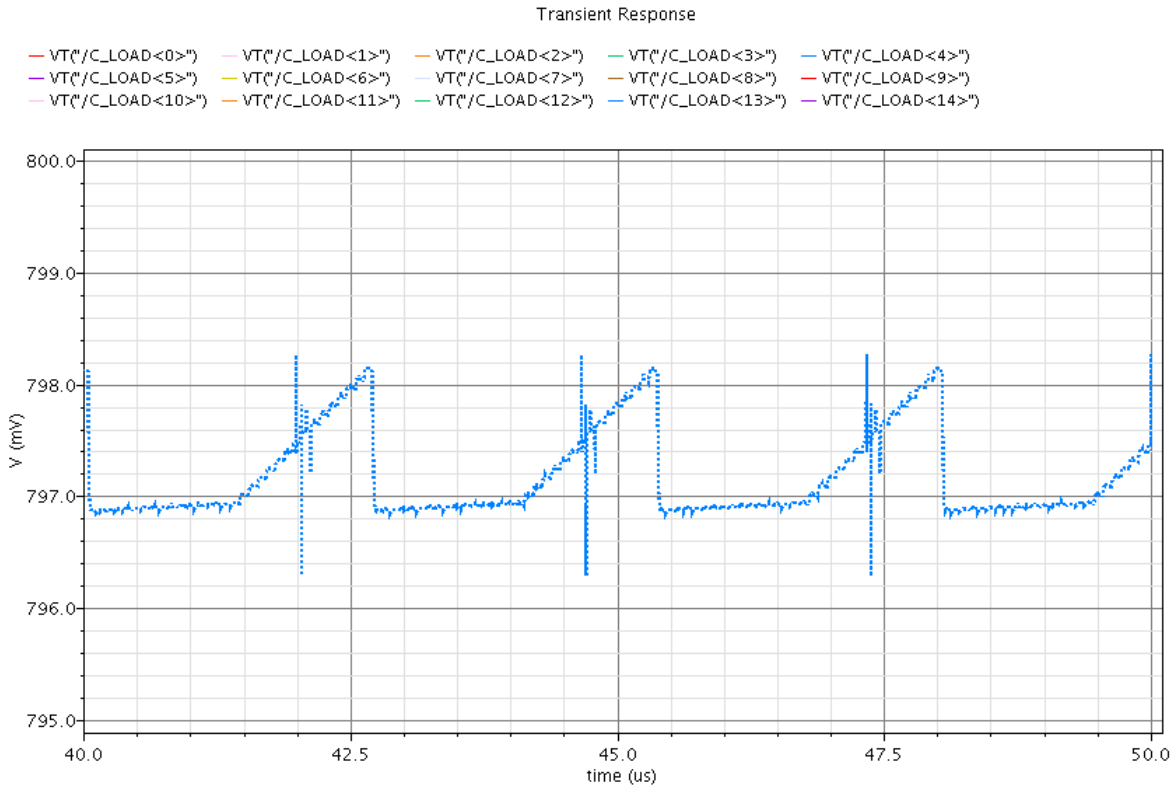


Figure 6.15: Plot of the deterministic voltage reference variation due to coupling between the unit capacitors and the reference voltage storage capacitor. The variation is coupled to all reference voltages and is therefore a common mode variation on the ADC references. Since the ADC utilizes differential references and input, this variation can be disregarded.

Additional short variations of the reference voltage are observed in Figure 6.15 during the 1.3 µs triangular variation. An example of the most significant of the short variations is shown in detail in Figure 6.16. The variation has a period of approximately 200 ps and a maximum deviation of approximately 1.2 mV. These variations correlate specifically to the beginning and end of phase 2 for the update cycle of the reference voltage shown, and are due to clock feedthrough. Initial schematic simulations showed a significantly larger variation of approximately 4 mV, but the reduction of the phase 2 clock risetime and falltime due to wiring capacitance resulted in a significant reduction in the clock feedthrough. Additional short variations, with reduced magnitude, occur during the beginning and end of phase 2 of update cycle for the reference voltage whose wiring is in close proximity to the reference. Since the magnitude of the variations is under 1.56 mV, one LSB, and the duration is less than a single bit period of received signal, 250 ns, the variation was deemed acceptable in the context of a variable reference ADC-based link.

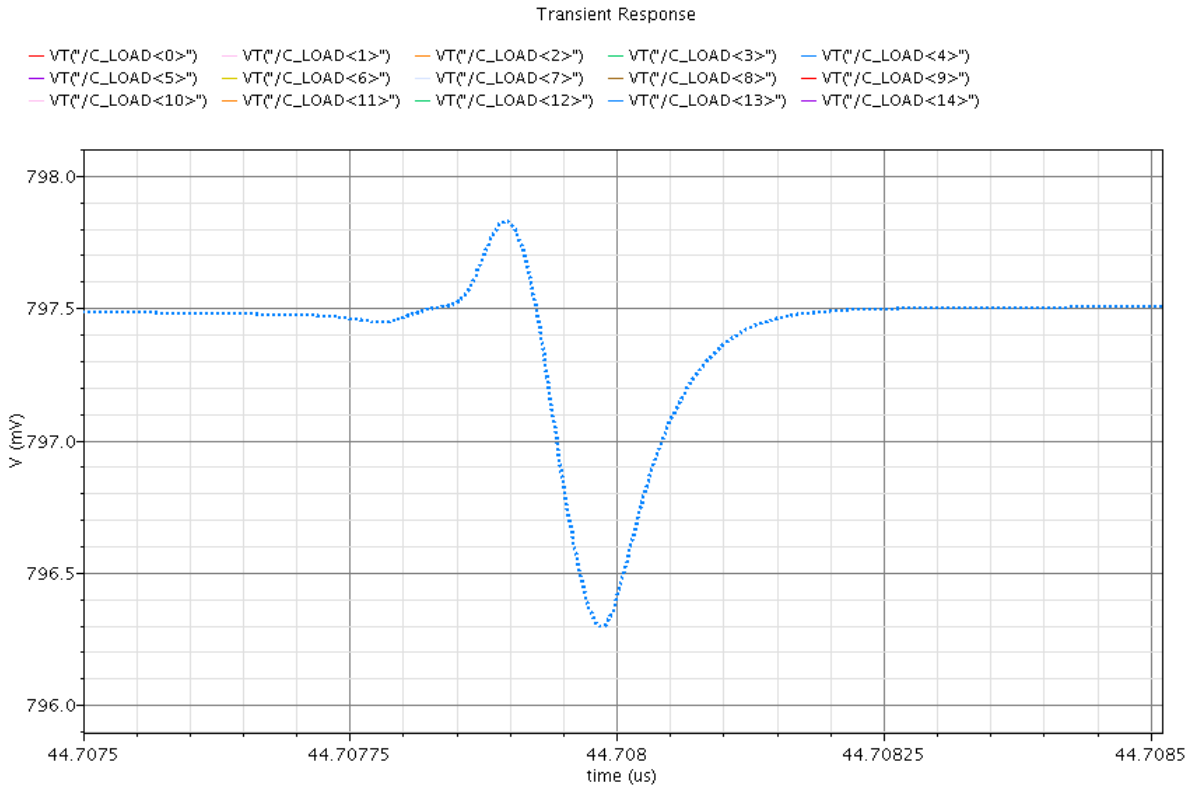


Figure 6.16: Plot of the short-duration reference voltage variation. The variation occurs due to the clock feedthrough at the beginning and end of phase 2 during the update cycle for the voltage reference and the neighboring voltage reference.

6.6.2 DAC LSB Step Size

In Figure 6.17, a post-extraction plot of the 17 lowest references from Figure 6.14 is shown. The 2.7 µs deterministic noise discussed in Section 6.6.1 is shown to couple to all references while the short-duration voltage variations due to clock feedthrough are correlated to the updates of the individual reference voltage and its neighbor. The simulation results show that the DAC step size consistency is acceptable with the extracted parasitic capacitances responsible for the majority of the variation.

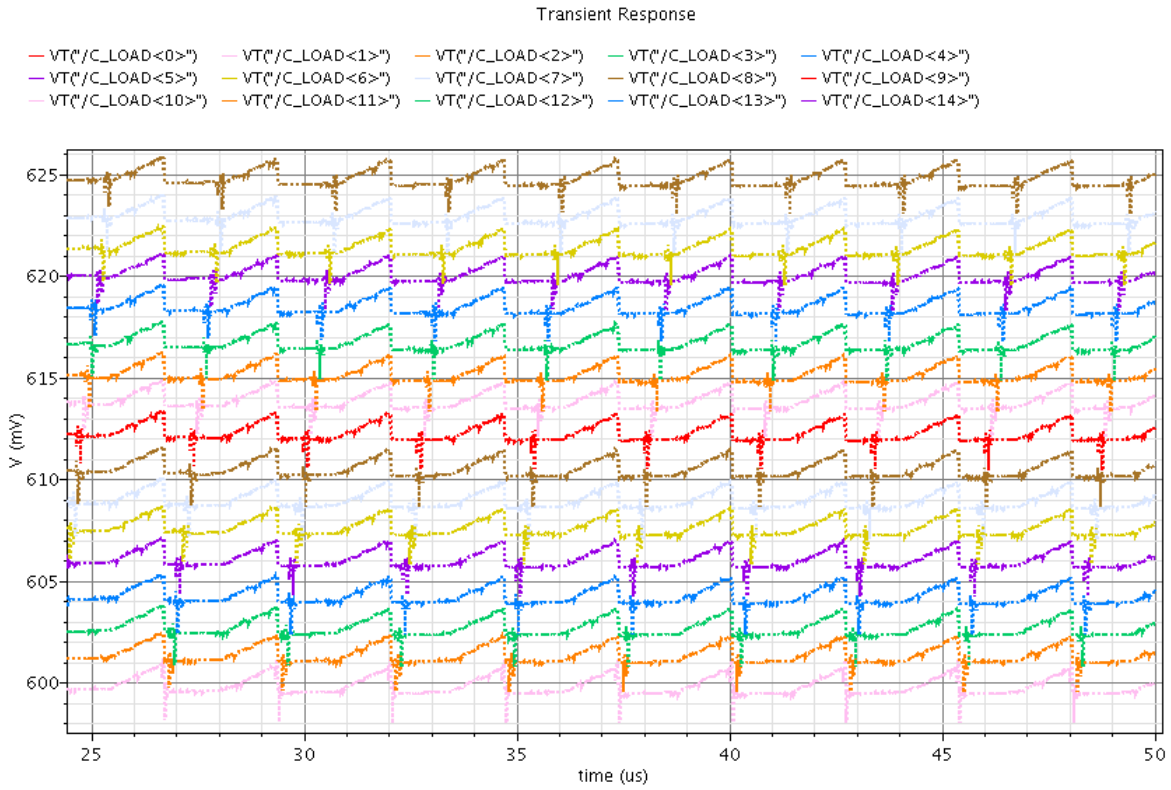


Figure 6.17: Post-extracted simulation of the 17 lowest DAC codes.

6.6.3 Random Reference Voltage Generation

In Figure 6.18, a post-extraction simulation of the DAC initialization for a random set of 32 voltage references spanning the output range is shown. The results show that the order and relation of reference voltages do not impact the DAC functionality. The random values do, on the other hand, determine the deterministic variation due to capacitive coupling between the unit capacitor nodes and the storage capacitor nodes. Due to the random ordering of the reference levels, the variation will track the reference levels being updated, demonstrating transitions with larger average magnitude even though the peak-to-peak variation will remain the same as in Figure 6.15.

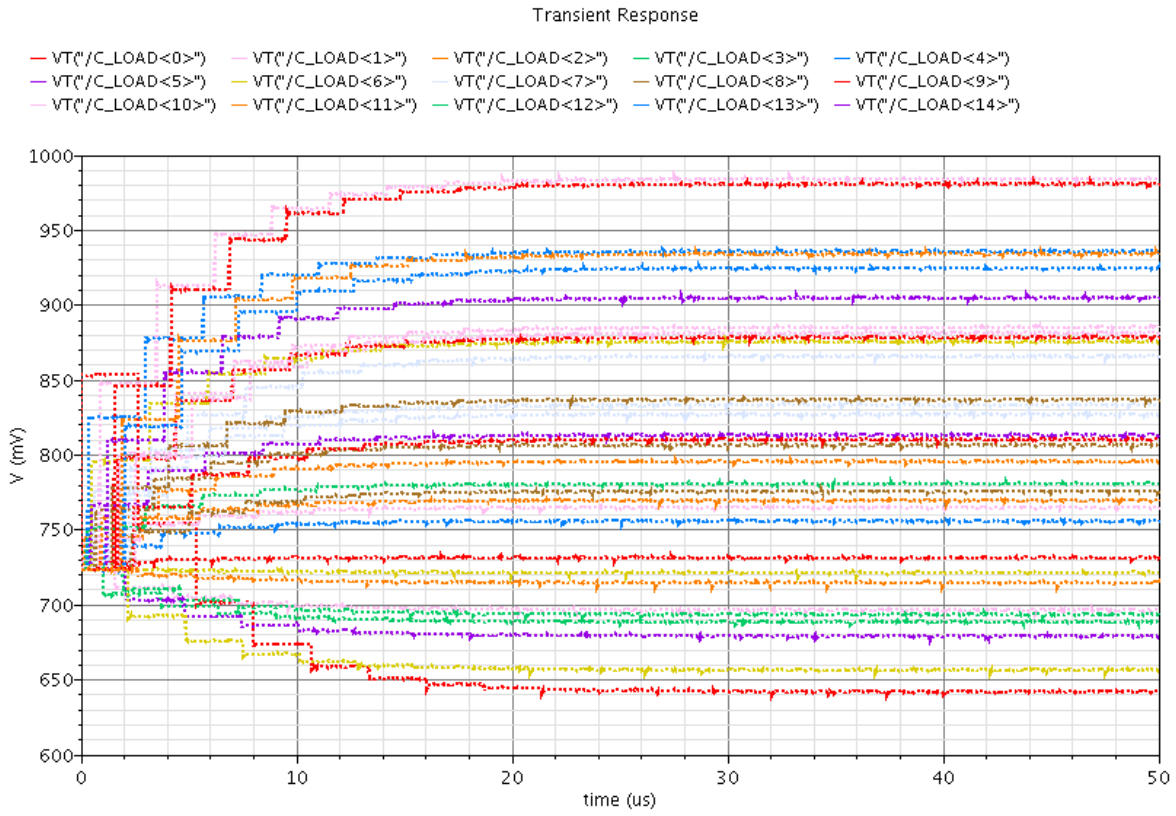


Figure 6.18: Post-extracted simulation of the reference voltages for a set of 32 random reference levels.

6.6.4 DAC Power Consumption

To determine the power consumption of the DAC, the average current flowing from the 1.2 V supply was measured during the steady-state region of the post-extraction simulation shown in Figure 6.18. Under nominal conditions, 48 series unit resistors in the DAC core unit connect VDD to VSS for a total resistance of 3 kΩ for a static power consumption of approximately 440 µW. The total simulated power consumption of the proposed DAC is approximately 0.5 mW with 90% of the power consumed by the resistor ladder. Since the 4-bit resistor ladder is required for a 4-bit uniform DAC, the power overhead for the implementation of the variable-reference DAC is 10%, approximately 50 µW.

6.7 DAC Measurement Results

Since the DAC performance cannot be measured directly, the performance results of the ADC can be utilized to demonstrate the functionality of the DAC. To demonstrate the ability to generate a variety of reference levels and the ability to calibrate the reference levels to cancel DAC and comparator variation, the ADC input is driven with a differential DC voltage from a semiconductor parameter analyzer. The DC voltage is swept to generate the differential ADC input voltage to ADC decimal output relationship shown in Figure 6.19. The pre-calibration results show that the step sizes between ADC threshold levels before calibration contain a high level of variation, indicating that the thresholds are deviating from their ideal voltage level. In fact, before calibration (for the specific die shown), the ADC output decimal level 1 is never observed. Changing the DAC reference voltage codes to calibrate the ADC threshold levels demonstrates a remarkable reduction in the step size variation, moving the thresholds close to their ideal voltage levels.

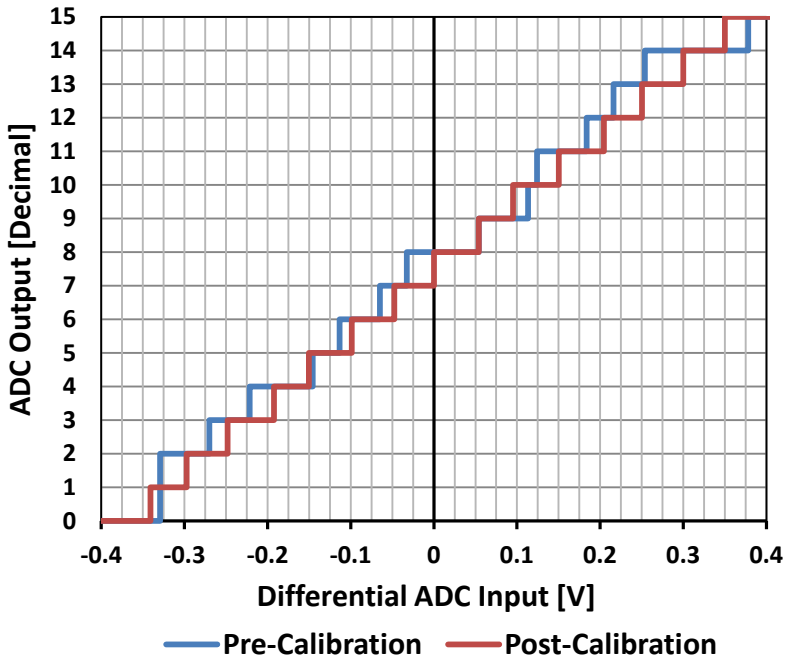


Figure 6.19: Measured pre- and post-calibration differential ADC input voltage to ADC decimal output relationship.

To demonstrate the performance of the ADC as a function of input signal frequency, the effective number of bits (ENOB) of the ADC is measured. ENOB is the precision of an ideal

ADC which would effectively equal the precision of the measured ADC. To measure ENOB, a full-range, single-frequency sinusoid is sampled by the ADC, in the 4-bit post-calibration reference level configuration, at the target sampling frequency, 4 Gb/s. The ADC output is then processed for the signal and the noise plus distortion powers. The two power levels are utilized to produce an effective precision, ENOB, in number of bits. The ENOB measurement results for the ADC are shown in Figure 6.20. The results demonstrate that at low frequencies, the ENOB is close to the 4-bit maximum for the 4-bit ADC configuration. As the input frequency is increased, the ENOB is reduced. The reduction in ENOB at higher frequencies is common, and can be a result of many non-idealities in the ADC design. The ENOB results are acceptable and imply that the DAC is operating correctly at the target operating frequency and for low and high input frequencies.

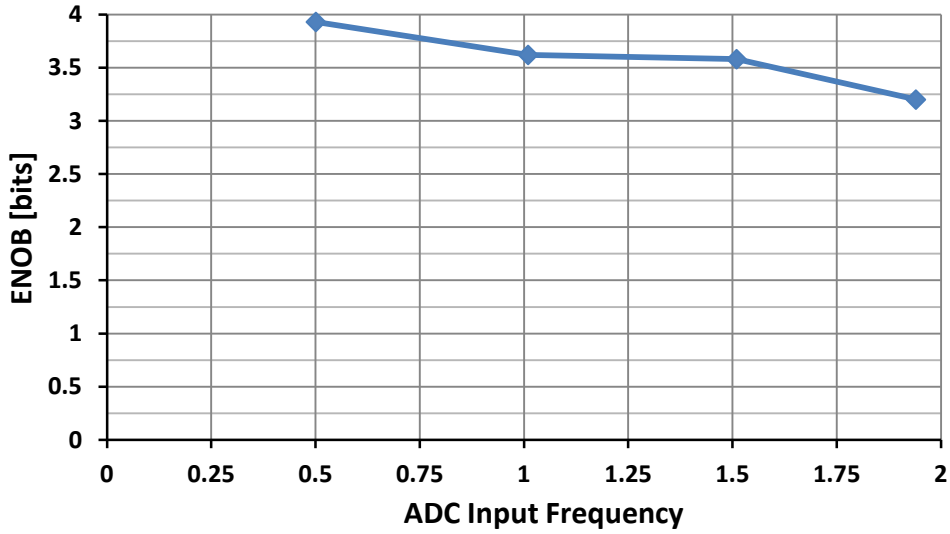


Figure 6.20: Measured ENOB of the ADC, in the 4-bit post-calibration reference level configuration.

CHAPTER 7

FEC-BASED 4 Gb/s BACKPLANE TRANSCEIVER IN 90 nm CMOS

7.1 Introduction

In this chapter, an FEC-based backplane transceiver was designed in order to demonstrate how error correction coding (ECC) can be utilized in high-speed I/O links to reduce power by (*i*) allowing for a reduced voltage swing at the transmit driver, and (*ii*) permitting the use of simpler equalization. By applying ECC, the SNR required to achieve a bit error rate of 10^{-15} drops by 3 to 6 dB, depending on the code. Thus, coding significantly reduces the bit error rate at a given slicer SNR. The relaxed SNR requirement within the receiver directly translates into a relaxed requirement on the signal swing. Since the transmitter driver has a low output impedance in order to match the characteristic impedance of a backplane channel, a reduction of the transit swing will significantly reduce the power overhead of the link. Other benefits of the diminished SNR requirements include reduced dependency on complex equalization schemes and relaxed jitter specifications on the clock jitter. An overview of the transceiver test chip design (post-extraction simulation) is presented in Table 7.1.

Table 7.1: Test Chip Design Specifications (Post-Simulation)

Data Rate:	6.25 Gb/s Line Rate 4.44 Gb/s Minimum Data Rate 6.25 Gb/s Maximum Data Rate (Uncoded)
Data Transmission:	Differential, NRZ
FEC Codes:	BCH Block Codes 0.57 Minimum Code Rate 0.9 Maximum Code Rate
Transmitter Equalization:	Pre-Emphasis (Non-Adaptive Gain Peaking)
Receiver Equalization:	None
Receiver Clock Recovery:	Full-Rate Bang-Bang PLL
Channel:	20 inch FR4
Package:	Chip-On-Board (COB)
Technology:	90 nm CMOS
Test Methodology:	On-Chip BER Counter

7.2 Test Chip Architecture and Circuits

7.2.1 System Design

The test chip design was optimized for a set of 12-to-15 dB loss (at 3.125 GHz) 20 inch channels based upon S-parameter data provided by an industrial liaison to the project. Due to the reduced operating frequency, a 20 inch channel was shown to be reliable without coding even at the minimum transmit swing of 0.25 V_{ppd}, regardless of the receiver gain setting. However, it is worth noting that the transfer characteristics of this physical 20 inch channel utilized during testing are far better than the characteristics assumed during the design process. The relatively good characteristics of this channel may be understood as follows. Four-layer test boards were manufactured using a low-cost, standard 62 mil stack-up, FR4 process. In such a process, the dielectric thicknesses are fixed, and, therefore, to produce a 50 Ω microstrip line, the width of the microstrip line is fixed. The width of the microstrip lines on the test boards is approximately 26 mils, which significantly reduces skin effect loss relative to that of the channel assumed during the design process. Therefore, the combination of reduced operating frequency and reduced channel loss (per unit length) motivates the use of a significantly longer trace length. To

achieve a channel with characteristics similar to those the test chip was optimized for, a 66 inch channel is used; the channel consists of three 20 inch and one 6 inch uncoupled 50 Ω microstrip line channel boards connected in series by means of edge mounted SMA connectors. At the maximum line rate of 4 Gb/s, this channel (see Figure 7.1), which will be referred to as the high-loss channel, has an insertion loss of 18.2 dB at the Nyquist frequency, 2 GHz.

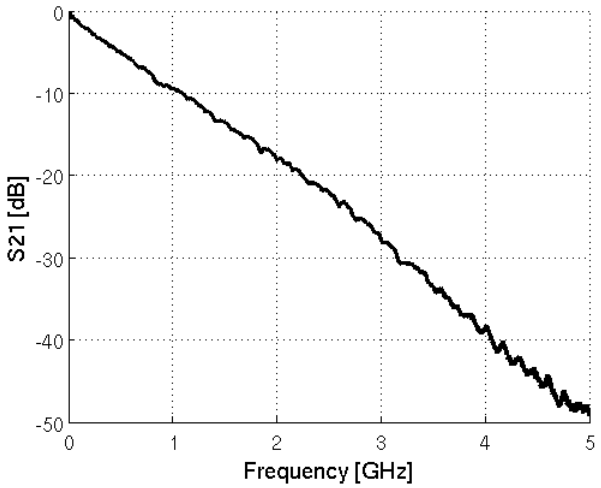


Figure 7.1. High-loss channel transfer function. Insertion loss is 18.2 dB at 2 GHz.

The effects of longer (more lossy) and shorter (less lossy) channels were also investigated. If the length of the channel is reduced to 60 inch, then reliable data transmission is achieved for all transceiver configurations (BER < 10^{-12}), whereas, if the channel length is increased to 80 inch, no transceiver configuration can achieve reliable data transmission. In the first case, the ISI is reduced, resulting in an SNR that is high enough to enable reliable data transmission even with low transmit swing, low receiver gain, and no ECC. Conversely, for the longer channel, the degradation of the channel surpasses the peaking response of the transmit driver, increasing the ISI and thereby lowering the SNR below the threshold where coding can reduce the BER to reliable levels. In summary, the transceiver design was optimized for a range of channel transfer characteristics, i.e., loss values. Using this transceiver, the benefits of ECC will not be evident for channels that are much more or less lossy than those for which the design was optimized.

The transceiver was tested across a second channel (see Figure 7.2) which includes a significant notch below 3 GHz; it will be referred to as the sub-Nyquist notch channel as line rates which place the notch minimum below Nyquist will be utilized. At the maximum line rate of 4 Gb/s, the channel has an insertion loss of 13.8 dB at the Nyquist frequency, 2 GHz. The

low-frequency notch has a maximum insertion loss of 20.5 dB at 1.58 GHz. Similar to the high-loss channel, the channel consists of an uncoupled 50 Ω microstrip lines on four-layer test boards manufactured using a low cost, standard 62 mil stack-up, FR4 process. The sub-Nyquist notch channel only utilizes a channel length of 20 inch. To generate the 1.58 GHz notch, a series combination of an 5.6 Ω 1206 surface mount resistor and 3.3 pF 1206 surface mount capacitor is used to shunt both traces at the end of the channel board closest to the receiver. While the pole generated by the surface-mount components dominates the low-frequency characteristic, the component parasitic inductances begins to dominate at 1.58 GHz. By 3 GHz, the parasitic inductance effectively isolates the channel from the shunt network, returning the insertion loss to that of an unmodified 20 inch channel for high frequencies. The high-frequency notch is likely a result of the low-cost SMA connectors utilized on the test channel boards, as the notch is seen on all test channel boards regardless of length.

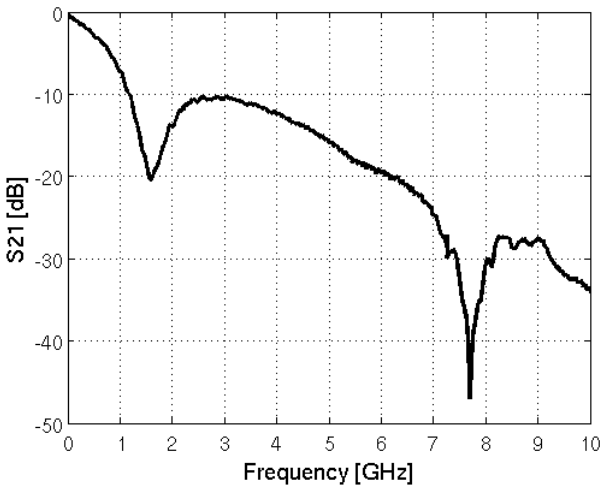


Figure 7.2. Sub-Nyquist notch channel transfer function. Low-frequency notch has a minimum (−20.5 dB) at 1.58 GHz and insertion loss is 13.8 dB at 2 GHz.

For both test channels, transmit pre-emphasis with 12 dB of peaking is implemented to boost the power of the high-frequency components compared to the low-frequency components [6]. For the high-loss channel, pre-emphasis is utilized to equalize the channel loss such that a pre-FEC BER of 10^{-4} to 10^{-6} can be achieved. For the sub-Nyquist notch channel, pre-emphasis is utilized equalize the 20 inch of channel loss and boost the high-frequency content, resulting in overshoot for transitions. The sampling point, for the sub-Nyquist notch channel, is then set within the overshoot region where the signal power is dominated by the boosted high-frequency

components. If consecutive bits are transmitted (no transition), the signal has had the opportunity to settle for at least one bit period, reducing the magnitude of ISI at the sampling location. Therefore, the sampling location generates a pre-FEC BER within the targeted range of 10^{-4} to 10^{-6}. For both test channels, short block length BCH block codes with variable code rates (CR) are employed to reduce the sampled pre-FEC BER of 10^{-4} to 10^{-6} to a post-FEC BER below 10^{-12}.

7.2.2 Transceiver Overview

The transceiver IC architecture, as tested, is shown in Figure 7.3. The analog inner transceiver consists of the transmit driver and receive amplifier, while the digital outer transceiver consists of all clocked circuitry, such as the FEC encoder and decoder.

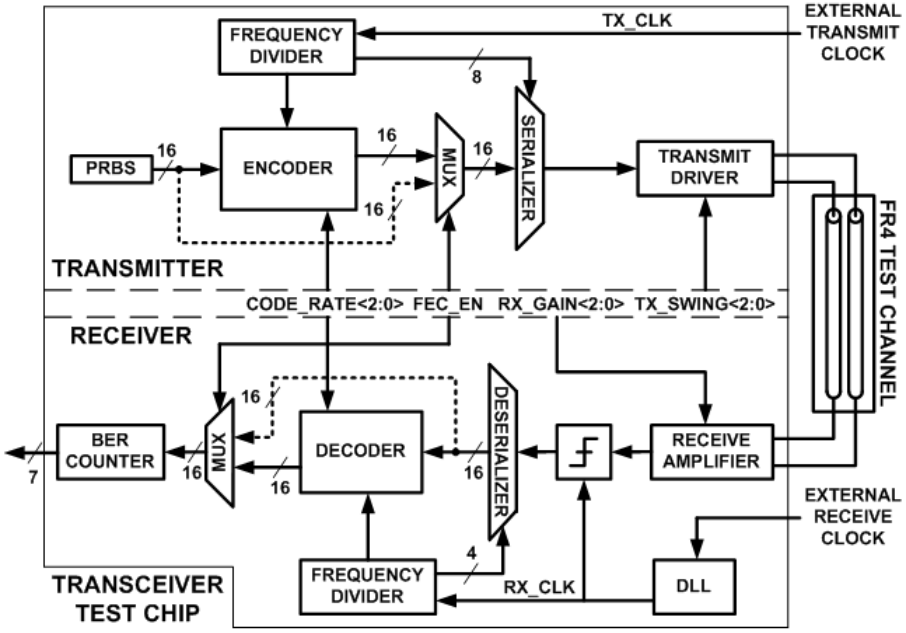


Figure 7.3. Transceiver IC architecture. Encoder and decoder bypass path for uncoded operation is designated with a dotted line.

The encoder and decoder have a 16-way parallel architecture, which makes the core clock frequency 1/16 of the line rate, and effects an intrinsic interleaving function for free, thereby compensating for burst errors [17]. The encoder and decoder can be bypassed to allow for uncoded transmission ($CR = 1$), in order to enable comparison with conventional transceivers.

The transceiver operates at a maximum line rate of 4 Gb/s, and employs external clock references. The external receive clock drives a DLL, which generates the internal sampling clock RX_CLK. To sample the data at the receiver, a ratioed CMOS SR-latch with an output shorting transistor is utilized [31].

The transceiver was originally designed to operate at 6.25 Gb/s, as shown in Table 7.1. While the transmitter was successfully operated up to the design frequency of 6.25 GHz, the receiver does not function correctly above 4 Gb/s due to an inadvertent frequency limitation within the clock multiplexer and frequency divider. Above this frequency, the low-frequency (digital) clock will have clock periods greater than the desired 16 high-frequency clock cycles, resulting in a catastrophic link failure. The tuning range of the CDR was designed to be 5.25 to 7.5 GHz; therefore, the DLL must be utilized to generate the sampling phase at the receiver at the chosen operating frequency of 4.0 GHz. Since the minimum operating frequency of the DLL design is 4.7 GHz, the delay line is locked to the maximum delay and an external SMA phase adjuster is utilized to delay the DLL clock reference, allowing the entire UI to be sampled. At 4.0 GHz operation, all of the receiver components required to complete the link data path operate as designed, resulting in a maximum transceiver line rate of 4.0 Gb/s.

In the transmitter, sixteen 2^9-PRBS units are interleaved to generate pseudo-random data which are then encoded. A 16:1 serializer multiplexes the encoded data, which are transmitted off-chip via a 100 Ω differential transmit driver.

In the receiver, an amplifier boosts the signal swing before sampling, followed by a 1:16 deserializer. The FEC decoder then detects and corrects errors in the data stream before passing the data to an on-chip 7-bit BER counter. The BER counter consists of sixteen 2^9-PRBS units employing the same polynomial as the transmitter. Errors are detected by comparing this sequence against the decoded data. The PRBS sequence in the BER counter is aligned with the decoded data through an initialization sequence. This process also detects the code word boundaries so that the BER counter compares only the systematic bits. The BER counter output is driven off-chip to allow for read-out of BER values between 10^{-7} and 10^{-12}. Since links with BER above 10^{-7} are considered unreliable and those with BER 10^{-12} are considered reliable, it was not considered worthwhile to precisely measure BERs outside the range of 10^{-7} to 10^{-12}.

The transceiver IC is a 2 mm x 4 mm die in a 90 nm CMOS technology. The die was mounted on a 5 inch, four-layer Rodgers 4003 dielectric PCB test board, shown in Figure 7.4,

using a chip-on-board (COB) process; connecting the on-chip bondpads to the gold on-board landing pads using gold bondwires. The test board traces feeding the transceiver high-speed inputs and outputs are grounded coplanar 50 Ω transmission lines. The die microphotograph is shown in Figure 7.5. The digital core is highlighted and consumes 1.11 mm^2 of die area.

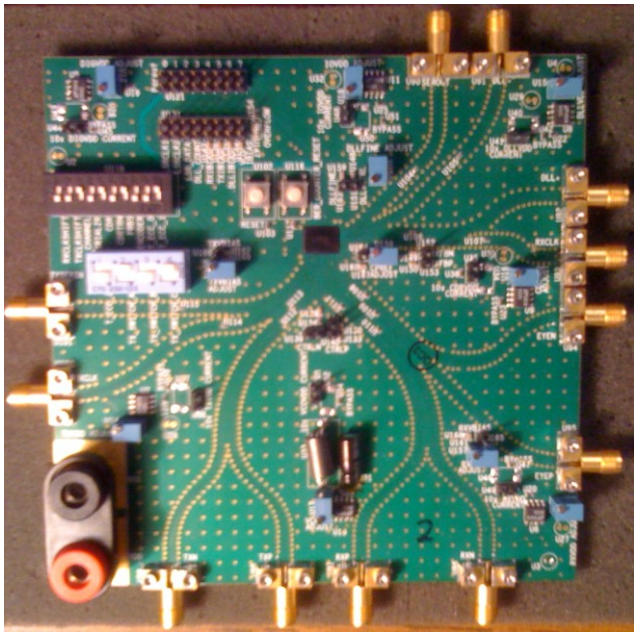


Figure 7.4. Five inch, 4-layer Rodgers 4003 dielectric PCB test board with test chip die mounted using a chip-on-board (COB) process. The die is protected underneath the black encapsulation rectangle at the center of the test board.

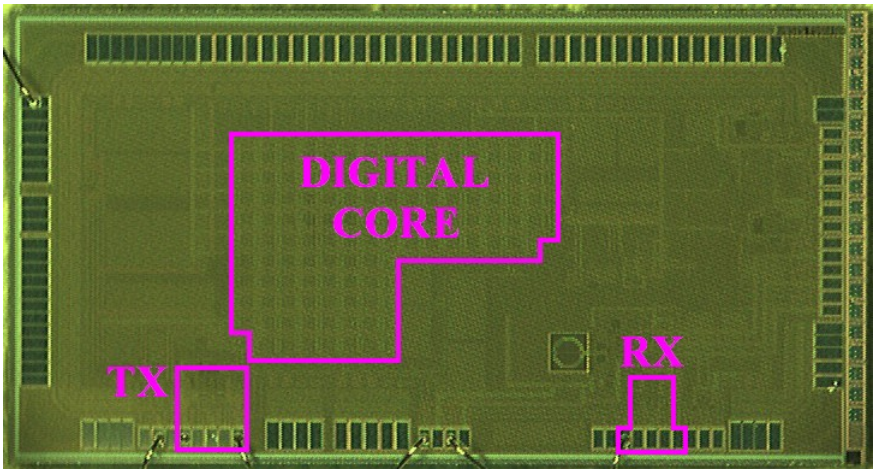


Figure 7.5. Die microphotograph with digital core, transmit driver (TX), and receive amplifier (RX) highlighted.

7.2.3 Transceiver Circuits

7.2.3.1 Transmit Driver

The 100 Ω differential off-chip driver [32], shown in Figure 7.6, implements pre-emphasis. The driver is partitioned into three transconductance stages which can be individually enabled to generate peak signal swings between 0.25 V_{ppd} and 1 V_{ppd} in 0.25 V_{ppd} increments. The RC source degeneration provides 12 dB of peaking at 3.125 GHz. The RC source degeneration requires the transconductance transistors to remain in saturation, so an H-bridge pre-driver is utilized to convert the full-swing signal output from the serializer to a small-swing signal for the transmit driver.

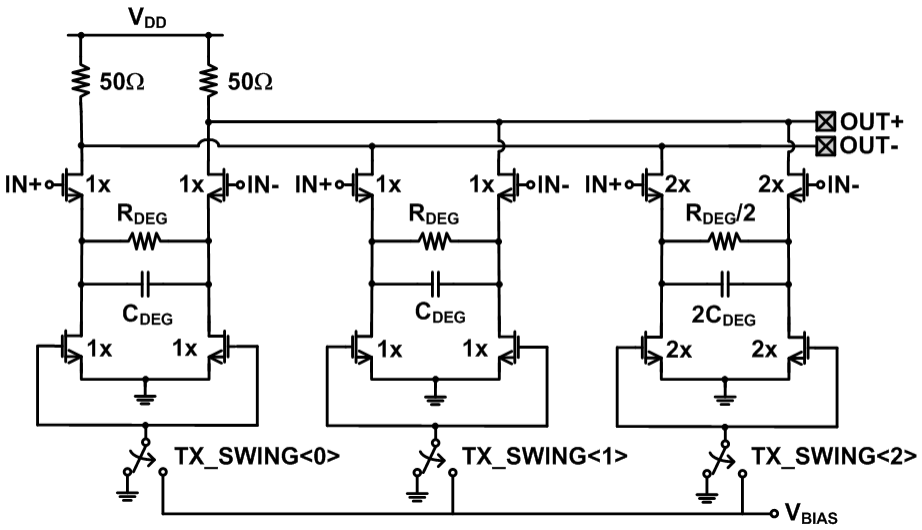


Figure 7.6. 100 Ω differential transmit driver with variable signal swings and RC source degeneration [32]. TX_SWING<2> adds 0.5 V_{ppd} to the transmit swing while TX_SWING<1:0> each add 0.25 V_{ppd} to the transmit swing.

7.2.3.2 Receive Amplifier

Due to pre-emphasis and channel loss, the signal at the input of the receiver is attenuated by one order of magnitude from the peak signal swing. The three-stage variable gain amplifier [32] shown in Figure 7.7 amplifies the signal swing before slicing. Digitally controlled source degeneration is used to control the low-frequency gain and an active inductor load in the last stage increases the bandwidth. The designed-for gain settings are 11 dB (low gain), 15.5 dB

(medium gain), and 21 dB (high gain). To mitigate DC offset within the amplifier, DC offset cancellation is implemented within the first stage.

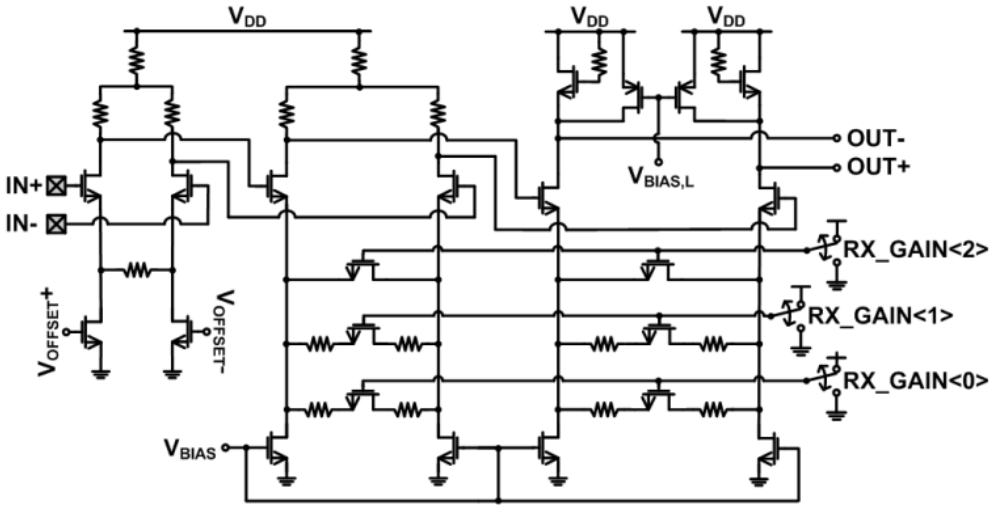


Figure 7.7. Variable gain receive amplifier [32]. RX_GAIN<2> automatically sets the receiver to high gain. Enabling only one of RX_GAIN<1:0> implements low gain, while enabling both RX_GAIN<1:0> implements medium gain.

7.2.3.3 Forward Error Correction Codec

The encoder implements five different BCH block codes: (64, 36, 5), (63, 39, 4), (63, 45, 3), (63, 51, 2), (63, 57, 1); where *n* is the codeword length, *k* is the data bits per codeword, and *t* is the error correction capability. The code rate (*CR*) of a code is defined as the data bits per codeword divided by the codeword length ($CR = k/n$). The number of information bits per code word ranges from 57 to 36, resulting in $0.57 \leq CR \leq 1$. Decreasing the CR reduces the information rate of the link. For example, it is lowered from 4 Gb/s to 3.62 Gb/s for a $CR = 0.9$ and to 2.29 Gb/s for the $CR = 0.57$. The five block codes have error correction capabilities that range from 1 to 5 bits. Systematic BCH encoding has the favorable property that the information bits are unchanged in the encoded bit-stream. Therefore, the encoder can be implemented using simple bit shifts and XOR operations. A low-power decoder is implemented on the receive side. The decoder architecture [18] is shown in Figure 7.8, where a simple error detector unit, similar to the encoder, gates the high-powered error correction unit. Due to the interleaving factor of 16, the encoder and decoder introduce a total end-to-end latency of approximately 2 kilobits.

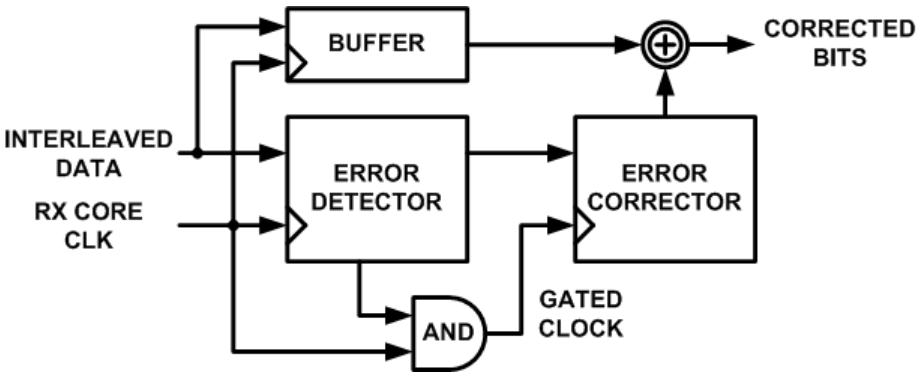
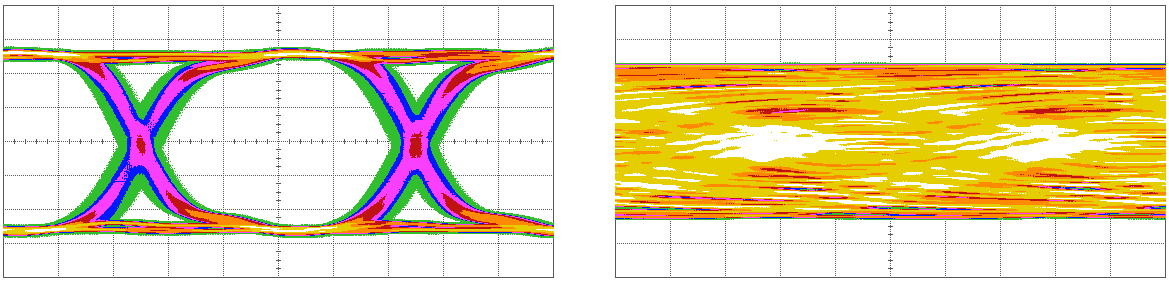


Figure 7.8. Decoder architecture.

7.3 High-Loss Channel Measurement Results

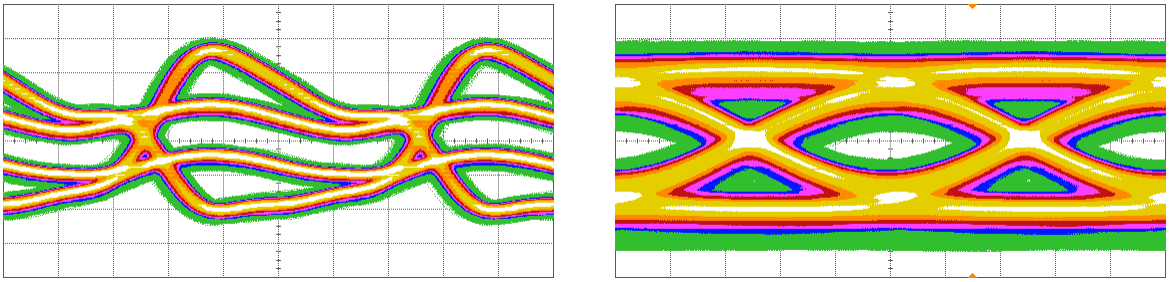
7.3.1 Eye Diagrams

To demonstrate the effect of the high-loss channel, Figure 7.1, on the signal integrity, an Altera high-speed signal integrity development board was used to generate a 4 Gb/s, 0.5 V_{ppd} PRBS signal without pre-emphasis, which was then passed through the test channel. The eye diagrams at both ends of the channel are shown in Figure 7.9; the channel is seen to cause a significant degree of eye closure (Figure 7.9(b)). The test chip transceiver is operated over this same channel. The eye diagrams measured at the test chip transmitter output and receiver input are shown in Figure 7.10; in this case, the transceiver was operated at a 4 Gb/s line rate with 0.5 V_{ppd} signal swing. A more open eye can potentially be received by test chip transceiver, as pre-emphasis was implemented at the transmitter, so as to shape the response by boosting the frequency content. A comparison of the eye diagrams shown in Figure 7.9(b) and 7.10(b) demonstrates the drastic signal integrity improvement achieved through the use of pre-emphasis. Transmit-side pre-emphasis has effectively generated a small eye opening at the far-end of the channel in exchange for a reduction in the signal swing. Of course, the signal integrity will be further degraded as it travels to the slicer.



(a) (b)

Figure 7.9: (a) Pre- and (b) post- high-loss test channel eye diagrams for a 4 Gb/s, PRBS, 0.5 V_{ppd} signal generated by an Altera high-speed signal integrity development board. [Horizontal scale: 50 ps/div.; vertical scale: 100 mV/div.].



(a) (b)

Figure 7.10: (a) Pre- and (b) post- high-loss test channel eye diagrams for a 4 Gb/s line rate with 0.5 V_{ppd} signal swing. Signals are generated by the test chip transceiver. The addition of pre-emphasis generates a small eye opening at the end of the channel. [Horizontal scale: 50 ps/div.; vertical scale: 75 mV/div. for (a) and 20 mV/div. for (b)].

Additionally, the eye diagram at the output of the transmitter (Figure 7.10(b)) shows a 4 GHz tone which is most easily identifiable by observing the variation on the logic high and logic low levels. The 4 GHz tone is attributable to coupling that occurs on the test board from the transmit clock transmission line to the transmitted data signal lines, even if the transmitter is unpowered. Due to the proximity of the transmit clock transmission line to the transmitted data signal lines, highlighted in Figure 7.11, the clock is coupled stronger to one line of the differential data line than the other, imparting a differential sinusoid on the data signal. However, the high loss of the test channel at 4 GHz (38.6 dB) significantly attenuates the noise due to clock coupling, so this does not put a limit on the operation of the data link.

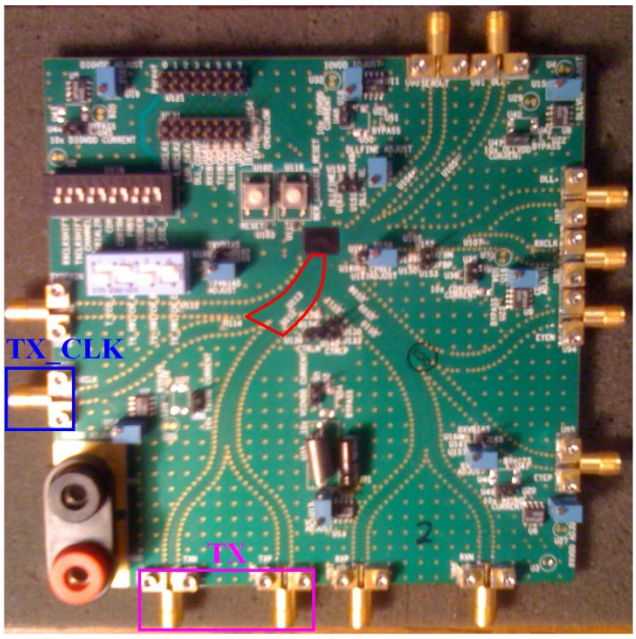


Figure 7.11: PCB test board with TX_CLK SMA connector (blue), TX SMA connector (purple), and area where the TX_CLK and TX traces are adjacent (red) highlighted.

7.3.2 BER Performance

7.3.2.1 Constant Line Rate BER Performance

To demonstrate the effect of coding at a given line rate, the transceiver was tested at a line rate of 4 Gb/s for two transceiver settings: (a) 0.25 V_{ppd} transmit swing with high receiver gain (low-swing setting), and (b) 0.5 V_{ppd} transmit swing with low receiver gain (high-swing setting). The BER bathtub curve for the low-swing and high-swing setting are shown in Figure 7.12 and Figure 7.13, respectively. The data demonstrates how the *CR* can be decreased to 0.71 or below to permit reliable transmission at a signal swing of 0.25 V_{ppd}. If the $CR = 0.81$ is desired, the data demonstrates how the signal swing can impact the BER performance requiring an increase in the signal swing from 0.25 V_{ppd} to 0.5 V_{ppd} to achieve a BER of 10^{-12}. Finally, for both transceiver settings, $CR = 1$ (uncoded) and $CR = 0.9$ resulted in an unreliable link (BER < 10^{-7} for all sampling locations). Therefore, those datasets were not shown.

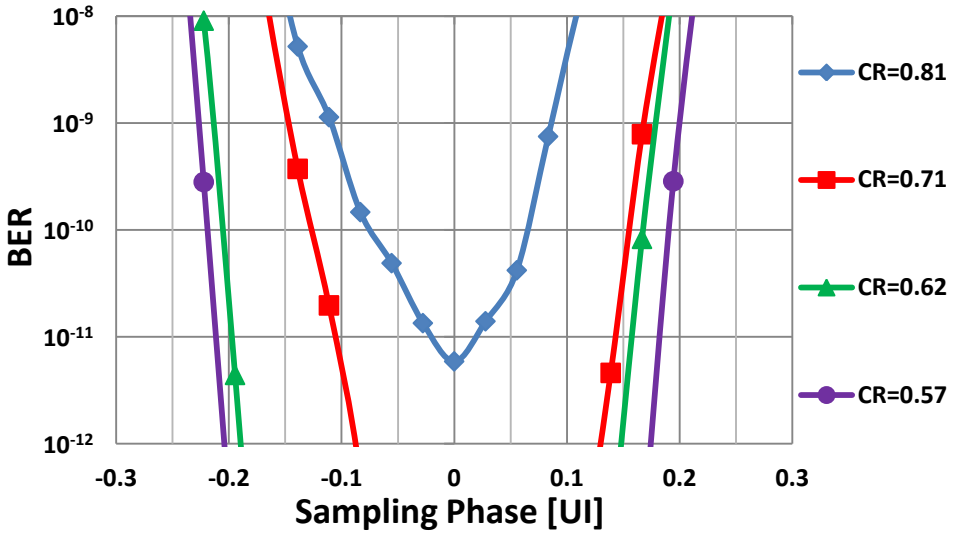


Figure 7.12: BER measurement results for 0.25 V_{ppd} transmit swing with high receiver gain at a line rate of 4 Gb/s.

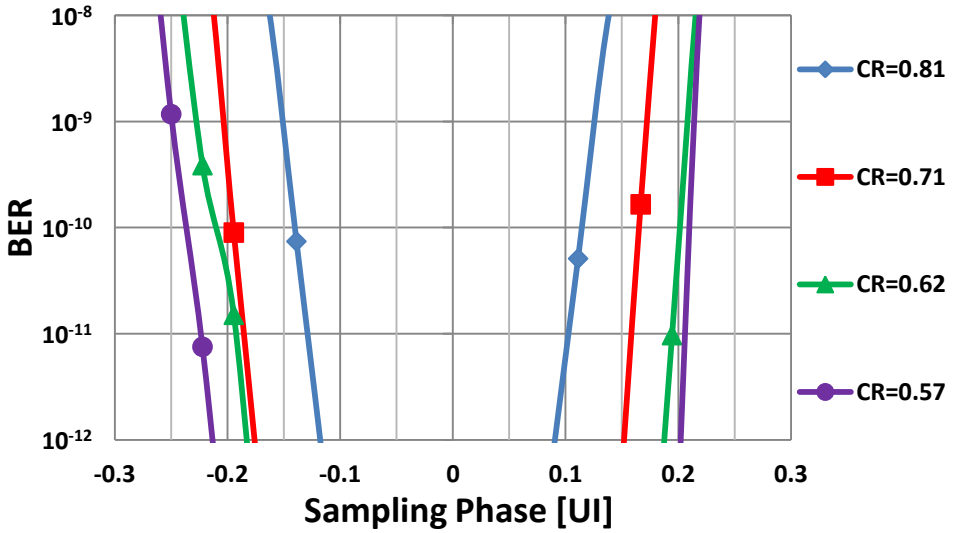


Figure 7.13: BER measurement results for 0.5 V_{ppd} transmit swing with low receiver gain at a line rate of 4 Gb/s.

7.3.2.2 Constant Information Rate BER Performance

To demonstrate how coding affects the performance of a link operating at a fixed information rate, the transceiver was tested at an information rate of 3.238 Gb/s. To do so, the link was first operated at a line rate of 3.238 Gb/s without ECC. Next, it was operated at a line rate of 4 Gb/s using a $CR = 0.81$. The measurement results for two link configurations are shown in Figure 7.14 for multiple transceiver swing settings: 0.25 V_{ppd}, 0.5 V_{ppd}, and 0.75 V_{ppd}. The results show that at low signal swings, coding improves the BER performance of the link, even

though the line rate is increased. For a 0.25 V_{ppd} swing, the minimum BER is reduced by a factor of 45. For a specific BER requirement, ECC can also be utilized to reduce the required signal swing. If we define a reliable link as one in which the BER is below 10^{-12}, the addition of ECC reduces the transmit swing requirement from 0.75 V_{ppd} to 0.5 V_{ppd}.

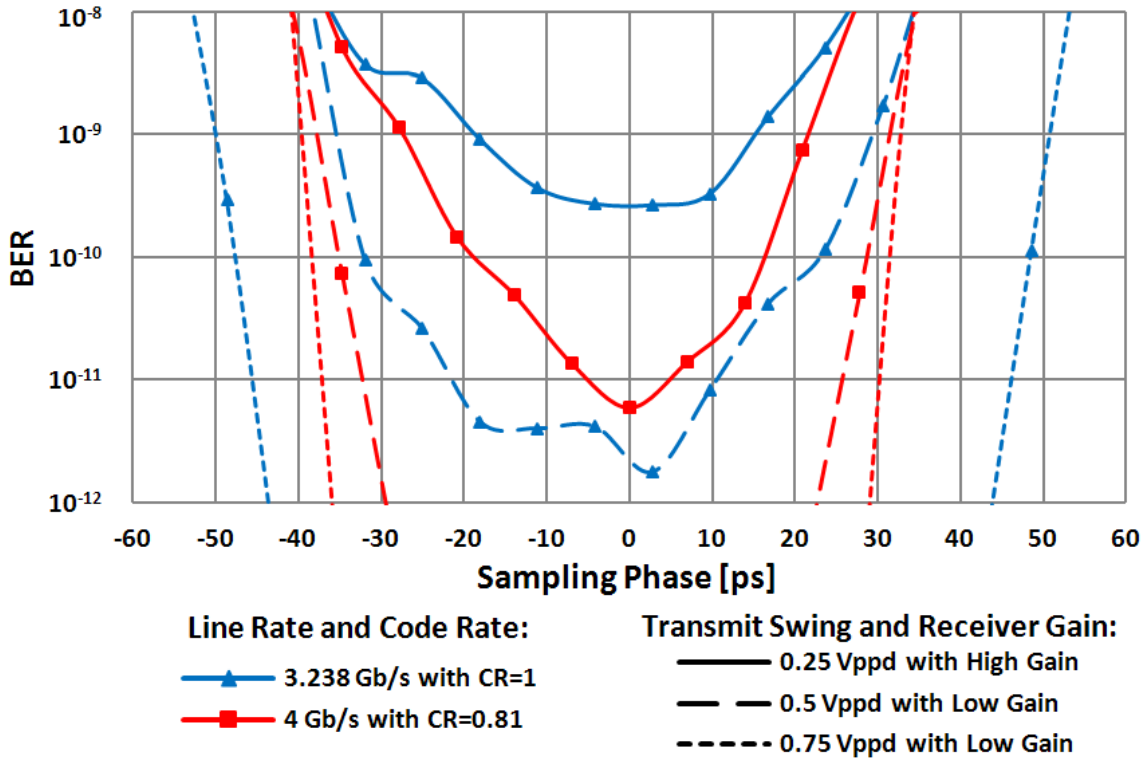


Figure 7.14: BER measurement results at a constant information rate of 3.238 Gb/s. Transceiver is configured with a 3.238 Gb/s line rate without coding (*CR* = 1) or 4 Gb/s line rate with *CR* of 0.81.

7.3.3 Jitter Tolerance

7.3.3.1 Constant Line Rate Jitter Tolerance

The data from Figures 7.12 and 7.13 demonstrates how increasing the *CR* can improve the allowable increase in peak-to-peak jitter in the link, as shown in Figures 7.15 and 7.16 for BER of 10^{-11} and 10^{-12}, respectively. The results also show how decreasing the *CR* is effective at increasing the allowable peak-to-peak jitter at low transmit signal swings and high code rates. For a 0.25 V_{ppd} and 0.5 V_{ppd} signal swing at a BER of 10^{-12}, decreasing *CR* from 0.81 to 0.71 increases the peak-to-peak jitter tolerance from 7 ps to 63 ps and from 55.6 ps to 83.3 ps,

respectively. The results also demonstrate how, for low *CR*, the utilization of coding negates the need to implement large swings, as the increase in allowable peak-to-peak clock jitter is very small. For a *CR* of 0.57 at a BER of 10^{-12}, doubling the signal swing from 0.25 V_{ppd} to 0.5 V_{ppd} only increases the jitter tolerance by 7%, from 104 ps to 111 ps.

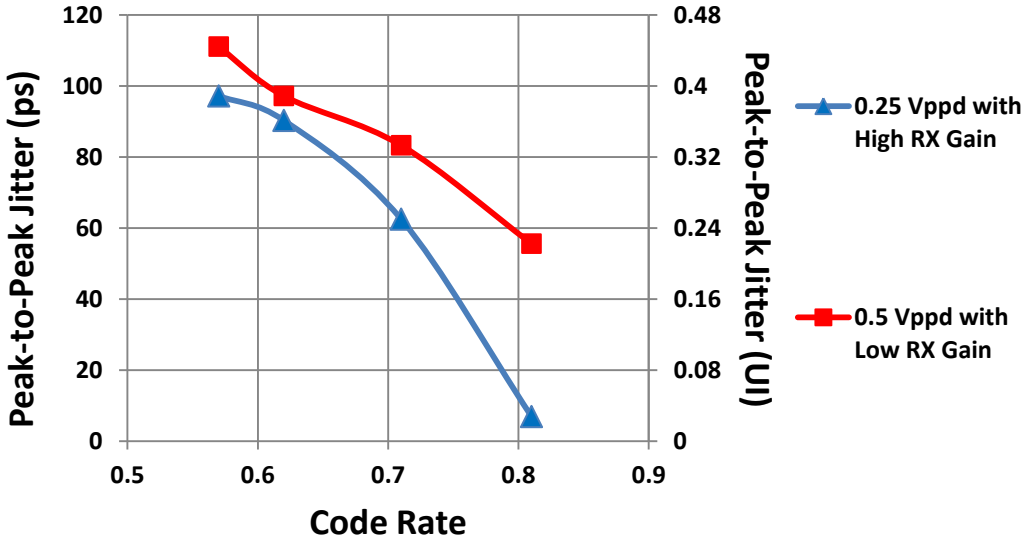


Figure 7.15: Allowable increase in peak-to-peak sampling clock jitter for a BER of 10^{-11} on the 4 Gb/s line rate transceiver.

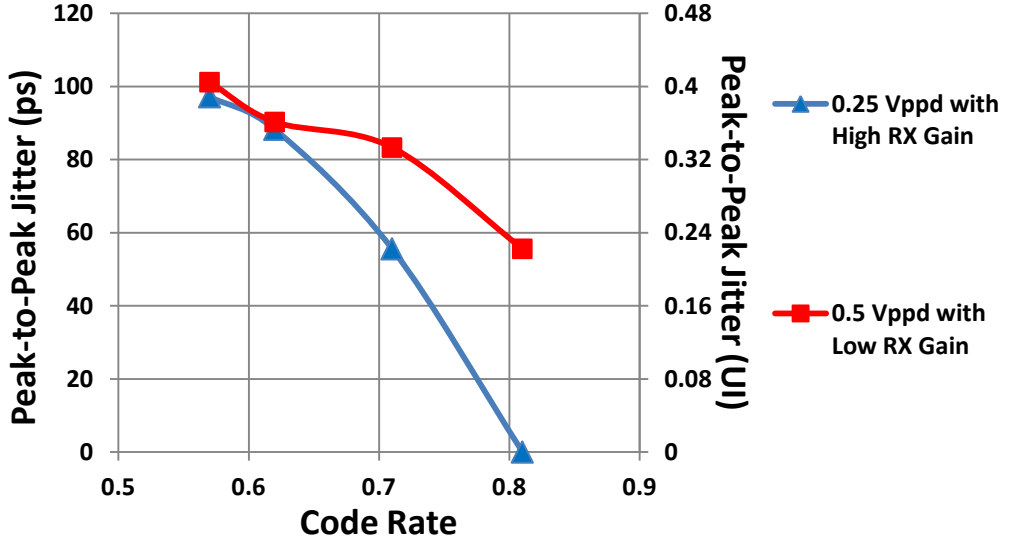


Figure 7.16: Allowable increase in peak-to-peak sampling clock jitter for a BER of 10^{-12} on the 4 Gb/s line rate transceiver.

7.3.3.2 Constant Information Rate Jitter Tolerance

ECC can also be utilized to improve the allowable increase in peak-to-peak sampling clock jitter for a constant information rate. The allowable increase in peak-to-peak sampling clock jitter permits a performance reduction in the clock circuitry, allowing for reduced clock circuitry power consumption. The allowable increase in peak-to-peak clock jitter, extracted from the data shown in Figure 7.14, is shown in Figure 7.17 and 7.18 for BERs of 10^{-11} and $10^{-12,}$ respectively. The results shows that, at low transmit swings, FEC allows the link to maintain a specified BER in the presence of higher peak-to-peak sampling clock jitter than in an uncoded link, for a fixed information rate. This is because the link is increasingly noise-limited at low transmit swings, which is where FEC performs best. At high transmit swings, the jitter performance is dictated by the ISI-induced zero-crossings. For the same information rate, the coded link has a higher line rate, and hence higher ISI, than the uncoded link. Therefore, for sufficiently high swings, the uncoded link's jitter performance will be superior to that of the coded link.

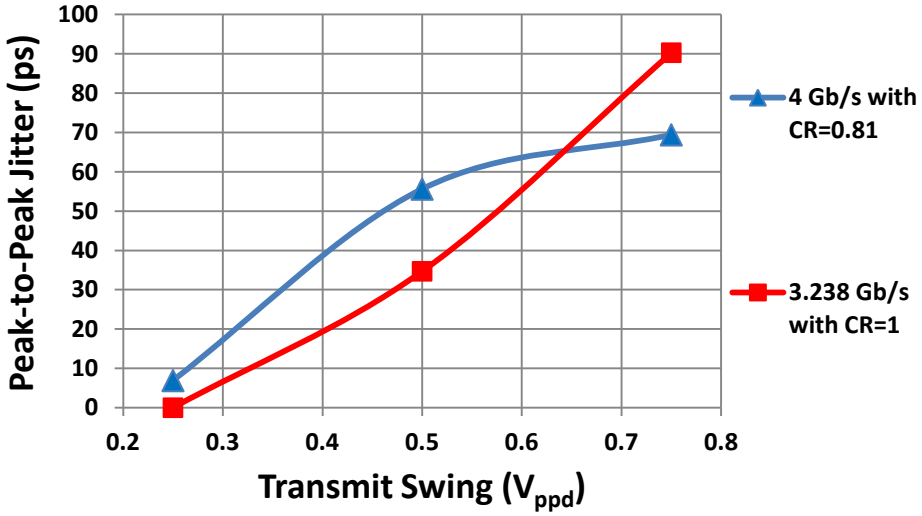


Figure 7.17: Allowable increase in peak-to-peak sampling clock jitter for a BER of 10^{-11} on the 3.238 Gb/s information rate transceiver. Transmit swings with 0 ps of allowable increase in peak-to-peak sampling clock jitter do not achieve a BER of 10^{-11} for any sampling location.

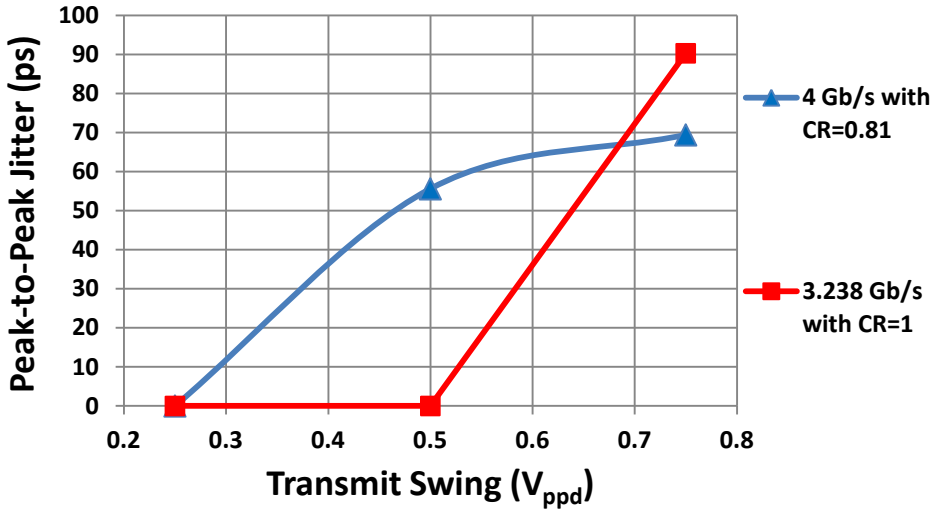


Figure 7.18: Allowable increase in peak-to-peak sampling clock jitter for a BER of 10^{-12} on the 3.238 Gb/s information rate transceiver. Transmit swings with 0 ps of allowable increase in peak-to-peak sampling clock jitter do not achieve a BER of 10^{-12} for any sampling location.

7.3.4 Power and Energy Consumption

7.3.4.1 Codec Power Consumption

The addition of coding requires additional power consumption within the transmitter and receiver, due to the inclusion of an encoder and decoder. The power consumption overhead of the encoder and decoder as a function of *CR* is shown in Figure 7.19. To generate the data, the transmitter and receiver are connected in a loopback configuration using SMA cables; this facilitates power measurement in the event of extremely low pre-FEC bit error rates. For extremely low pre-FEC BERs, few errors are present after slicing, and the decoder error correction unit, which is clock gated, is rarely operated and consumes insignificant power. Therefore, the power consumption shown in Figure 7.19 is essentially the power consumed by the encoder and decoder error detector units. The data show that the power dissipated by the encoder and error detector ranges from 1.5 mW/Gb/s to 5.4 mW/Gb/s, depending on the *CR*. This power is expected to scale with process node.

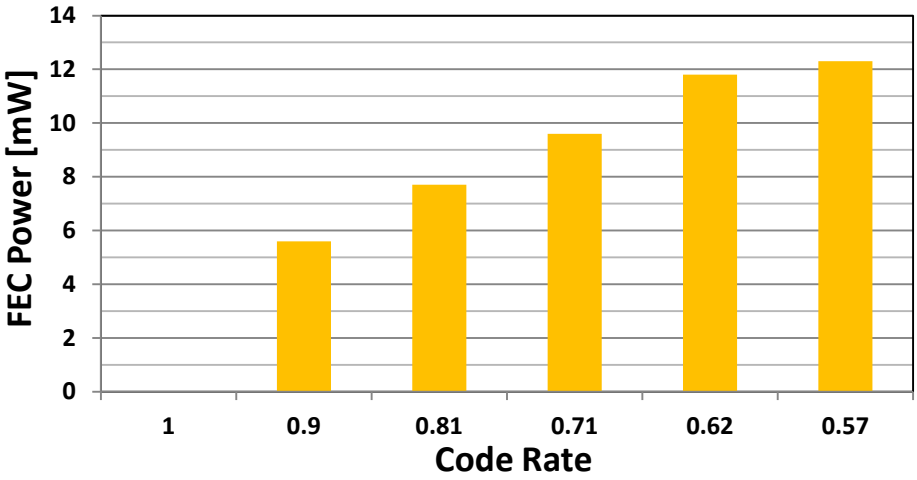


Figure 7.19: Codec power consumption for a line rate of 4 Gb/s as a function of code rate (*CR*). The codec power is measured at an extremely low pre-FEC BER.

In the event of a high pre-FEC BER, the error correction unit is operated to improve the BER of the link at the expense of additional power consumption. The power consumption of the FEC for three pre-FEC BERs is shown in Figure 7.20; three transceiver configurations were used: cable loopback with 0.5 V_{ppd} transmit swing and low gain, the test channel with 0.25 V_{ppd} transmit swing and high receiver gain, and the test channel with 0.5 V_{ppd} transmit swing and low receiver gain. The data points were obtained using a sampling phase that is estimated to produce a BER of at least 10^{-11} post-FEC. Referring to the data in Figures 7.12 and 7.13, one concludes that the pre-FEC BER is smallest for cable loopback and greatest for the test channel with 0.25 V_{ppd} swing. The power measurement results thus indicate that the power consumed by the error correction unit is correlated to the pre-FEC BER, as expected. The power required to correct the errors and drastically reduce the BER is small, on the order of 1 mW. Therefore, the main power overhead of FEC is attributed to the power consumption of the encoder and error detector units, as shown in Figure 7.19.

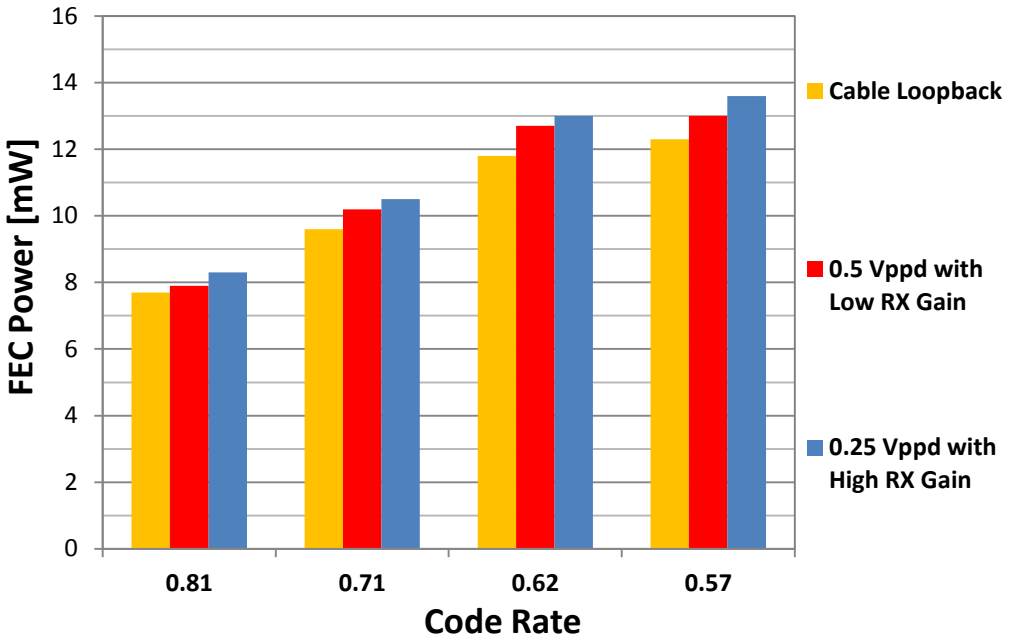


Figure 7.20: Codec power consumption for a link rate of 4 Gb/s as a function of the code rate (CR). The transceiver settings for 0.25 V_{ppd} with high receive (RX) gain and 0.5 V_{ppd} with low receive gain are measured utilizing the test channel at the estimated minimum BER sampling location. The post-FEC BER of each sampling location is below 10^{-9}.

7.3.4.2 Transmitter Power Consumption

In Figure 7.21, the power consumption of the transmit driver is plotted as a function of the differential transmit swing. The linear relationship between the power and transmit swing is a property of the current mode nature of the driver and places an upper bound on the FEC power overhead. For the transmitter on the test chip, the power required to change the signal swing by 0.25 V_{ppd} is approximately 11 mW and is larger than the power consumption required to implement FEC at all but the lowest code rates.

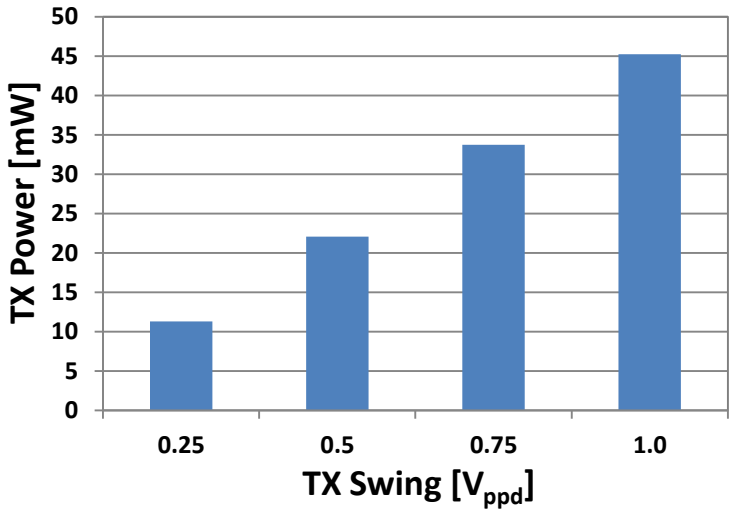


Figure 7.21: Transmit driver power consumption as a function of signal swing.

7.3.4.3 Constant Line Rate Inner Transceiver and Codec Power and Energy

The power dissipation of the inner transceiver (transmit driver and receive amplifier) and codec is shown in Figure 7.22 as a function of CR for BER < 10^{-12}. These results show that a decrease in the CR allows for power reduction by reducing the transmit signal swing, but only up to a point. Once the minimum transmit swing is achieved, further reduction of the CR results in a slight increase in the power consumption. This is because the codec power increases from 7.9 mW to 13.6 mW as CR is reduced from 0.81 to 0.57.

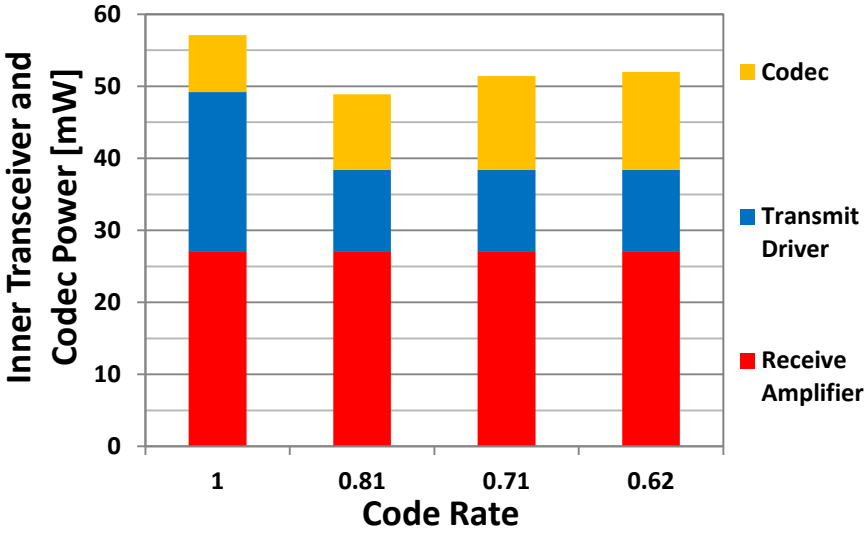


Figure 7.22: Inner transceiver and codec power as a function of CR for a reliable (BER < 10^{-12}) 4 Gb/s line rate transceiver.

In Figure 7.23, the power consumption of the 4 Gb/s line rate transceivers from Figure 7.22 are normalized by the information rate to demonstrate the energy per information bit, E_b. The energy per bit demonstrates the balance between power reduction in the transmitter and decrease in information rate. For high code rates, there is a slight reduction in the power per information rate as the code is increased. This is a result of the significant power savings in the transmit driver slightly overcoming the reduction in information rate and power overhead of the codec. On the other hand, at low code rates, the lack of power reduction in the transmitter is overwhelmed by the reduction in information rate, resulting in an increase in the power per information rate. The results show that FEC with high code rates can reduce the power consumption of the link per information rate, but that drastically reducing the information rate with a low *CR* can increase the power consumption of the link per information rate. By reducing the CR from 0.81 to 0.71, the energy per bit is reduced by 0.5 pJ while reducing the CR from 0.71 to 0.62 increases the energy per bit by 3.6 pJ.

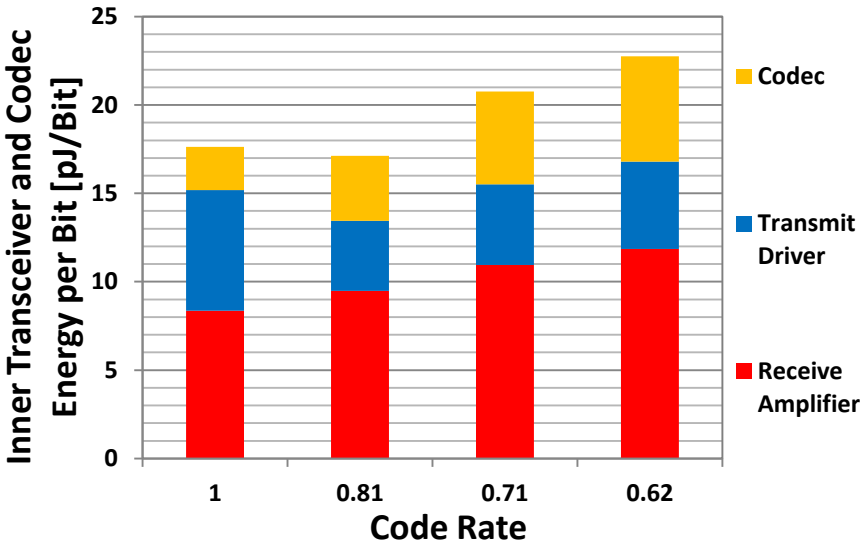


Figure 7.23: Inner transceiver and codec energy per bit E_b as a function of *CR* for a reliable (BER < 10^{-12}) 4 Gb/s line rate transceiver.

Coded links with a fixed line rate will have an information rate that varies with *CR*. Though Figure 7.23 captures the effect of this reduced information rate in computing energy per bit, it is of interest to normalize this energy with respect to the channel loss. For a fair comparison with uncoded systems, we normalize E_b with the channel loss at the Nyquist frequency determined by the information rate, 0.5 times the information rate, as shown in Figure 7.24. The results in

Figure 7.24 lead to a different conclusion than in Figure 7.23, namely, that decreasing the *CR* decreases the performance of the link in terms of energy per bit per dB channel loss. The energy per unit loss data suggests that only the highest code rates should be utilized since the penalty for reducing the *CR* is small at high code rates and increases drastically at low code rates. The results of Figures 7.22 through 7.24 all demonstrate that low code rate FEC should not be used on the test channel.

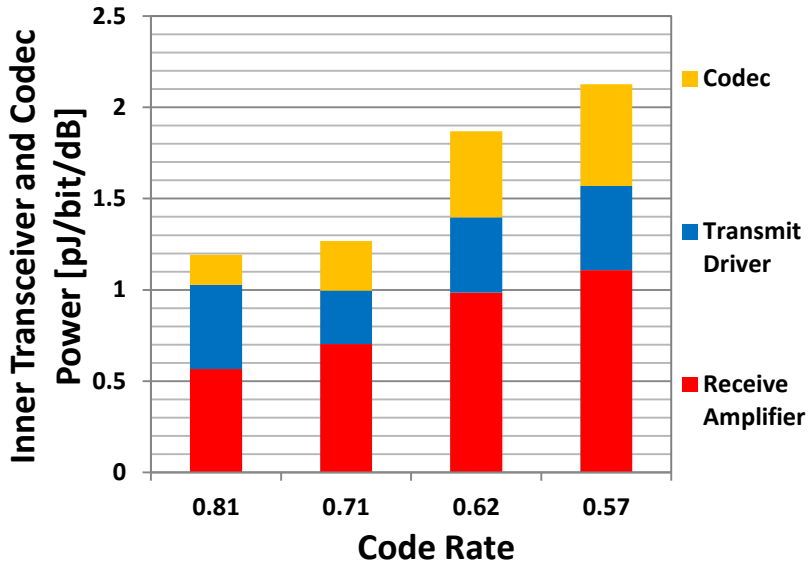


Figure 7.24: Inner transceiver and coded power per information rate per dB of channel loss as a function of *CR* for a reliable (BER < 10^{-12}) 4 Gb/s line rate transceiver.

7.3.4.4 Constant Information Rate Inner Transceiver Power and Energy

Figure 7.25 shows the inner transceiver and codec energy per information bit, E_b, for two different information rates, 3.238 Gb/s and 2.857 Gb/s, for both the coded and uncoded links. Figure 7.25 indicates that coding reduces E_b, provided the information rate is high enough while the CR is low enough to achieve a reliable BER. In these cases, the effective SNR improvement due to FEC combined with the reduced transmit swing and driver power compensate for the increased ISI (in terms of BER) and the additional power overhead due to the codec. At an information rate of 3.238 Gb/s, the coded link with CR = 0.81 reduces E_b by 1.1 pJ (equivalent to 1.1 mW/Gb/s), while at an information rate of 2.857 Gb/s, the coded link with CR = 0.71 increases E_b by 3.7 pJ. Therefore, there exists a minimum information rate, IR_0, above which the introduction of FEC will provide inner transceiver energy savings. The results of Figure 7.25

demonstrate that IR_0 is somewhere between 2.857 Gb/s and 3.238 Gb/s for the high-loss channel of Figure 7.1. If additional information rates between 2.857 Gb/s and 3.238 Gb/s were measured, then the exact location of IR_0 could be discerned.

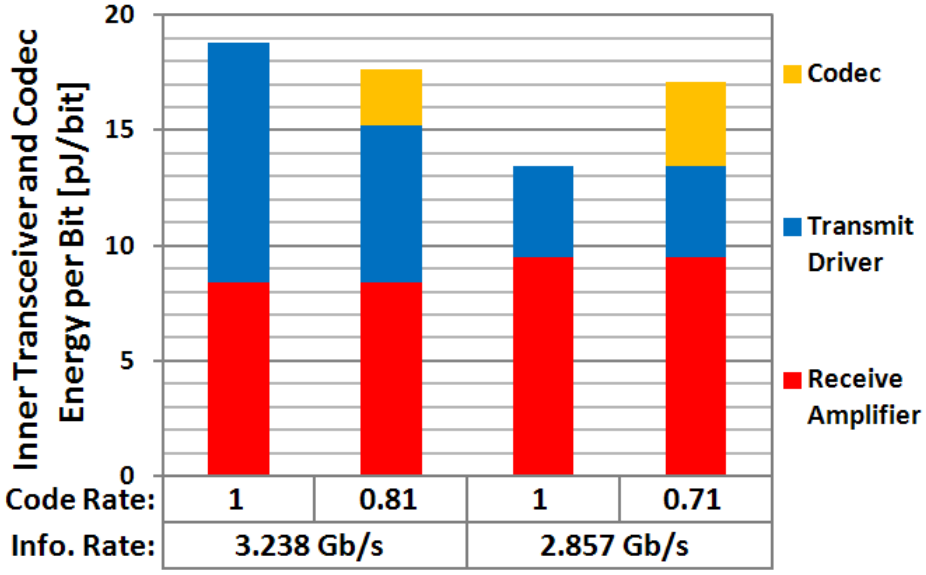


Figure 7.25: Inner transceiver and codec energy per information bit for reliable (BER < 10^{-12}) transceivers at two information rates of 3.238 Gb/s and 2.857 Gb/s.

Since all of the digital components utilize the same power supply and, with the exception of the codec, cannot be individually disabled, the digital power consumption of the data path cannot be measured. Therefore, it is important to note that the power penalty due to the increase in line rate in the digital data path (excluding the codec) has not been included in the analysis. Inclusion of power and energy increases in the serializer, deserializer, and sampler will increase the power and energy consumption of the complete link. The increase will result in a minimum information rate above IR_0 where the introduction of FEC to the complete link will begin to provide energy savings.

7.4 Sub-Nyquist Notch Channel Measurement Results

7.4.1 Eye Diagrams

To demonstrate the effect of the sub-Nyquist notch channel, Figure 7.2, on the signal integrity, the post-channel differential eye diagrams for a 0.5 V_{ppd} PRBS signal generated by the

transceiver IC are shown in Figure 7.26. Due to the presence of the notch, line rates with significant signal frequency components in the vicinity of 1.58 GHz will see increased degradation. Therefore, as the line rate is increased toward 3.16 Gb/s, the signal integrity quickly degrades. For line rates above 3.16 Gb/s, the signal integrity improves as the line rate is increased toward the maximum operating frequency of the transceiver, 4 Gb/s.

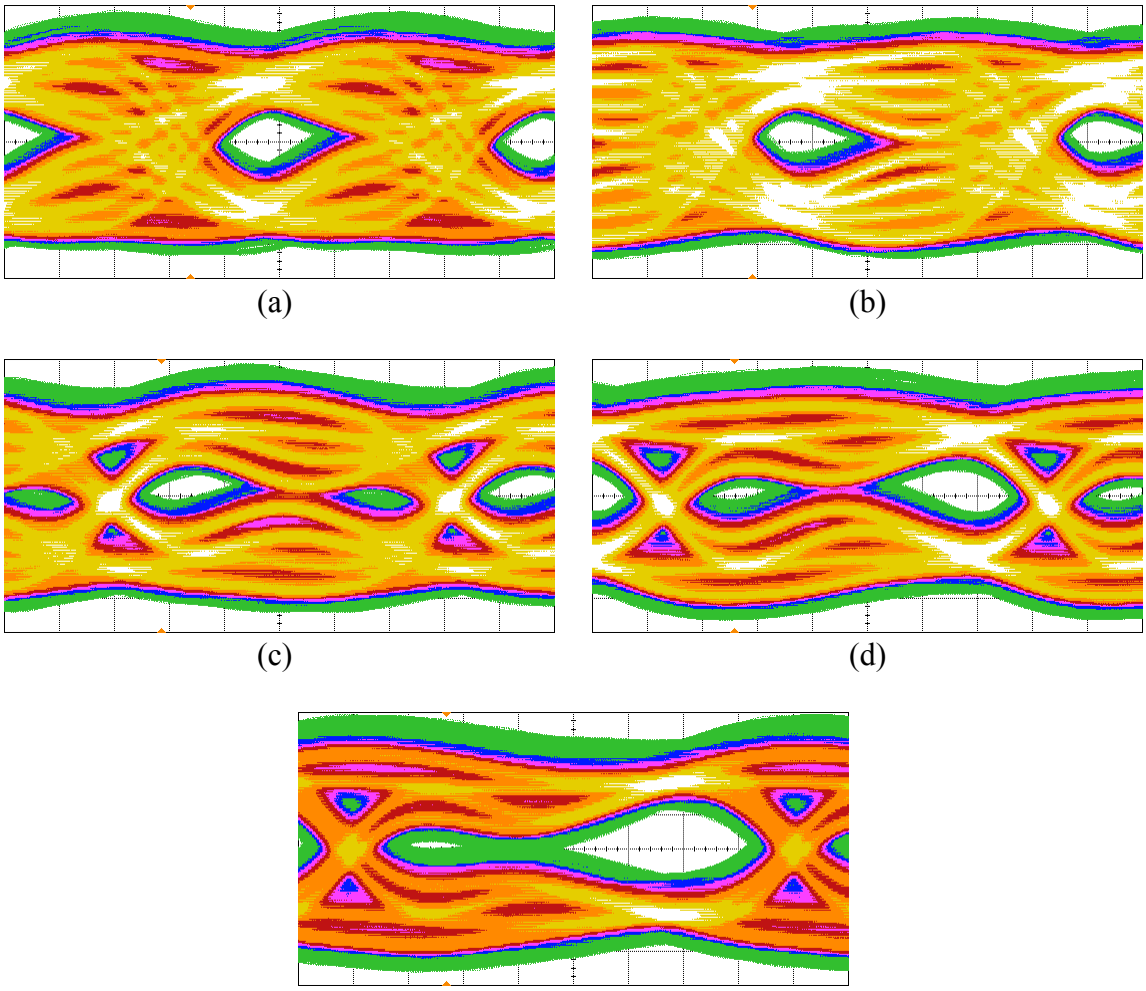


Figure 7.26: Eye diagrams at the output of the sub-Nyquist notch test channel for (a) 4 Gb/s, (b) 3.619 Gb/s, (c) 3.238 Gb/s, (d) 2.857 Gb/s, and (e) 2.476 Gb/s. Signal swing is 0.5 V_{ppd}. Signals are generated by the test chip transceiver. [Horizontal scale: 50 ps/div.; vertical scale: 25 mV/div.].

7.4.2 BER Performance

In Figure 7.27, the maximum achievable information rate (BER < 10^{-12}) is shown as a function of the line rate. At low line rates, the transceiver can be operated reliably without

coding. When the Nyquist frequency approaches the notch minimum frequency, the maximum achievable data rate drops dramatically, requiring the introduction of FEC (CR = 0.71) to achieve BER < 10^{-12}. As the line rate is increased further, the maximum CR which achieves the desire BER increases, resulting in a rapid increase in the maximum achievable information rate, surpassing the maximum achievable information rate of uncoded links. Therefore, the introduction of coding across the sub-Nyquist notch test channel permits a 27% increase in the information rate of the link, from 2.857 Gb/s to 3.619 Gb/s.

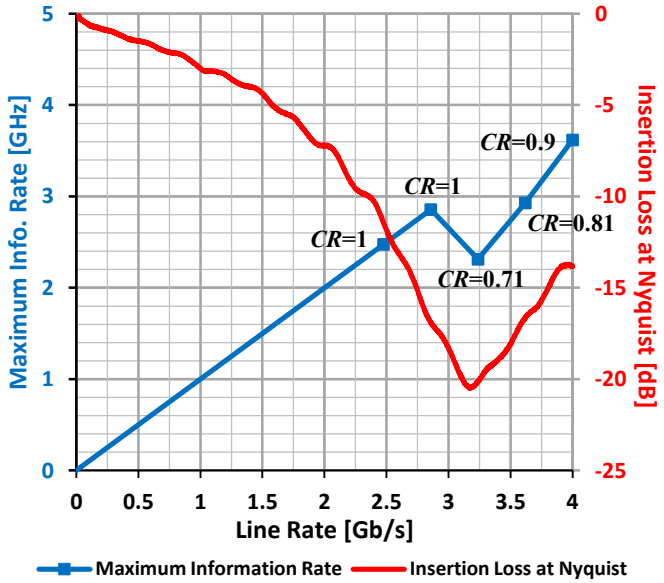


Figure 7.27: Maximum information rate (BER < 10^{-12}) and insertion loss at the Nyquist frequency as a function of line rate. For all measured maximum information rates, the corresponding CR is labeled. For line rates with Nyquist frequency above the notch minimum, 1.58 GHz, uncoded links are not reliable. The introduction of coding permits an increase in the achievable information rate.

The ability to operate reliably at a higher CR implies that the pre-FEC BER is lower at line rate of 4 Gb/s than at line rates of 3.238 and 3.619 Gb/s. As such, if coding is implemented, it is more efficient to increase the line rate to 4 Gb/s than to operate at 3.238 or 3.619 Gb/s. In Table 7.2, the minimum-power transceiver configurations as a function of CR which achieve a BER < 10^{-12} for a 4 Gb/s line rate are shown. As the CR is decreased, the required transmit swing is reduced from 0.75 V_{ppd} to the minimum transmit swing of 0.25 V_{ppd}. The table also provides BER information for the uncoded transceiver at the same information rates. For CR of 0.9 and 0.81, the corresponding uncoded transceivers are unreliable. For CR of 0.71 and 0.62,

corresponding uncoded transceivers achieve a BER < 10^{-12} given the minimum power configurations listed in the table.

Table 7.2: BER Measurement Results for Uncoded and 4 Gb/s Line Rate Coded Transceiver Configurations. Minimum Transmit Swing and Receive Gain Shown for Reliable (BER < 10^{-12}) Configurations.

INFORMATION RATE	UNCODED TRANSCEIVER				CODED TRANSCEIVER				
	LINE RATE	BER	TX SWING	RX GAIN	*CR*	LINE RATE	BER	TX SWING	RX GAIN
4.0 Gb/s	4.0 Gb/s	> 10^{-7}	N/A	N/A	N/A	N/A	N/A	N/A	N/A
3.619 Gb/s	3.619 Gb/s	> 10^{-7}	N/A	N/A	0.9	4.0 Gb/s	< 10^{-12}	0.75 V_{ppd}	Medium
3.238 Gb/s	3.238 Gb/s	> 10^{-7}	N/A	N/A	0.81	4.0 Gb/s	< 10^{-12}	0.5 V_{ppd}	Medium
2.857 Gb/s	2.857 Gb/s	< 10^{-12}	0.75 V_{ppd}	Low	0.71	4.0 Gb/s	< 10^{-12}	0.25 V_{ppd}	High
2.476 Gb/s	2.476 Gb/s	< 10^{-12}	0.25 V_{ppd}	Low	0.62	4.0 Gb/s	< 10^{-12}	0.25 V_{ppd}	Medium

7.4.3 Power and Energy Consumption

7.4.3.1 Codec Power Consumption

In Figure 7.28, the power consumption of the encoder and decoder as a function of *CR* for three channels is presented. For all three channels, the power consumption is shown for the minimum power, 4 Gb/s transceiver configuration which achieves a BER < 10^{-12}. The three channels are the high-loss channel (see Figure 7.1), the sub-Nyquist notch channel (see Figure 7.2), and an SMA cable loopback from the TX test board connectors to the RX test board connectors. The power consumption of the codec is dependent on the pre-FEC BER because the high powered error corrector operation is gated when no error is detected in the code word. The effectively error-free cable loopback power consumption characterizes the power consumption of the codec with the error corrector disabled. Therefore, the additional codec power consumption for the two test channels, on the order of 1 mW, is the power consumption of the error corrector unit reducing the post-FEC BER below 10^{-12}. No data is presented for the high-loss channel with *CR* = 0.9 since no reliable link configurations exist.

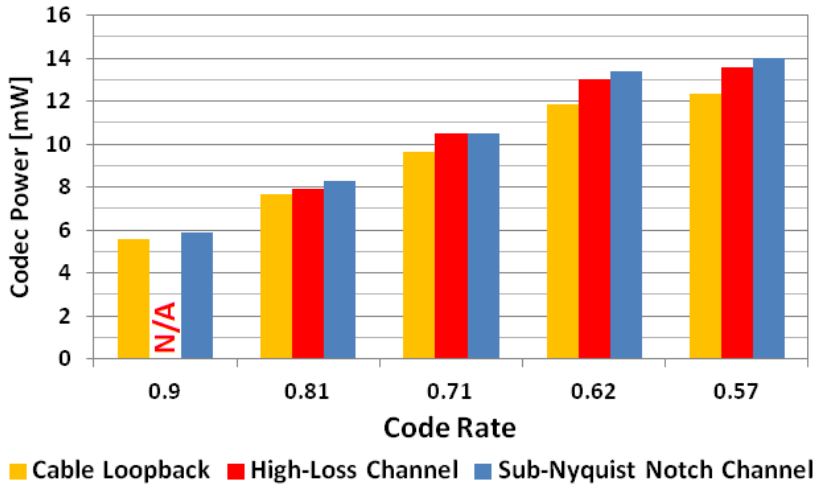


Figure 7.28: Codec power consumption for reliable (BER < 10^{-12}) 4 Gb/s line rate transceiver configurations. The cable loopback setup approximates the power of the codec with the error corrector disabled. The additional power consumption of codec for the high-loss and sub-Nyquist notch channels is a result of enabling the decoder error corrector in order to reduce the BER < 10^{-12}.

7.4.3.2 Constant Line Rate Inner Transceiver Power and Energy

In Figure 7.29, the power dissipation of the inner transceiver (transmit driver and receive amplifier) and codec is shown as a function of CR for BER < 10^{-12}. These results show that a decrease in the CR reduces power through transmit signal swing reduction until the minimum transmit swing is achieved at a CR of 0.71. As the CR is reduced past 0.71, the power consumption increases due to codec power consumption increases

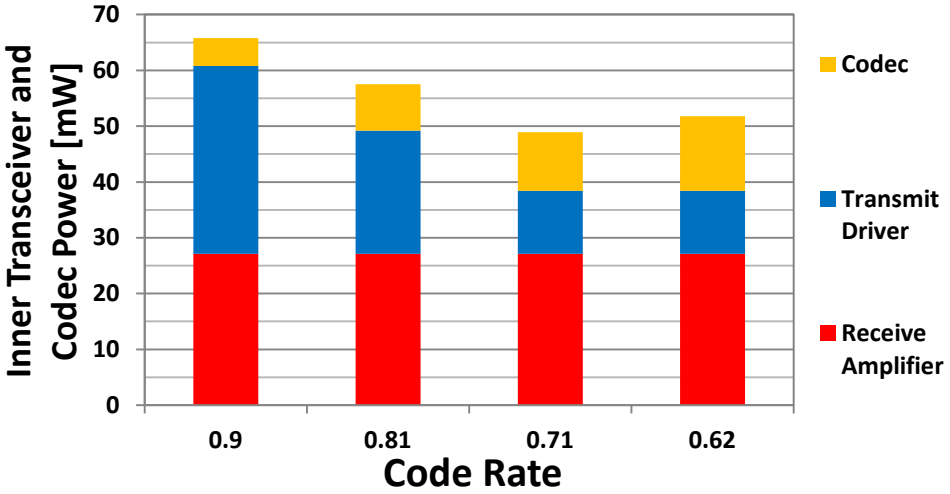


Figure 7.29: Inner transceiver and codec power as a function of CR for a minimum power, reliable (BER < 10^{-12}), 4 Gb/s line rate transceiver.

In Figure 7.30, the energy dissipation per bit, E_b, of the inner transceiver and codec is shown for the minimum power coded transceiver configurations in Table 7.2. The results show that at high *CR*, decreasing the *CR* results in small decreases in energy consumption. The reduction in transmitter power more than compensates for the reduction in information rate. Once the minimum transmit swing is achieved at a *CR* of 0.71, further reduction in *CR* results in a dramatic increase in energy since the information rate is reduced without transmit driver power reduction. Reducing the *CR* from 0.9 to 0.71 reduces the energy per bit by 1.1 pJ, while further reducing the *CR* from 0.71 to 0.62 increases the energy per bit by 3.8 pJ.

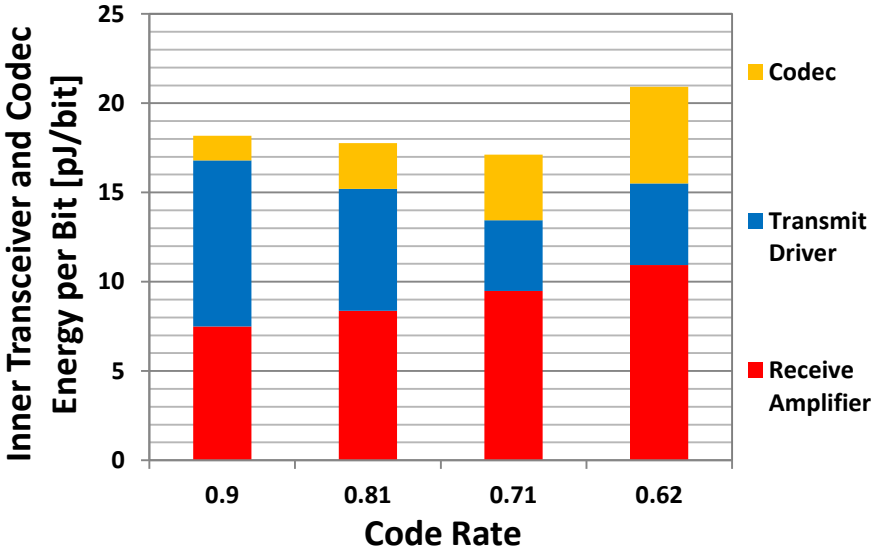


Figure 7.30: Inner transceiver and codec energy per bit E_b as a function of *CR* for a reliable (BER < 10^{-12}) 4 Gb/s line rate transceiver.

7.4.3.3 Constant Information Rate Inner Transceiver Power and Energy

In Figure 7.31, the inner transceiver and codec energy per bit are shown for the reliable uncoded and coded transceivers listed in Table 7.2 with information rates of 2.857 Gb/s and 2.476 Gb/s. At an information rate of 2.476 Gb/s, the uncoded transceiver can achieve a BER < 10^{-12} with the minimum transmit swing. Therefore, the introduction of coding substantially increases the energy consumption 5.4 pJ/bit due to the codec power overhead. On the other hand, at an information rate of 2.857 Gb/s, the introduction of coding reduces the transmit swing requirement from 0.75 V_{ppd} to the minimum swing of 0.25 V_{ppd}. The reduction in transmit driver power is significantly larger than the codec power overhead, resulting in a

substantial 4.2 pJ/bit (20%) decrease in the inner transceiver and codec energy consumption. Therefore, there exists a minimum information rate, IR_0, where the introduction of FEC will provide inner transceiver energy savings. The results of Figure 7.31 demonstrate that IR_0 is somewhere between 2.476 Gb/s and 2.857 Gb/s for the sub-Nyquist notch channel of Figure 7.2. If additional information rates between 2.476 Gb/s and 2.857 Gb/s were measured, then the exact location of IR_0 could be discerned.

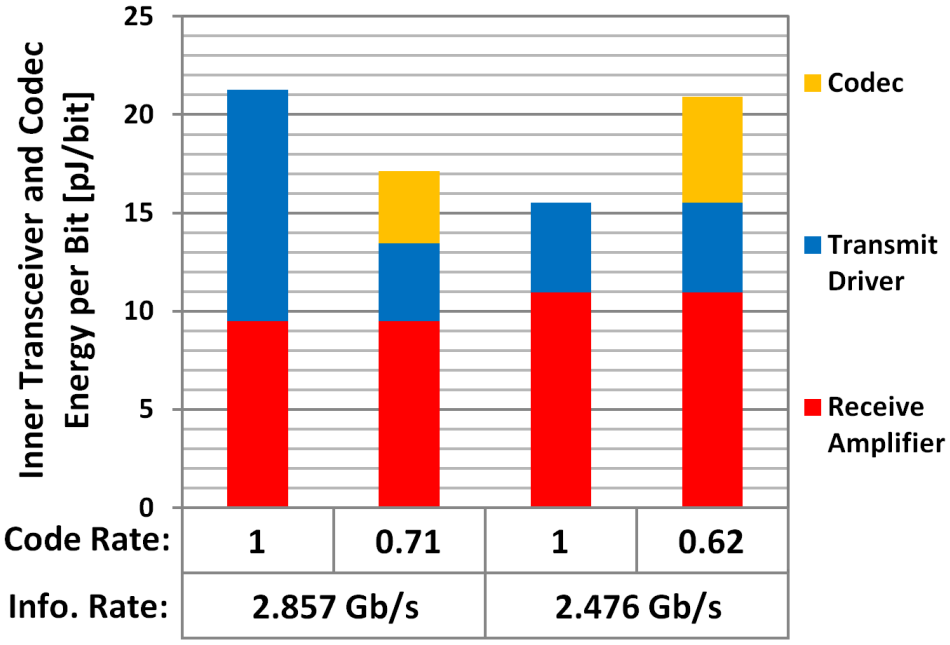


Figure 7.31: Inner transceiver and codec energy per information bit for reliable (BER < 10^{-12}) transceivers for two information rates: 2.857 Gb/s and 2.476 Gb/s.

Since all of the digital components utilize the same power supply and, with the exception of the codec, cannot be individually disabled, the digital power consumption of the data path cannot be measured. Therefore, it is important to note that the power penalty due to the increase in line rate in the digital data path (excluding the codec) has not been included in the analysis. Inclusion of power and energy increases in the serializer, deserializer, and sampler will increase the power and energy consumption of the complete link. The increase will result in a minimum information rate above IR_0 where the introduction of FEC to the complete link will begin to provide energy savings.

7.5 Conclusions

The utilization of FEC within backplane links has been demonstrated to improve reliability, increase jitter tolerance, reduce power consumption, and reduce the design complexity of a transceiver when applied to the appropriate class of high-speed I/O links. The greatest improvement in these metrics is seen at low signal swings (low SNR) and high code rates. At high SNRs, the link is ISI-limited, thereby reducing the benefits of coding. For ISI dominated links, limited pre-emphasis or peaking may be used to partially equalize the channel response, and then FEC can be employed to achieve a reliable BER. This combination moves the response at the receive slicer closer to being noise limited and away from being ISI dominated. Another interesting scenario is with parallel links where crosstalk from adjacent channels can act as random noise, making the study of ECC in parallel links worthwhile. Furthermore, in more advanced process nodes, the power and area overhead of the codec will scale down, thereby increasing the energy benefits of FEC-based links.

In general, random noise of the link will have a significant impact on the BER in an operational window determined by the channel, line rate, noise/crosstalk power, and transmit peaking response. In this window, the use of coding can drastically improve the reliability of the link. For this specific transceiver, it was a challenge to determine the class of channels over which coding shows benefits. This challenge arose from two sources: (1) reduction of the line rate from 6.25 Gb/s to 4 Gb/s, and its interaction with the fixed transmit peaking, which was optimized for 6.25 Gb/s operation, and (2) the low noise power of our serial link test setup, which consisted of a test chip with a single transceiver.

For the high-loss backplane channel, which is not ISI dominated due to the usage of pre-emphasis, and has noise which limits the BER from being reliable, ECC coding was shown to improve the BER performance, improve the allowable increase in peak-to-peak clock jitter, and reduce power. For a 4 Gb/s line rate, ECC is the only way to transmit data reliably. The results show that reducing the *CR* improves the allowable increase in peak-to-peak clock jitter for a given line rate. For a 3.238 Gb/s information rate, ECC was shown to reduce the BER. Additionally, ECC was shown to improve the allowable increase in peak-to-peak clock jitter and reduce the power consumption for a given BER through the reduction of transmit driver power.

For the sub-Nyquist notch channel, ECC was shown to increase the maximum achievable information rate and reduce energy consumption. For information rates with Nyquist frequency at or below the notch minimum frequency, the utilization of coding was shown to be superior to the uncoded link for reliability and power reasons.

REFERENCES

[1] D. Derickson and M. Müeller, *Digital Communications Test and Measurement: High Speed Physical Layout Characterization*. Upper Saddle River, NJ: Prentice Hall, 2007.

[2] H. Johnson and M. Graham, *High-Speed Signal Propagation:Advanced Black Magic*. Upper Saddle River, NJ: Prentice Hall, 2003.

[3] ESD Association, "White paper 1: A case for lowering component level HBM/MM ESD specifications and requirements," Revision 2.0, Aug. 2007.

[4] ESD Association, "White paper 2: A case for lowering component level CDM ESD specifications and requirements," Revision 2.0, April 2010.

[5] *Human Body Model (HBM) - Component Level*, JS-001-2010, 2010.

[6] P. Hanumolu, G. Wei, and U. Moon, "Equalizers for high-speed serial links," *Int. J. High Speed Elec. Syst.*, vol. 15, no. 2, Jun. 2005.

[7] A. Faust, "The effect of on-chip ESD protection on reliable high-speed I/O link equalization power consumption," M.S. thesis, University of Illinois at Urbana-Champaign, 2011.

[8] S. Galal and B. Razavi, "Broadband ESD protection circuits in CMOS technology," *ISSCC Dig. Tech. Papers*, pp. 182-183, Feb. 2003.

[9] S. Galal and B. Razavi, "40-Gb/s amplifier and ESD protection circuit in 0.18-µm CMOS technology," *IEEE J. Solid-State Circuits*, vol. 39, no. 12, pp. 2389-2396, Dec. 2004.

[10] A. Srivastava, "Semidigital clock-data recovery system and bandwidth extension for ESD-protected high-speed extension for ESD-Protected high-speed IO circuits," M.S. thesis, University of Illinois at Urbana-Champaign, 2008.

[11] M. Harwood et al., "A 12.5Gb/s SerDes in 65 nm CMOS using a baud-rate ADC with digital receiver equalization and clock recovery," *ISSCC Dig. Tech. Papers*, pp. 436-437, 2007.

[12] P. Schvan et al., "A 24GS/s 6b ADC in 90 nm CMOS," *ISSCC Dig. Tech. Papers*, 2008.

[13] H. Chung et al., "A 7.5-GS/s 3.8-ENOB 52-mW flash ADC with clock duty cycle control in 65 nm CMOS," *IEEE Symp. on VLSI Circuits*, pp. 268-269, 2009.

[14] R. Narasimha, "System-aware design of energy-efficient high-speed I/O links," Ph.D. dissertation, University of Illinois at Urbana-Champaign, 2011.

[15] J. F. Bulzacchelli et al., "A 10-Gb/s 5-tap DFE/4-tap FFE transceiver in 90-nm CMOS Technology," *IEEE J. Solid-State Circuits*, vol. 41, no. 12, pp. 2885-2900, Dec. 2006.

[16] D. J. Costello Jr. and G. D. Forney Jr., "Channel coding: The road to channel capacity," *Proc. IEEE*, vol. 95, no. 6, pp. 1150-1177, Jun. 2007.

[17] R. Narasimha and N. Shanbhag, "Forward error correction for high-speed I/O," in *Asilomar Conference on Signals, Systems and Computers*, 2008, pp. 1513-1517.

[18] R. Narasimha and N. Shanbhag, "Design of energy-efficient high-speed links via forward error correction," *IEEE Trans. Circuits Syst. II*, vol. 57, no. 5, pp. 359-363, May 2010.

[19] C. Richier et al., "Investigation on different ESD protection strategies devoted to 3.3 V RF applications (2 GHz) in a 0.18 µm CMOS process," *Proc. EOS/ESD Symp.*, pp. 251-259, 2000.

[20] A. Hajimiri, "Generalized Time- and Transfer-Constant Circuit Analysis," *IEEE Trans. Circuits and Systems I*, vol. 57, no. 6, pp. 1105-1121, June 2010.

[21] E. Rosenbaum, H.-M. Bae, K. S. Bhatia, and A. C. Faust, "Moving signals on and off chip," *Proc. IEEE CICC*, pp. 585-592, 2009.

[22] G. F. Franklin, J. D. Powell, and A. Emami-Naeini, *Feedback Control of Dynamic Systems*. Upper Saddle River, NJ: Pearson Education, 2006.

[23] R. E. Ziemer and W. H. Tranter, *Principles of Communication: Systems, Modulation, and Noise*, 6th ed. Hoboken, NJ: John Wiley & Sons, Inc., 2009.

[24] A. C. Faust, A. Srivastava, and E. Rosenbaum, "Effect of on-chip ESD protection on 10 Gb/s receivers," in *Proc. EOS/ESD Symp.*, 2011, pp. 106-115.

[25] J. R. Andrews, "RZ vs. NRZ," *Appl. Note AN-12*, Picosecond Pulse Labs, 2001.

[26] J. Craninckx and M. Steyaert, "A 1.8-GHz CMOS low-phase-noise voltage-controlled oscillator with prescalar," *IEEE J. Solid-State Circuits*, vol. 30, no. 12, pp. 1474-1482, Dec. 1995.

[27] R. Merrill and E. Issaq, "ESD design methodology," *Proc. EOS/ESD Symp.*, pp. 233-237, 1993.

[28] D.J. Allstot, X. Li, and S. Shekhar, "Design considerations for CMOS low-noise amplifiers," *Proc. IEEE RFIC Symp.*, pp. 97-100, 2004.

[29] C.-M. Hsu, M.Z. Straayer, and M.H. Perrott, "A low-noise wide-BW 3.6-GHz digital delta-sigma fractional-N frequency synthesizer with a noise-shaping time-to-digital converter and quantization noise cancellation," *IEEE J. Solid-State Circuits*, vol. 43, pp. 2776-2786, Dec. 2008.

[30] C.-M. Hsu, "Techniques for high-performance digital frequency synthesis and phase control," Ph.D. dissertation, MIT, Cambridge, MA 2008.

[31] J. M. Rabaey, A. Chandrakasan, and B. Nikolic, *Digital Integrated Circuits: A Design Perspective*, 2nd ed. Upper Saddle River, NJ: Printice Hall, 2003.

[32] K. S. Bhatia, "Design and ESD protection of wideband, radio frequency integrated circuits in CMOS technologies," Ph.D. dissertation, University of Illinois at Urbana-Champaign, 2006.