

DESIGN OF RELIABLE AND ENERGY-EFFICIENT
HIGH-SPEED INTERFACE CIRCUITS

BY

MIN-SUN KEEL

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, 2015

Urbana, Illinois

Doctoral Committee:

Professor Elyse Rosenbaum, Chair
Associate Professor Pavan Kumar Hanumolu
Professor José E. Schutt-Ainé
Professor Naresh R. Shanbhag

Abstract

The data-rate demand in high-speed interface circuits increases exponentially every year. High-speed I/Os are better implemented in advanced process technologies for lower-power systems, with the advantages of improved driving capability of the transistors and reduced parasitic capacitance. However, advanced technologies are not necessarily advantageous in terms of device reliability; in particular device failure from electrostatic discharge (ESD) becomes more likely in nano-scale process nodes. In order to secure ESD resiliency, the size of ESD devices on I/O pads should be sufficiently large, which may potentially reduce I/O speed. These two conflicting requirements in high-speed I/O design sometimes require sacrifice to one of the two properties.

In this dissertation, three different approaches are proposed to achieve reliable and energy-efficient interface circuits. As the first approach, a novel ESD self-protection scheme to utilize “adaptive active bias conditioning” is proposed to reduce voltage stress on the vulnerable transistors, thereby reducing the burden on ESD protection devices. The second approach is to cancel out effective parasitic capacitance from ESD devices by the T-coil network. Voltage overshoot generated by magnetic coupling of the T-coil network can be suppressed by the proposed “inductance halving” technique, which reduces mutual inductance during ESD. The last approach employs system-level knowledge in the design of an ADC-based receiver for high intersymbol interference (ISI) channels. As a system-level performance metric, bit-error rate (BER) is adopted to mitigate a bit-resolution requirement in “BER-optimal ADC”, which can lead to $2\times$ power-efficiency in the flash ADC and achieve a better BER performance.

Acknowledgments

When I first came to the United States as a “student” after working for a company, I had several plans in my mind for the academic accomplishments. Now, I cannot say that all my goals are achieved, but my study and life as a graduate student are pretty satisfactory, in my thought. All my accomplishments are thanks to many people.

First of all, I would like to express my gratitude to my advisor, Prof. Elyse Rosenbaum, for her guidance and patience. She always helped me to develop my crude ideas in academic and practical ways. Especially when I was at a loss in the middle of my graduate study, she was the one who encouraged me to go on. I would also like to thank Prof. Pavan Hanumolu, for his invaluable teaching and advice on my research. I am also grateful to Prof. Naresh Shanbhag for his mentoring on how to attack a technical problem in a different paradigm, and Prof. Jose Schutt-Aine for his teaching on high-speed theory and measurement.

Many thanks are due to my friends and colleagues. In particular, my office-mate, Dr. Adam Faust, helped me to prepare for life as well as study in the United States. Dr. Nathan Jack was always kind to answer my questions and gave me one of the research ideas. I also would thank Dr. Karan Bhatia for his invaluable advice and discussion on my research topics. My friends at UIUC also deserve my thanks: Dr. Vrashank Shukla, Robert Mertens, Kuo-Hsuan Meng, Nick Thomson, Yang Xiu, Zaichen Chen, Collin Reiman and Dabin Zhang, all from Prof. Rosenbaum’s group. I also have many friends in the other circuit groups, whom I really thank: Mingu Kang, Dr. Eric Kim, Yingyan Lin, Sai Zhang, Wooseok Choi, Seong Joong Kim, Mrunmay Talegaonkar, Saurabh Saxena and Tejasvi Anand, to name a few.

Finally, I would like to express my deepest gratitude to my wife, Hyunjoo, my daughter, Dayeon, and my parents for their endless support, patience and love. My work at UIUC would not have been possible without them. This dissertation is dedicated to my beloved family.

Table of Contents

Chapter 1. Introduction	1
Chapter 2. Adaptive Active Bias Conditioning for ESD Self-Protection	4
2.1 Motivation	4
2.2 Bidirectional SSTL Transceiver	7
2.3 ESD Protection Strategy	8
2.4 AABC Circuit	10
2.5 Chip Implementation	13
2.6 Overview on ESD Measurement and Failure Analysis Method	14
2.7 ESD Measurement Results	16
2.8 Effectiveness of AABC	20
2.9 Summary	25
Chapter 3. T-Coil Network for I/O Bandwidth Extension	26
3.1 Motivation	26
3.2 T-Coil Techniques	27
3.3 CDM Hazard in T-Coil	35
3.4 CDM-Reliable T-Coil	40
3.5 Receiver Circuit Implementation	58
3.6 IC Layout and Chip Microphotograph	87
3.7 Simulation Results	90
3.8 Test Board Design	104
3.9 Measurement Results	107
3.10 Summary	110
Chapter 4. BER-Optimal ADC for ADC-Based Receivers	112
4.1 Motivation	112
4.2 BER-Optimal ADC	113
4.3 ADC Design	115
4.4 Full-Chip Simulation	123
4.5 Full-Chip Floorplan and Layout	124
4.6 Test Board Design	126
4.7 Measurement Results	127
4.8 Summary	132
Chapter 5. Conclusion and Future Work	134
5.1 Conclusion	134
5.2 Future Work	135

Appendix A. Power-On I-V Curve Method.....	137
Appendix B. T-Coil Parameter Calculation from Multi-Port S-Parameters	141
B.1 Definitions of Inductance and Quality Factor	141
B.2 Parameter Extraction from Multi-Port Inductors	143
B.3 Calculation of Tap-Resistance	146
References.....	148

Chapter 1. Introduction

High-speed wireline I/O design is one of the most challenging areas in electrical engineering. It is an enabling technique for high-capacity data communication or computer networks, ranging from long-distance Ethernet, to medium-reach backplane, or even shorter-reach chip-to-chip interface. Especially nowadays, tremendous volumes of data are generated from all over the world, which calls for high data-rate wireline communication links. Figure 1.1 shows the increasing trend of per-pin data-rate in various wireline I/O standards [1]. It is apparent that the data rate increases by approximately $2\times$ every four years. In order to keep up with this rapid rising trend of data rate, process technologies with faster and smaller devices are needed for low-power system implementation. The data-rate trend with process nodes in Figure 1.2 evidently shows that higher data-rate I/Os are implemented on more advanced processes.

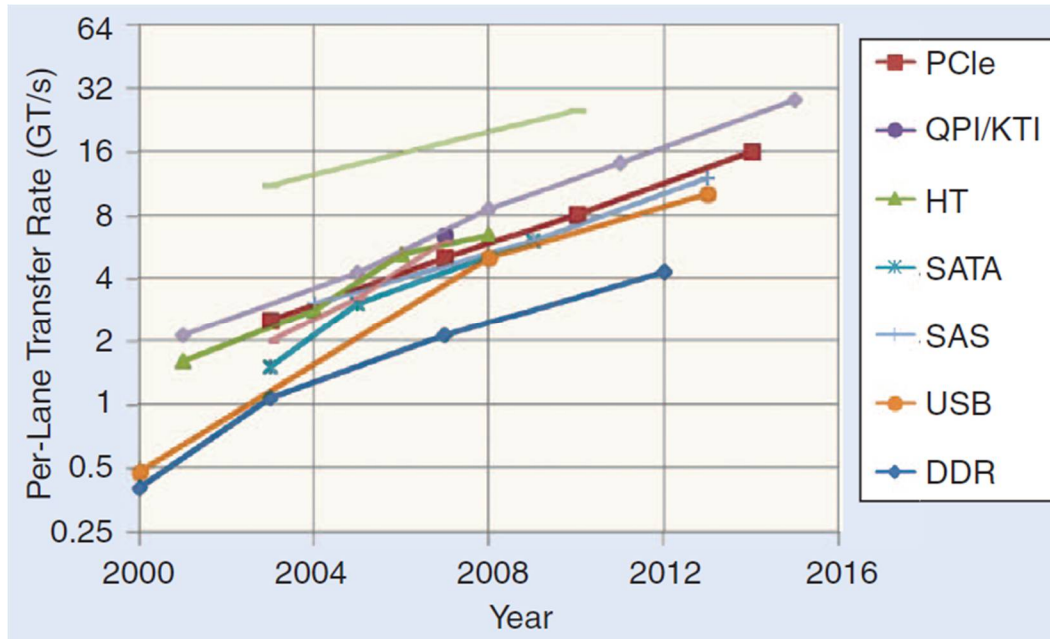


Figure 1.1. Data rate per pin vs. year for various wireline I/O standards [1].

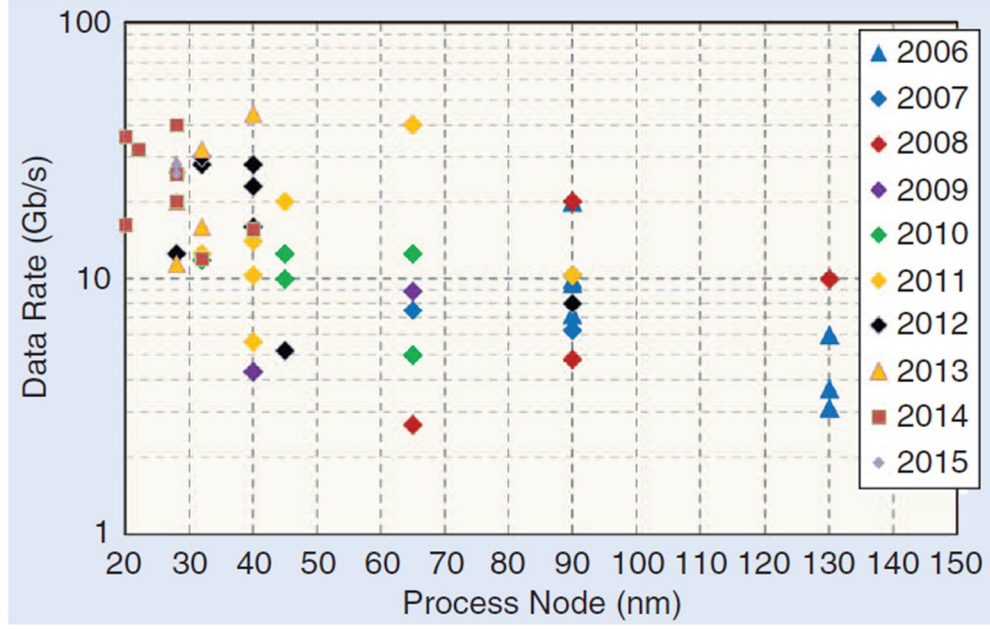


Figure 1.2. Data-rate trends vs. process technology nodes and year in wireline communication links [1].

However, advanced technologies do not necessarily guarantee device reliability. In particular, electrostatic discharge (ESD) reliability is always one of the major issues with process scale-down. ESD devices should operate above IC supply voltage and below IC breakdown voltage, which defines the “ESD design window” [2], which is redrawn in Figure 1.3. The rectangles labeled as the minimum feature size of each process generation represent the ESD voltage margin by the width, whose upper limit and lower limit are determined by breakdown voltage and IC supply voltage, respectively. The height of each rectangle denotes the current withstand level of the ESD clamps. It is obvious that the ESD design window reduces with advanced process nodes, as a reduction rate of breakdown voltage is much faster than that of supply voltage. Therefore, the use of advanced process technologies is not always favorable to ESD resiliency.

In order to overcome this dilemma in high-speed I/O design, a new design approach is necessary to accomplish both energy-efficiency and ESD-reliability. In this dissertation, three design concepts are proposed: (1) ESD self-protection, (2) ESD parasitic cancellation, and (3) circuit design metric redefinition toward system performance. These approaches will be discussed in more detail from Chapter 2 to Chapter 4, and verified by simulation and measurements results.

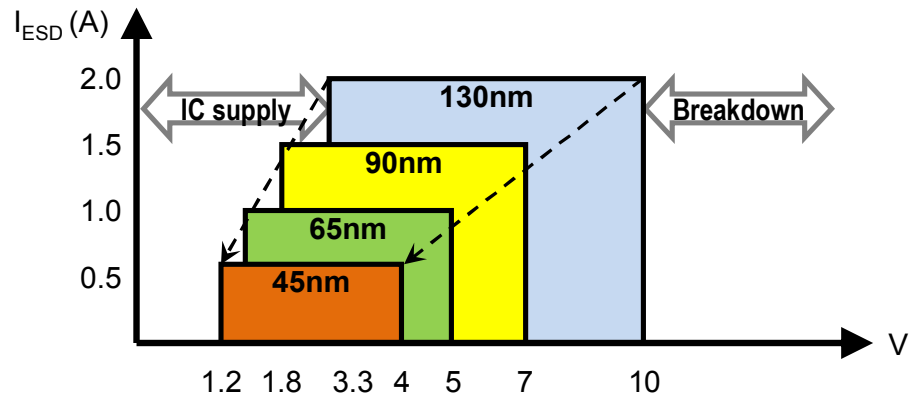


Figure 1.3. ESD design window with process scaling (redrawn from [2]). The ESD voltage margin is represented by the width, and current carrying capability of the clamp circuits is indicated by the height of the rectangle in each process node.

Chapter 2. Adaptive Active Bias Conditioning for ESD Self-Protection

2.1 Motivation

In high-speed I/O design, there is a stringent trade-off relationship between I/O bandwidth and ESD protection level. As shown in Figure 2.1 [3], a higher data-rate limits the capacitive loading budget on the I/O pad, which corresponds to the decreased ESD protection level. This trend becomes even worse for the over-10 Gb/s I/Os, which are currently the main stream applications. Moreover, in order to support a higher data rate, ICs should be fabricated in nano-scale CMOS technologies, in which good ESD reliability is difficult to achieve because of the reduced gate oxide breakdown voltage (BV_{ox}), reduced drain-source breakdown voltage (i.e., V_{ti} , the lateral BJT trigger voltage) and increased resistance of thinner interconnection metal layers [2]. Hence, sometimes high-speed I/O designers have to sacrifice ESD reliability to achieve the required bandwidth.

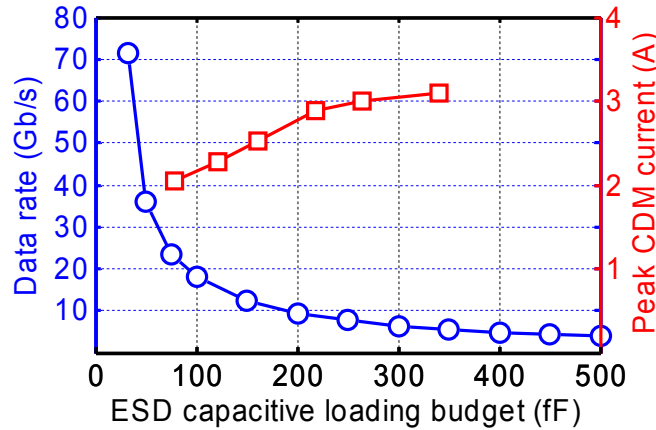


Figure 2.1. Relationship between data-rate of high-speed serial links and allowable capacitive loading requirements with CDM protection level. The peak CDM current on the right y-axis shows maximum sustainable CDM current by dual-diodes, with the capacitive loading budget in the x-axis (the original plot of Figure 12 in [3] is redrawn).

The most common ESD target voltages are 2 kV for Human Body Model (HBM) [4] and 500 V for Charged Device Model (CDM) [3]. The HBM and CDM models reproduce the rise time and peak current of actual ESD waveforms; the HBM or CDM tester pre-charge voltage can be set to the target value. In

the old process technologies, these ESD target levels can be easily satisfied using known ESD protection devices and the proper current shunting networks. However, conventional ESD design approaches do not work well in the advanced technology nodes, especially for high-speed I/O design, and it can be difficult, or even impossible, to achieve 2 kV HBM and 500 V CDM. Given this situation, the “practical” ESD target levels were recently reconsidered by the Industry Council on ESD Target Levels [3], [4], which recommended to lower the HBM target level from 2 kV to 1 kV [4], and the CDM level from 500 V to 250 V [3], taking advantage of improved factory controls.

However, even the new ESD target levels may be unrealizable for very high-speed serial links (> 20 Gb/s) in advanced processes, as an example will show. As an ESD protection network, primary dual-diodes at the I/O pads and an active rail clamp are assumed [5]. A BGA package of 300 mm² size is assumed, so that a 250 V CDM discharge will produce a 3 A peak current [3]. The induced PAD voltage (V_{PAD}) from ESD current (I_{ESD}) can be calculated as follows:

$$V_{PAD} = I_{ESD} (R_{dio} + R_{clp} + R_{bus}) + V_{dio,on} + V_{clp,on}, \quad (2.1)$$

where R_{dio} and $V_{dio,on}$ are on-resistance and turn-on voltage of dual-diodes, respectively, and R_{clp} and $V_{clp,on}$ represent on-resistance and turn-on voltage of a rail clamp, respectively. Bus resistance on the VDD and VSS rails are denoted by R_{bus} . The model parameters for dual-diodes are obtained from the 130-nm CMOS process design kit (PDK) and series resistance, which is proportional to capacitance, is scaled based on the ESD capacitance budget according to the serial link data-rate for 5 Gb/s and 20 Gb/s as shown in Figure 2.1. Dual-diodes parameters are further modified for CDM simulation to emulate transient voltage overshoot by simply multiplying the series resistance of the dual-diodes by 1.5. As shown in Figure 2.2, I_{ESD} is plotted with respect to V_{PAD} to show the I-V characteristic. The peak ESD currents for 1 kV HBM and 250 V CDM are indicated by the horizontal lines; V_{PAD} at the crossing point between an I-V curve and the horizontal line is required to be smaller than device breakdown voltages. The maximum tolerable V_{PAD} is set equal to V_{tl} for HBM and BV_{ox} for CDM; these values are process-dependent, and two process nodes are studied: 130 nm and 45 nm. Not only transistors, but the ESD

devices (e.g., dual-diodes and rail clamps) can also have process-dependency on the ESD levels. However, in this analysis, it is assumed that ESD devices are more robust than the transistors, so that ESD failure always comes from the transistors. As shown in Figure 2.2(a), a 5-Gb/s link should easily pass 1 kV HBM at both the 130- and 45-nm nodes, but a 45-nm 20-Gb/s link only marginally meets the target. Figure 2.2(b) indicates that a 5-Gb/s link in 130-nm CMOS can pass 250 V CDM, but the 45-nm version will be marginal, and a 45-nm 20-Gb/s link cannot reach the CDM target at both process nodes. In this simple analysis, the ESD device model parameters were based on 130-nm PDK and not modified for 45-nm process. Thus, the predicted results for 45-nm technology might be overly optimistic.

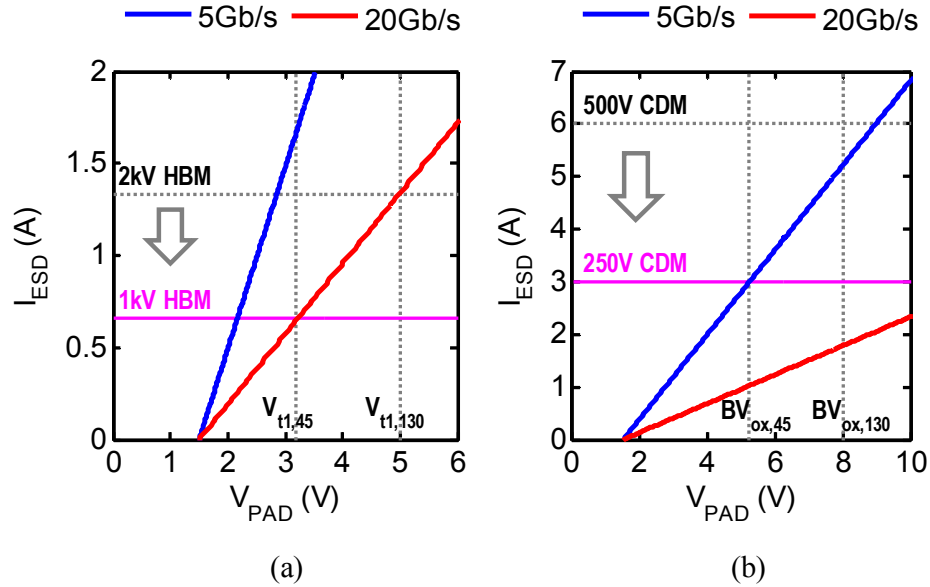


Figure 2.2. ESD current vs. peak I/O PAD voltage for 5-Gb/s and 20-Gb/s serial links during (a) HBM and (b) CDM. Breakdown voltages for 130 and 45-nm CMOS technologies are indicated.

In response, various ESD self-protected I/Os have been proposed. One of the promising techniques is an active biasing scheme in which the gate biases of the ESD-susceptible transistors are controlled to reduce the voltage stresses at the gate-oxide and drain-source region [6], [7], [8], [9], [10]. Only [6] addressed the case of bi-directional transceivers, but its solution is effective only for positive I/O stress relative to VSS (PS mode), and it does not mitigate negative I/O stress that is applied with respect to

VDD (ND mode). ND-mode ESD primarily stresses PMOS, which could be problematic in nano-scale processes, where the breakdown voltages of PMOS approach those of NMOS [11].

In this dissertation, *Adaptive Active Bias Conditioning* (AABC) is proposed. It provides ESD polarity-aware protection by controlling the gate voltages of the vulnerable transistors based on the direction of ESD current flow. AABC is demonstrated on a Stub Series Terminated Logic (SSTL) transceiver [12].

2.2 Bidirectional SSTL Transceiver

The proposed ESD protection concept is illustrated in Figure 2.3. ESD current shunting is provided by dual-diodes at the I/O and an active rail clamp. The AABC circuit generates appropriate gate voltages for the ESD-vulnerable transistors. The AABC circuit is composed of two blocks: gate-bias controller and direction detector, which are triggered by an ESD detector inside an active rail clamp only when a power-off ESD event occurs.

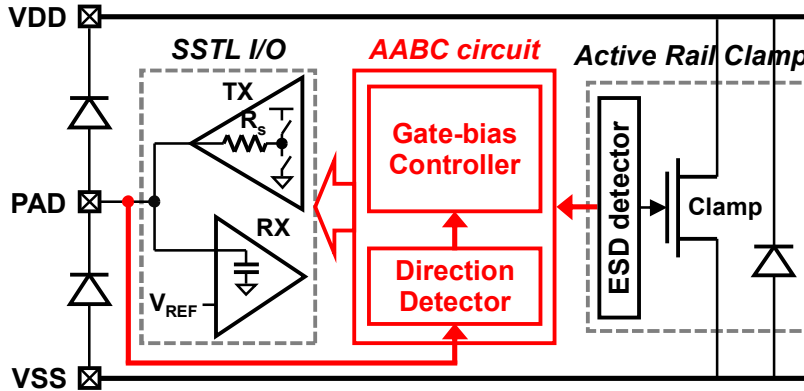


Figure 2.3. Block diagram of ESD-protected transceivers. The AABC circuit is triggered by a fast transient on VDD when the chip is in the power-off state, and it controls the gate biasing of the vulnerable transistors.

The schematic of the single-ended bidirectional SSTL transceiver is drawn in Figure 2.4. The receiver (RX) performs a comparison between received data at PAD and a reference voltage (V_{REF}). The RX comparator is based on a complementary self-biased differential amplifier [13], which generates its tail current locally using negative feedback. The original topology is modified by adding transistors MP3,

MN3 and MN4 to support AABC control and to eliminate stand-by leakage current. Appropriate number of inverters will follow /RXOUT to make a full signal swing with the correct logic polarity. The transmitter (TX) is an array of fifteen voltage-mode drivers with on-die termination (ODT) controlled by a 4-bit control register, “ZQ[3:0].” ODT provides 50-Ω series termination during TX mode, and 50-Ω parallel termination (or, Thevenin termination) for RX mode by a resistor divider with two 100-Ω resistors. This resistor divider at RX mode also generates a common-mode voltage for AC coupling at the RX input. Either the receiver or the transmitter is selected via the “ENABLE_RX” signal. The ESD control signals, ESDP, ESDN, and ESD1-3, are off the main signal path.

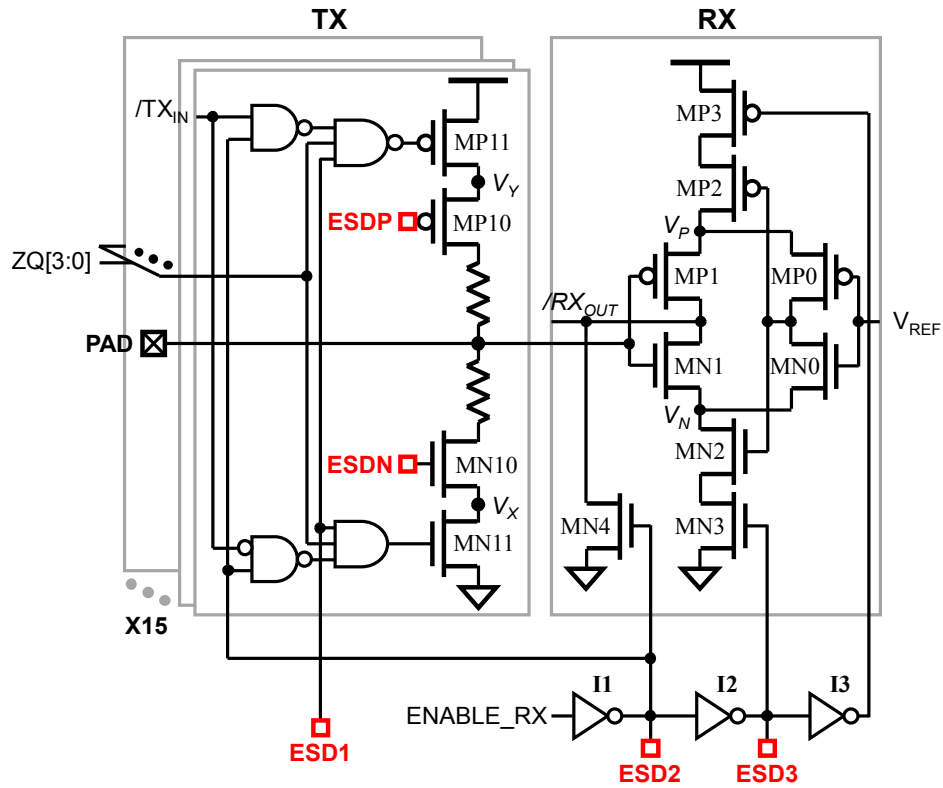


Figure 2.4. Schematic of the single-ended bidirectional SSTL transceiver with ESD control signals (marked in red).

2.3 ESD Protection Strategy

Multi-Gb/s operation will require the circuit to be implemented with thin gate-oxide transistors. Analysis is performed to identify the most vulnerable transistors during HBM ESD [14]. ESD current

discharges from PS mode and ND mode create the most severe voltage stress [15], because ESD current flows through the discharge path with the largest aggregate resistance. Figure 2.5(a) shows PS-mode current flow, which causes a large positive voltage drop between PAD and VSS, resulting in gate-oxide stress on MN1 in the receiver and drain-source stress on MN10 (or MN11) in the driver. Similarly, as shown in Figure 2.5 (b), MP1 and MP10 (or MP11) are vulnerable during ND mode stress with the largest negative voltage drop between PAD and VDD.

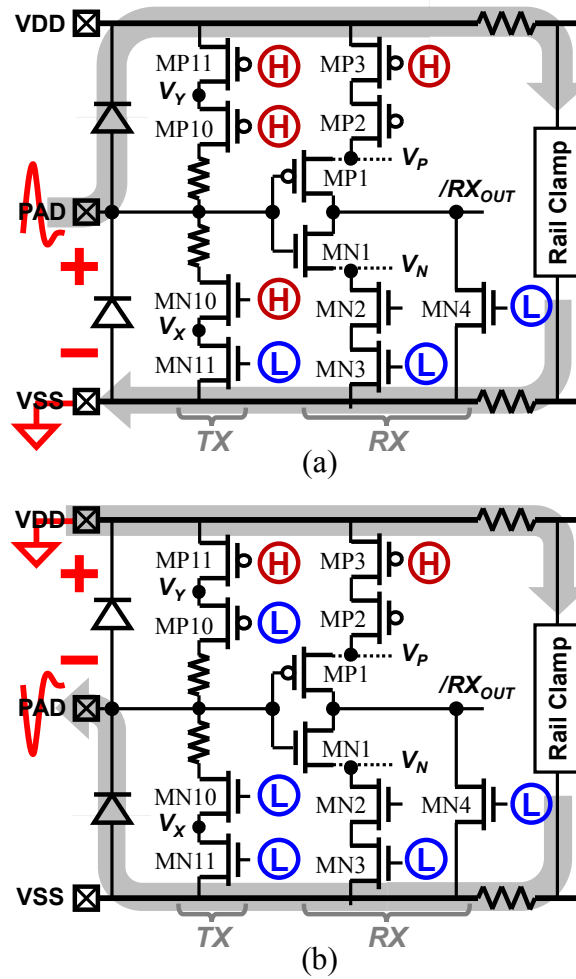


Figure 2.5. HBM stress modes: (a) PS mode, and (b) ND mode. The gate bias states that would maximize the resilience of this circuit to the ESD stress are indicated as circled letters: H as logic-high (VDD) and L as logic-low (VSS).

The receiver transistors, MP1 and MN1, can be protected by turning off transistors MP3, MN3 and MN4 during ESD; this causes nodes V_P , V_N and $/RX_{OUT}$ to be set to an intermediate voltage between

VDD and VSS, thereby minimizing V_{GS} and V_{GD} for MN1 and MP1 [6]. The driver transistors should be protected from drain breakdown; the transistors' drain-source breakdown voltage can be maximized by biasing them in the off state, as this retards the lateral BJT turn-on. However, setting $V_{G,MN10}$ to VSS during PS mode stress or $V_{G,MP10}$ to VDD during ND mode would create gate-oxide overstress, because $V_{DG} = V_{PAD}$ in both cases. To prevent this, the gates of MN10 and MP10 should be pulled high during PS mode stress and low during ND mode as indicated in Figure 2.5. This results in $V_X = VDD - V_{th,n}$ in PS mode, and $V_Y = VSS + |V_{th,p}|$ in ND mode; if the ESD diode and rail clamp have similar voltage drops, $V_{DS,MN10}$ and $V_{DS,MN11}$ will be roughly equal to $V_{PAD}/2$ at PS mode, and $V_{DS,MP10} = V_{DS,MP11} \approx V_{PAD}/2$ at ND mode, which best suppresses drain junction stress. The desired ESD control signals are summarized in Table 2.1, based on the required gate bias states in Figure 2.5 and the logic configuration in Figure 2.4.

Table 2.1. Desired ESD control signals for different operation modes.

Signal Mode		ESDP	ESDN	ESD1	ESD2	ESD3
ESD	PS mode	H	H	L	L	L
	ND mode	L	L			
Normal	TX mode	L	H	H	H	L
	RX mode			H	L	H

2.4 AABC Circuit

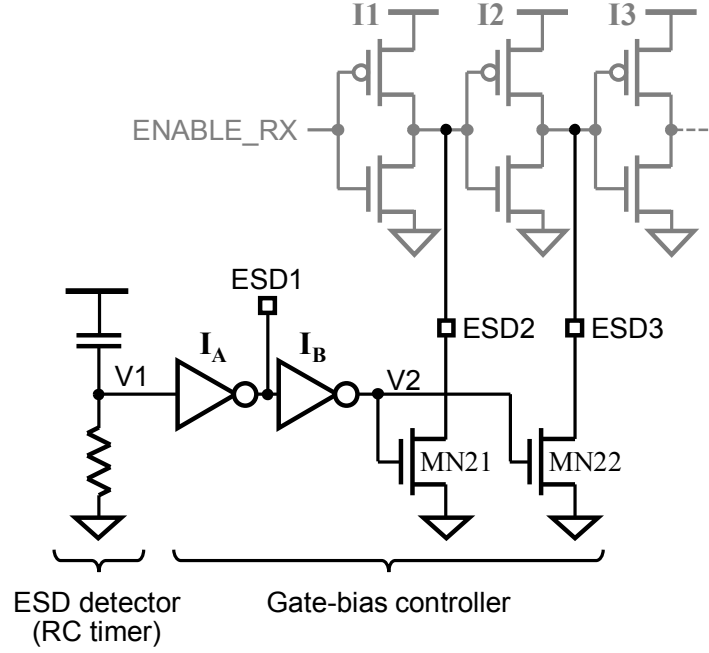
The AABC circuit implementation is shown in Figure 2.6. The RC timer in Figure 2.6(a) is used as an ESD detector, which detects an ESD fast transient on the supply and pulls up node V1 to generate logic low signals for ESD1-3 in the gate-bias controller. In this study, no dedicated RC timer circuit is designed, but the V1 signal is reused from the existing rail clamp in the power I/O cells, which are located near the SSTL I/Os. However, if the distance between the SSTL I/O and the rail clamp circuit is far away in a practical implementation, the RC timer should be included in the AABC circuit for a faster ESD detection. In Figure 2.6(a), the inverter I_A should be sized sufficiently large to deliver the signal ESD1 to many transistor gates in the driver within an ESD pulse duration. ESD2 and ESD3 should both be logic-

low during an ESD event, while they are to be controlled by the ENABLE_RX signal in a normal operation; therefore, the drain nodes of MN21 and MN22 are directly connected to the outputs of inverters I1 and I2, respectively, as shown in the gray-colored circuits in Figure 2.6(a). Thus, during an ESD event, one of MN21 or MN22 will be in contention with I1 or I2, respectively. MN21 [MN22] must be carefully sized to overcome the driving capability of the PMOS in I1 [I2].

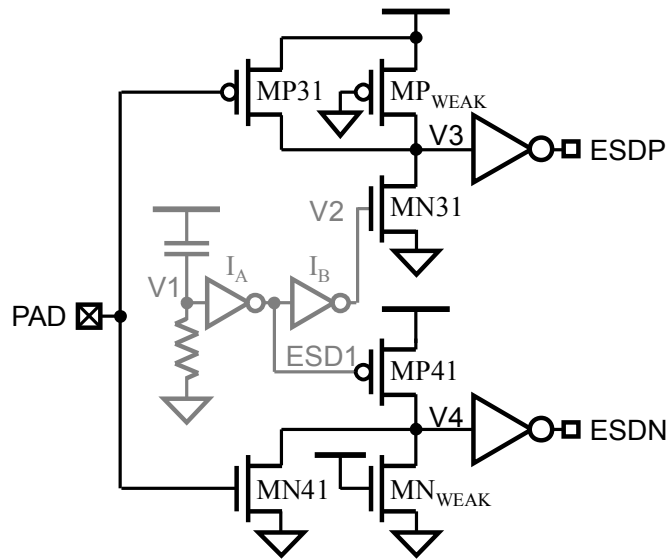
The direction detector in Figure 2.6(b) makes use of the fact that $V_{PAD} > VDD$ in PS mode and $V_{PAD} < VSS$ in ND mode. The most important feature of the direction detector is low latency because an ESD pulse is a very short one-shot pulse—1 ns in case of CDM [16]. ESD control signals must be generated quickly before the ESD pulse reaches its peak amplitude. Thus, the circuit is implemented using a ratioed logic style [17] for its speed from simple circuit topology and its ability to detect out-of-supply-range signals ($V_{PAD} > VDD$ or $V_{PAD} < VSS$). Careful sizing of transistors minimizes contention at internal nodes, V3 and V4. The combined current driving capability of MP31 and MP_{WEAK} [MN41 and M_{NWEAK}] should be larger than MN31 [MP41], while a single MP_{WEAK} [M_{NWEAK}] should be weaker than MN31 [MP41]. For PS-mode ESD, MP31 is off and MP41 fully turns on, to pull the nodes V3 and V4 down; this drives ESDP and ESDN high. In ND mode, ESDP and ESDN are driven low. During normal operation, MN31 and MP41 are off; V3 is pulled up by MP_{WEAK} and V4 is pulled down by M_{NWEAK}, eliminating static current. The direction detector adds less than 5 fF parasitic capacitance to the PAD, which is negligible relative to the ESD diodes and bond pad. Thus, the AABC circuit should not impact signal integrity.

To verify the transient response of the AABC circuit, circuit simulation was performed using Spectre [18] for the worst-case conditions on the parasitic-extracted circuit netlist from the actual layout; transistors are in slow corner, and temperature is set to 135 °C. A CDM-like current pulse of $t_{pulse} = 1$ ns, $t_{rise} = 100$ ps and 3A peak current is injected; the two slowest signals ESDP and ESDN are plotted in Figure 2.7 along with PAD and VDD, all with respect to VSS. ESDP and ESDN settle to the desired

states (logic-high at PS mode and logic-low at ND mode) within 200 ps. This indicates the direction detector can work on the CDM time-scale.



(a)



(b)

Figure 2.6. AABC schematic: (a) gate-bias controller with RC timer from existing rail clamp circuit (the inverters I1-3 in Figure 2.4 are repeated), and (b) direction detector (part of (a) is repeated here to clarify signal connection).

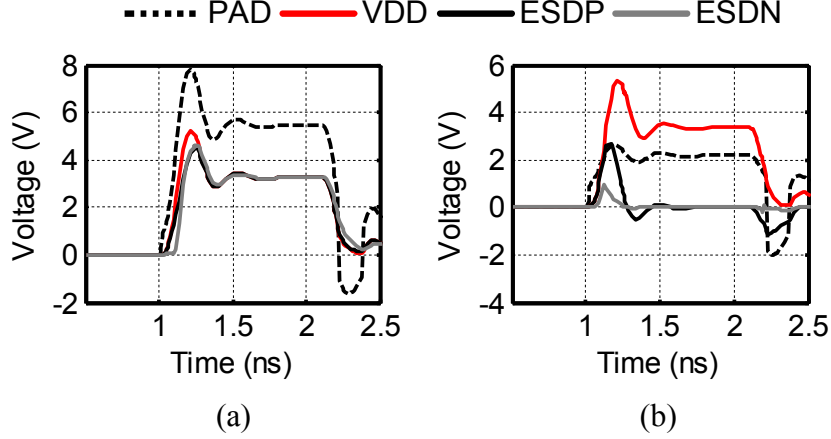


Figure 2.7. Transient simulation of direction detector response to 1 ns pulse stress: (a) PS mode, (b) ND mode. Slow process corner, 135 °C, post-layout parasitics included. The ESD control signals settle within 200 ps.

2.5 Chip Implementation

To evaluate the efficacy of AABC, the four SSTL transceivers described in Table 2.2 were implemented on a 130-nm low-power CMOS process. A 2 Gb/s data-rate was targeted. All four transceivers (TRX) have dual-diode primary ESD protection. TRX1 is a baseline circuit, which has only primary dual-diodes but does not have secondary ESD protection and active-biasing schemes. TRX2 uses the active biasing circuit presented in [6] which protects only against PS-mode ESD. TRX3 contains the proposed AABC, shown in Figure 2.6. TRX4 does not utilize active biasing, but has secondary ESD protection, consisting of a series resistor and grounded-gate NMOS (GGNMOS) at the receiver input. The chip microphotograph is shown in Figure 2.8. The transceiver core occupies 0.0116 mm², and the AABC circuit is 0.0012 mm²—only 10% area overhead. The test chip is assembled in an 80-pin, 12×12 mm² QFN package. Transceivers in the test chip are programmed by the 10-bit shift-register via a three-wire interface, similar to the Serial Peripheral Interface (SPI); the shift-register configures the RX or TX mode for the bidirectional TRXs, and changes the driving strength and termination resistance. To study the effect of AABC on signal integrity, the TX eye diagrams are measured at the output of a 3-dB attenuation channel; results are shown in Figure 2.9 for a 2-Gb/s 2⁷−1 pseudo-random bit sequence (PRBS). TRX1 and TRX3 have similar eye openings, proving that the AABC circuit does not degrade signal integrity.

Table 2.2. SSTL transceiver splits.

	TRX1 (prim)	TRX2 ([5])	TRX3 (AABC)	TRX4 (prim+2 nd)
Primary protection	Yes	Yes	Yes	Yes
Secondary protection	No	No	No	Yes
Active biasing [ESD polarity]	No	Yes [PS]	Yes [PS, ND]	No

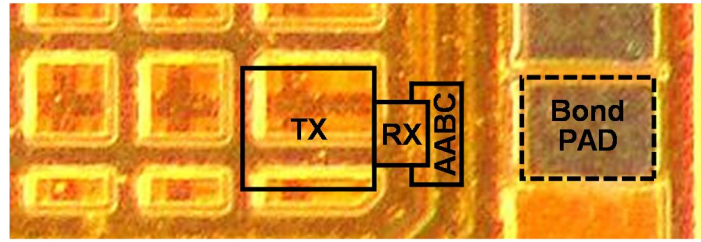


Figure 2.8. Chip microphotograph.

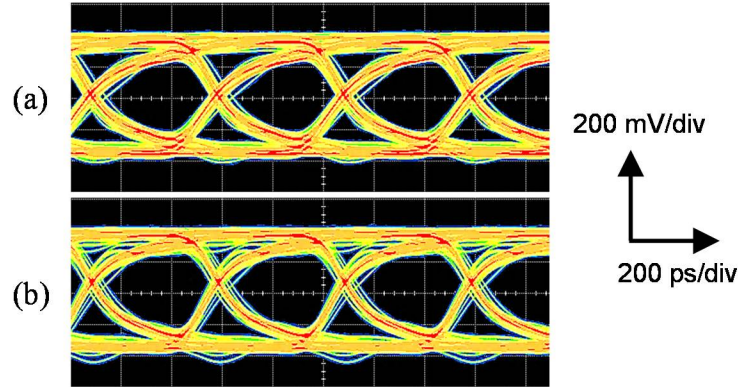


Figure 2.9. Measured TX eye diagram after 3-dB attenuation channel with 2^7-1 PRBS at 2 Gb/s: (a) SSTL I/O without AABC (TRX1), and (b) SSTL I/O with AABC (TRX3).

2.6 Overview on ESD Measurement and Failure Analysis Method

ESD tests are performed on the packaged parts, and failures are detected with the criteria of increased pin leakage current, deviation on the electrical parameters (e.g., TX amplitude, termination impedance, RX common-mode voltage, etc.) and TRX function failure from eye diagrams. Figure 2.10 shows two examples of RX eye diagrams from the failed chips after the ESD zap, for the same PRBS

pattern as Figure 2.9. Zero-crossing points are shifted up in Figure 2.10(a), and the RX does not function in Figure 2.10(b). The many “slips” on the waveforms in Figure 2.10(b) are a clear sign of RX malfunction; the ESD-damaged RX transistors does not perform the comparison function well, due to switching speed degradation or increased leakage current at the gate of the transistor.

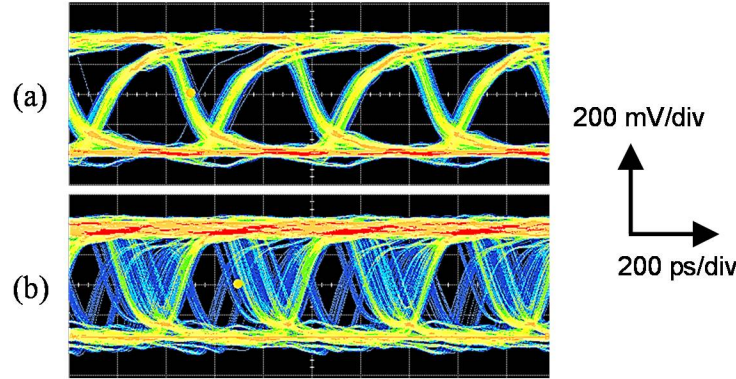


Figure 2.10. Examples of failed RX output eye diagrams after the ESD stress: (a) zero-crossing distortion, and (b) RX malfunction.

Pin leakage current was not increased in case of Figure 2.10(b), which implies that a simple leakage test is not sufficient to detect all ESD-induced failures. Moreover, pin leakage measurement on an unpowered chip did not give much information in this work. The test chip already flows some pin-leakage current even prior to ESD stress, because a few TX driver array turn on unexpectedly during DC pin voltage sweep. The possible cause is a charged-up power supply through the top diodes during power-off ESD, when the PAD-to-VSS voltage is swept; the gate voltages of some NMOS transistors in the TX driver may be accidentally pulled up to flow high leakage current. To avoid such unexpected leakage current, pin leakage is measured from I-V curves on a power-on board by disabling any unnecessary current through shift-register control before DC voltage sweep. This specific “power-on I-V curve” gives additional useful information on which section of core circuit or ESD device is failed, without resorting to any visual inspection by Scanning Electron Microscope (SEM). ESD-induced failure can be modeled as shunt resistance as shown in Figure 2.11. A finite top resistance (R_T) represents failure at the top diode or the PMOS transistor in I/O circuit, while a finite bottom resistance (R_B) models failure at the bottom diode or the NMOS transistor. In PS mode (Figure 2.11(a)), ESD current (I_{ESD}) flows through the top

diode and maximum stress voltage (V_{max}) is applied between PAD and VSS; in this case, finite R_T and R_B are caused by failure of the top diode and NMOS, respectively, because reverse breakdown voltage of the bottom diode is typically higher than V_{tl} or BV_{ox} of thin gate-oxide NMOS transistors. In ND mode, finite R_T and R_B represent failures from the PMOS and bottom diode, respectively, as shown in Figure 2.11(b). These two shunt resistance, R_T and R_B , can be calculated from the proposed power-on I-V curve method. More discussion is presented in Appendix A.

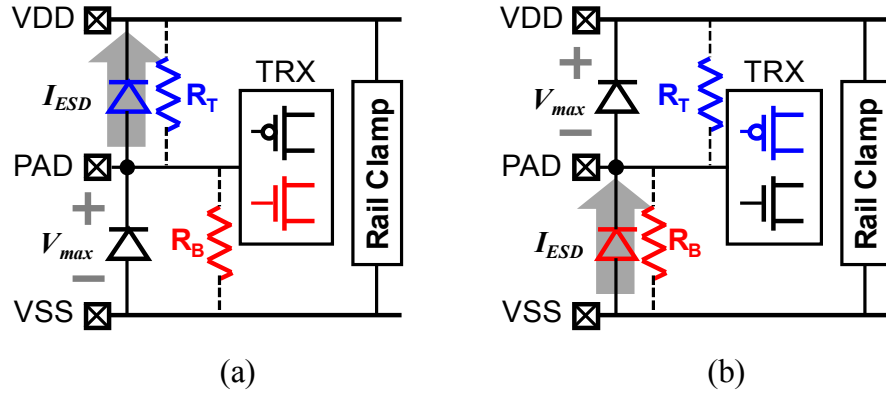


Figure 2.11. ESD failure model for (a) PS mode to show R_T caused from top diode failure and R_B from NMOS failure in transceiver circuit, and for (b) ND mode with R_T from the PMOS and R_B from the bottom diode failure. Critical ESD current, I_{ESD} , and maximum stress voltage, V_{max} , are indicated in the figure.

2.7 ESD Measurement Results

2.7.1 Transmission line pulsing (TLP) results

HBM performance is estimated by 100 ns Transmission Line Pulsing (TLP) with $t_{rise} = 10$ ns [19]. As presented in Table 2.3, all four TRX have very similar failure currents (I_{fail}). In each case, electrical failure analysis indicates that the failing device is an ESD diode. The ESD diodes each have a TLP I_{fail} of 2.7 A; I_{fail} of the SSTL I/O exceeds 3 A because the driver sinks some of the stress current. Given the relation, $V_{HBM} \approx 1500 \times I_{fail, TLP}$, all four transceivers were expected to pass HBM testing up to the tester limit of ± 4 kV.

An initial assessment of CDM performance is obtained from very-fast TLP (VF-TLP) [20]: $t_{\text{pulse}} = 5$ ns and $t_{\text{rise}} = 600$ ps. Results are given at the bottom two rows in Table 2.3. Failure analysis results for PS-mode and ND-mode VF-TLP are presented in Table 2.4. and Table 2.5., respectively. The failed items and the calculated shunt resistors from the proposed power-on I-V curve are listed, which will resolve the failure points. Under PS-mode VF-TLP stress, TRX1 fails due to failure at the receiver NMOS transistors. Figure 2.12 shows the eye diagrams of TRX1 at RX mode, before and after the PS-mode VF-TLP stress. As shown in Figure 2.12(b), the zero-crossing point has been significantly shifted to near the output-high level, which shows the receiver failure. However, there are no distinctive signs of increased leakage current and the shifted TX electrical characteristics. Hence, maybe the NMOSs of TRX1 are damaged slightly, not affecting leakage current and TX functionality, but damaged sufficiently for RX malfunction. TRX2 to TRX4 survive until the (top) ESD diode fails, and they have a 32% higher failure current than TRX1, as shown in Table 2.3. The failure point can be inferred from small R_T and failure of all electrical parameters, as shown in Table 2.4.. Under ND-mode VF-TLP stress, the AABC-protected TRX3 improves I_{fail} by 28% and 19% relative to TRX1 and TRX2, respectively. These results verify that the concept of AABC works, which protects regardless of the stress polarities. As shown in Table 2.5., TRX1 to TRX3 fail due to gate-oxide breakdown in the receiver PMOS, while failure in TRX4 is caused by the bottom diode with $R_B = 400\Omega$.

Table 2.3. Summary of failure current under TLP and VF-TLP stresses.

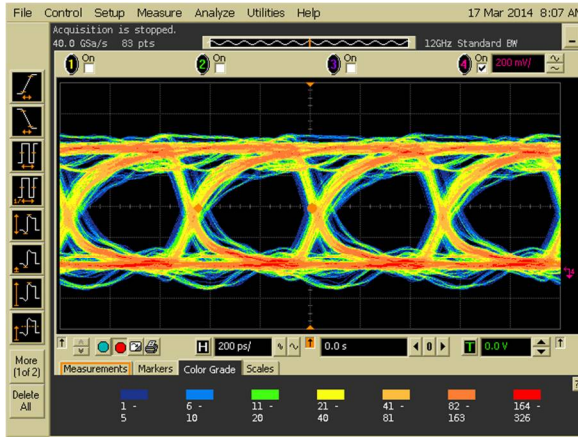
Mode \ TRX split		TRX1 (prim)	TRX2 ([6])	TRX3 (AABC)	TRX4 (prim+2 nd)
100ns TLP	PS mode	3.71 A	3.60 A	3.70 A	3.70 A
	ND mode	3.36 A	3.37 A	3.37 A	3.37 A
5ns VF-TLP	PS mode	9.26 A	12.17 A	12.18 A	12.24 A
	ND mode	6.87 A	7.39 A	8.76 A	9.14 A

Table 2.4. Failure analysis after PS-mode VF-TLP.

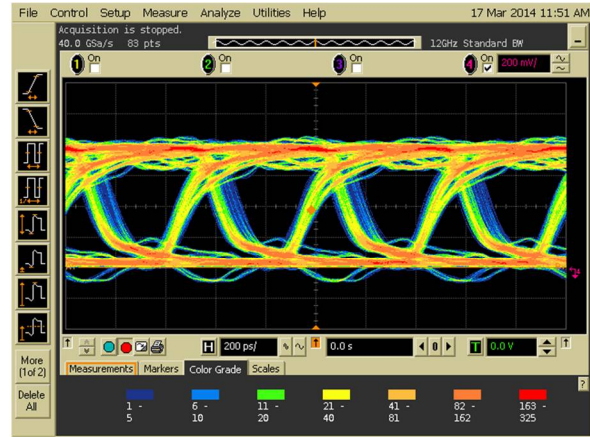
Items \ TRX		TRX1 (prim)	TRX2 ([6])	TRX3 (AABC)	TRX4 (prim+2 nd)
Failed items		•RX funct.	•Leakage •Elec. par. •RX funct.	•Leakage •Elec. par. •RX funct.	•Leakage •Elec. par. •RX funct.
Shunt Resistor	R_T	infinity	228 Ω	125 Ω	167 Ω
	R_B	infinity	626 Ω	infinity	infinity
Failure point		RX(NMOS)	top diode RX(NMOS)	top diode	top diode

Table 2.5. Failure analysis after ND-mode VF-TLP.

Items \ TRX		TRX1 (prim)	TRX2 ([6])	TRX3 (AABC)	TRX4 (prim+2 nd)
Failed items		•RX funct.	•RX funct.	•RX funct.	•Leakage •Elec. par. •RX funct.
Shunt Resistor	R_T	infinity	infinity	infinity	infinity
	R_B	infinity	infinity	infinity	400 Ω
Failure point		RX(PMOS)	RX(PMOS)	RX(PMOS)	bottom diode



(a)



(b)

Figure 2.12. RX eye diagrams for TRX1: (a) before the PS-mode VF-TLP stress, and (b) after the PS-mode VF-TLP.

2.7.2 HBM and FICDM results

HBM and field-induced CDM (FICDM) [16] tests are performed and summarized in Table 2.6.. HBM levels exceed 4 kV (HBM tester limit) at both positive and negative zaps for all transceivers, as expected from the TLP results (Table 2.3). Clearly, HBM protection of 130-nm CMOS technology does not present a special challenge, as was also shown in Figure 2.2.

Due to a limited number of samples and limited time on the FICDM tester, only negative CDM stress could be applied. Three chips were tested using 250 V steps. The ESD current direction of negative CDM is similar to that of PS-mode VF-TLP [21]. As shown in Table 2.6., TRX2 and TRX3 have a -1000V CDM failure level, which is 33% larger than that of TRX1; this is consistent with the PS-mode VF-TLP results. AABC is thus proven to enhance CDM reliability. Failure analysis results for negative CDM are summarized in Table 2.7.. During CDM test, only electrical parameters are measured because of a limited access to the required test equipments at the ESD laboratory. However failure point can be inferred from the failed electrical characteristics. For example, after each CDM stress, common-mode voltages at RX-mode were measured. Near CDM failure voltage, RX common-mode level had been decreased from its nominal level (~780 mV). This fact implies that the NMOS gate leakage to GND lowered RX common-mode level. From these observations, TRX1-3 show receiver NMOS failure. Different from PS-mode VF-TLP (Table 2.4.), no ESD diode failure is observed during negative CDM. This discrepancy may come from the quality of the probes used in the two ESD test environments. In FICDM tester, discharge is performed through a pogo pin which can support a short rise-time less than 400 ps (typically, ~200 ps) [16], [22]. However, on-package VF-TLP test setup uses a single-ended probe with a poor ground return path, because the fine-pitch (~150 μm) coplanar waveguide RF probes, which are commonly used in on-wafer VF-TLP, cannot fit to the wide pin pitch (~0.5 mm) of the QFN package. This band-limited probe setup in VF-TLP leads to attenuated voltage overshoot compared to FICDM tester, which results in retarded gate-oxide breakdown (or, relatively earlier diode failure) in VF-TLP results.

TRX4 survives -1500 V CDM. The superior CDM performance of TRX4 was not predicted by the VF-TLP results because of early failure in ESD diode. It is also attributed to the shorter rise-time of the CDM pulse [22] as explained before. The rapid discharge causes transient voltage overshoot in the ESD diode (forward recovery) and the secondary protection responds even more quickly than AABC.

For all VF-TLP and CDM stresses, AABC enhances the protection level relative to that obtained using primary ESD protection, but it does not provide better reliability than secondary ESD protection. However, in applications which do not permit secondary protection, such as RF or very high-speed SerDes [3], AABC would be advantageous.

Table 2.6. Summary of HBM and CDM results.

Model \ TRX split	TRX1 (prim)	TRX2 ([6])	TRX3 (AABC)	TRX4 (prim+2 nd)
Positive HBM	$> 4\text{ kV}$			
Negative HBM	$< -4\text{ kV}$			
Negative CDM	-750 V	-1000 V	-1000 V	$< -1500\text{ V}$

Table 2.7. Failure analysis after negative CDM.

Items \ TRX	TRX1 (prim)	TRX2 ([6])	TRX3 (AABC)	TRX4 (prim+2 nd)
Failed items	•Elec. par.	•Elec. par.	•Elec. par.	-
Failure point	RX(NMOS)	RX(NMOS)	RX(NMOS)	-

2.8 Effectiveness of AABC

2.8.1 Revisiting ESD target levels

From the measurement results in the previous sections, the AABC technique is verified to guarantee better ESD reliability. In this section, the effectiveness of ESD resiliency by AABC will be investigated in the context of ESD target level study, which was conducted in Section 2.1 (Figure 2.2). With AABC,

drain-source overstress ($V_{DS,peak}$), mostly related to HBM performance, can be determined as follows in case of two-transistor-stacked drivers (refer to Figure 2.5(a)):

$$V_{DS,peak} = \max(VDD - V_{th,n}, V_{PAD} - VDD + V_{th,n}). \quad (2.2)$$

The induced VDD during ESD can be calculated as follows, where all parameter definitions are same as those in (2.1):

$$VDD = I_{ESD} (R_{clp} + R_{bus}) + V_{clp,on}. \quad (2.3)$$

The gate-oxide overstress, which is generally related to CDM reliability, can be assumed as 30% reduction by the aid of the AABC scheme, according to the measurement results. For this statement, it is further assumed that the gate-oxide breakdown mostly occurs between the gate and the source of the input transistors, which is simply a function of V_{PAD} . Thus peak gate-oxide voltage ($V_{ox,peak}$) can be calculated as follows with AABC:

$$V_{ox,peak} = 0.7 \times V_{PAD}. \quad (2.4)$$

The two expressions in (2.2) and (2.4) signify the actual voltage stresses at the drain-to-source and the gate-oxide of the vulnerable transistors, respectively; they are overlapped with the previous I-V curves in Figure 2.2, to show the effectiveness of the AABC scheme as given in Figure 2.13. The simulation parameters are exactly same as those in Figure 2.2. The additional dashed lines in Figure 2.13 represent I-V curves of the I/Os assisted by AABC. In HBM case (Figure 2.13(a)), both 5-Gb/s and 20-Gb/s I/Os can pass 1-kV HBM target with enough margin on both 130-nm and 45-nm process technologies. For CDM in Figure 2.13(b), the voltage margin for 5 Gb/s on 45-nm process is significantly improved, while 20-Gb/s link still cannot satisfy 250-V CDM target. The achievable CDM level for 20-Gb/s links on 45-nm process nodes is expected to be 125 V with the AABC technique.

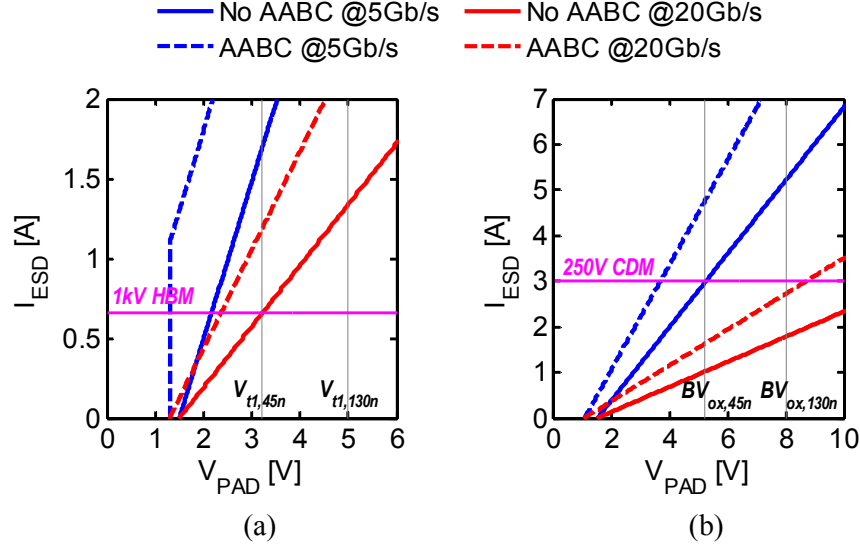


Figure 2.13. ESD current vs. peak PAD voltage for (a) HBM and (b) CDM. The solid lines show I-V characteristics without AABC (same as Figure 2.2), and the dashed-lines depict modified I-V characteristics by the aid of AABC. Peak PAD voltage, V_{PAD} in a given I_{ESD} is reduced for both HBM and CDM with the AABC technique to guarantee more ESD margins.

2.8.2 Dual-diodes resizing

From the results in the previous sections, it is expected that the primary ESD diode size can be reduced by the benefit of the improved CDM reliability, based on the simple assumptions that the gate-oxide breakdown is only caused by the V_{GS} overstress of transistors and ESD failure always occurs from the transistors, not from the other ESD devices. The induced PAD voltage, V_{PAD} , is composed by voltage drop at dual-diodes (V_{d-d}) and the other terms (V_o); V_{d-d} relates only to the series resistance portion which is inversely proportional to the diode perimeter, and the constant term (diode turn-on voltage, $V_{dio,on}$) is included in V_o . If contribution of dual-diodes on V_{PAD} is α ($0 < \alpha < 1$), V_{PAD} can be expressed as follows:

$$V_{PAD} = V_{d-d} + V_o = \alpha V_{PAD} + (1 - \alpha) V_{PAD}. \quad (2.5)$$

For the AABC scheme with a CDM level improvement by β (e.g., $\beta = 0.3$ for 30% CDM improvement), the allowable V_{PAD} (V_{PAD}') can be effectively increased:

$$V_{PAD}' = (1 + \beta) V_{PAD}. \quad (2.6)$$

If the size of dual-diodes is reduced by γ , the voltage drop at dual-diodes becomes $V_{d-d}/(1-\gamma)$, and V_{PAD}' can be expressed using the terms in (2.5) as follows:

$$V_{PAD}' = \frac{V_{d-d}}{1-\gamma} + V_O = \frac{1}{1-\gamma} \alpha V_{PAD} + (1-\alpha) V_{PAD}. \quad (2.7)$$

Equating (2.6) and (2.7), the dual-diode reduction factor, γ can be calculated as follows:

$$\gamma = \frac{\beta}{\alpha + \beta}. \quad (2.8)$$

For $\beta = 0.3$ from the results in Table 2.6., the dual-diode size can be reduced by the curve shown in Figure 2.14(a). For example, if dual-diodes contributes to one-half of V_{PAD} ($\alpha = 0.5$), the dual-diode size can be reduced by 35% ($\gamma = 0.35$), and when α becomes smaller, γ increases. Figure 2.14(b) gives the contour plot by varying both α and β ; the color bar in the right-hand side shows the γ values. Evidently, as α becomes smaller and β becomes larger, the dual-diode size can be reduced significantly. However, this analysis is an extremely simple case which ignores ESD failure at ESD protection devices. In reality, significantly reduced diode size can lead to earlier ESD failure than the transistors. Thus, the upper limit of dual-diode size reduction should also be determined considering the failure current of dual-diodes.

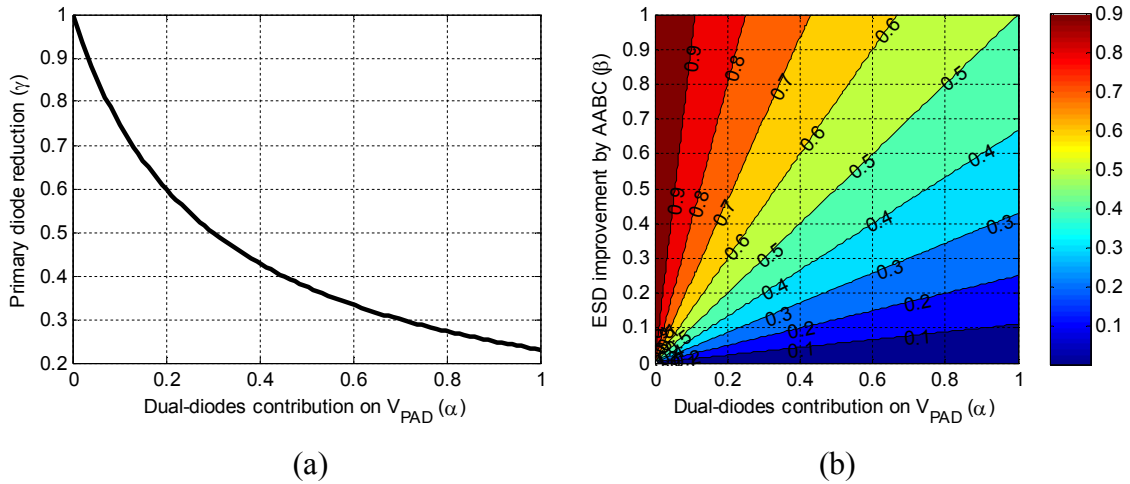


Figure 2.14. (a) Dual-diode size reduction (γ) with respect to the V_{PAD} contribution on dual-diodes (α), for 30% CDM improvement by AABC ($\beta = 0.3$). (b) The amount of dual-diode size reduction with different α and β .

2.8.3 Possible applications

The AABC techniques can be most effective for the transceiver circuits composed of stacked transistors. AABC cannot be applied for the transceivers with only a single transistor at each pull-up and pull-down network (e.g., a simple inverter composed of a PMOS and an NMOS). Even in the transceivers with stacked transistors, the benefit of AABC for the analog-like circuits will not be substantial compared to when AABC is used in the digital-style circuits where the gates of the transistors can be stuck to one of the power rails. Figure 2.15 is an example of a practical differential amplifier with the enable switches to support stand-by mode. The tail current transistor (MN3) and current mirror circuitry (MN4, MP1 and MP2) are shown, with the enable switches, MX1-5, which are controlled by the complementary signals, “EN” and “ENB.” During ESD, the AABC circuit forces ENB to VDD and EN to GND to pull the node V1 down to ground. However, even in the worst case, it is unlikely that the node V1 is stuck to VDD in this analog amplifier.

If a receiver is designed by only NMOS transistors, which is very typical for multi-gigabit receiver front-end circuits, the direction detection in AABC is not required. However, transmitters, especially for very high-speed serial links, are usually designed by voltage-mode push-pull drivers with CMOS transistors as shown in Figure 2.4. In this case, the direction detector in AABC is mandatory to protect both PMOS and NMOS transistors.

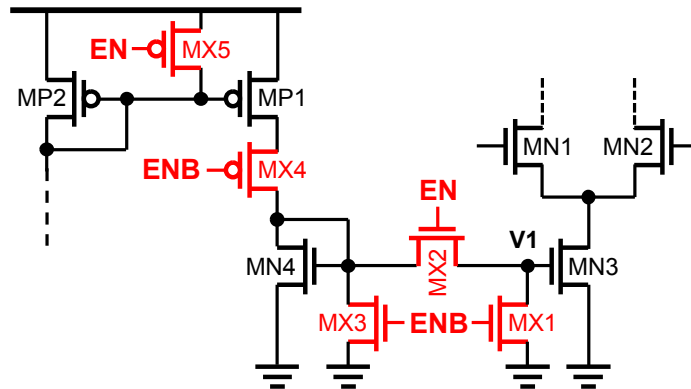


Figure 2.15. Schematic of a practical differential amplifier with enable transistors.

2.9 Summary

A novel, ESD polarity-aware protection technique, called Adaptive Active Bias Conditioning (or AABC), enhances the immunity of I/O circuits to voltage overstress. AABC was implemented and verified in a 130-nm CMOS process, where it was found to enhance the resilience of I/Os to VF-TLP and CDM stresses by 30%, relative to the case of primary ESD protection. With AABC, the transceiver front-end area increases by only 10% and signal integrity is unaffected. For ESD failure analysis, the power-on I-V curve measurement is proposed and applied to indicate exact failure points, without resorting to visual inspection.

High data-rate applications, e.g. over 20 Gb/s, require one to use both an advanced CMOS process and fairly small primary protection devices to achieve the necessary I/O bandwidth. In advanced process nodes, such as 65-nm and beyond, the transistor failure currents and gate-oxide breakdown voltages are reduced relative to 130-nm CMOS [23], and techniques such as AABC will be required to meet ESD targets. The need will be manifest first for CDM ESD but eventually for HBM, too, unless the targets are lowered below 1kV (refer to Figure 2.2).

Chapter 3. T-Coil Network for I/O Bandwidth Extension

This chapter presents a T-coil network, which utilizes inductive peaking for bandwidth extension, especially by canceling ESD device capacitance at I/O pads. The properties of different T-coil structures will be discussed. A potential ESD hazard of T-coils will be studied, and a new T-coil scheme to achieve both high bandwidth and ESD reliability will be proposed. To compare performance in the context of practical I/O circuits, 25-Gb/s wireline receiver is implemented.

3.1 Motivation

Speed limitation of amplifier circuits predominantly comes from the amount of capacitive loading in comparison with transistor's driving capability, especially in high-speed I/Os. In order to overcome bandwidth limitation of circuits, various bandwidth extension techniques are proposed. One of the viable techniques is “inductive peaking” [24] which utilizes moderately damped resonance (peaking) between inductors and load capacitors. The inductive peaking is a relatively old technique which has been used for several tens of years in high-speed circuits (e.g., oscilloscope circuitry) [25] with discrete inductors. However, by virtue of the practical implementation of planar spiral inductors in CMOS technologies, these techniques are intensively used in high-speed transceiver ICs nowadays [26], [27], [28], [29], [30], [31]. Four basic inductive peaking schemes are illustrated in Figure 3.1 [25], according to the relative position of inductors to load capacitors. The efficiency of bandwidth extension from these schemes depends on the ratio between self-capacitance (mostly, drain capacitance of a transistor, C_d) and load capacitance (C_L). Among the four techniques in Figure 3.1, bridged “T-coil” in Figure 3.1(d) shows the best bandwidth extension—ideally, $2.72\times$ in a maximally-flat group-delay condition, compared to an uncompensated RC network [31]. T-coil networks also have different variations according to load and driver/receiver positions. More detailed discussion on the various T-coil schemes will be given in Section 3.2.

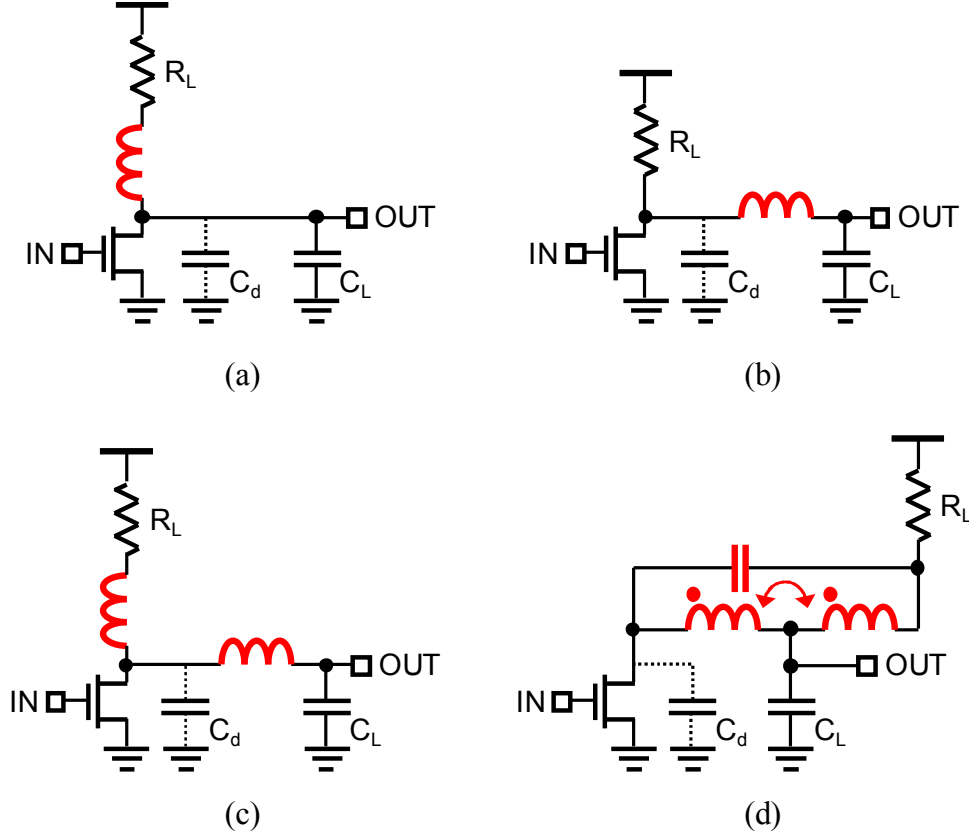


Figure 3.1. Various inductive peaking schemes: (a) shunt peaking, (b) series peaking, (c) shunt-and-series peaking, and (d) bridged T-coil network.

3.2 T-Coil Techniques

Figure 3.2(a) shows an equivalent circuit of a T-coil. The T-coil is a 3-port transformer, in which the primary (L_1) and secondary (L_2) windings of a transformer share a common node (P3). Usually, the required magnetic coupling factor (k) between the primary and secondary inductors is in a moderate range: around 0.5 for maximally flat group delay [25], [31], which can be easily obtained in silicon spiral transformers. As shown in Figure 3.2(b), T-coil can be implemented by a center-tapped spiral inductor to utilize lateral coupling for the required coupling factor. As the magnetic fields generated from L_1 and L_2 are in the same direction (i.e., both in out-of-page directions, as shown in Figure 3.2(b)), the dot positions of L_1 and L_2 should be matched with the current flow (from P1 to P2 via P3) as shown in Figure 3.2(a), according to the dot convention in a transformer [32]. The bridging capacitor (C_B) is an inherent capacitor in a spiral inductor, but for the better bandwidth, an additional capacitor can be placed between P1 and P2.

The T-coil has four design parameters: the primary and secondary inductance (L_1 and L_2), coupling factor (k), and bridging capacitor (C_B).

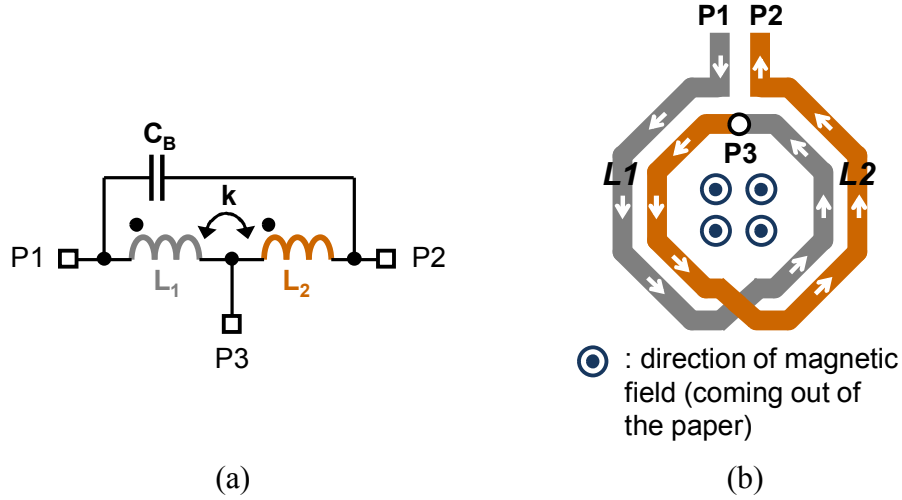


Figure 3.2. T-coil structure: (a) equivalent circuit and (b) layout with an on-chip spiral inductor. Two inductors, L_1 and L_2 , are magnetically coupled, as shown in the dot direction with a coupling factor, k . The bridging capacitor C_B represents an intrinsic parasitic capacitance, or intentionally added one.

From the recent researches, T-coils are adopted in the wireline and wireless transceiver circuits to nullify parasitic capacitance at an I/O pad, which is mainly contributed by the ESD protection circuit; it achieves broadband impedance matching and bandwidth extension [29], [30], [31], [33]. Figure 3.3 shows one of the examples of high-speed transceivers assisted by the T-coil. Typically P3 node is connected to a huge load capacitance such as ESD devices (dual-diodes DP and DN in Figure 3.3), and a termination resistor (R_T) is located at P2 node during normal operation but can be disabled when the I/O circuit is not in use. Improved impedance matching (or bandwidth extension) from T-coils can be explained by a transmission line analogy as shown in Figure 3.4 [34]. Magnetically coupled L_1 and L_2 are decoupled using mutual inductance, M , according to the dot convention [32], as shown in Figure 3.4; M is expressed as $M = k\sqrt{L_1 L_2}$. The parasitic capacitance at P3 from ESD diodes and circuits is lumped into C_{ESD} . To achieve perfect impedance matching, the T-coil parameters can be chosen to satisfy (3.1) to form an artificial transmission line with characteristic impedance, Z_0 (ignoring C_B for simplicity):

$$Z_0 = R_T \approx \sqrt{\frac{L_1 + L_2 + 2M}{C_{ESD}}}, \quad (3.1)$$

where R_T is a termination resistor, which will be located at P2. Thus, when (3.1) is satisfied, the incoming signal can pass through this T-coil network without any reflection, and C_{ESD} can be completely hidden from the signal path.

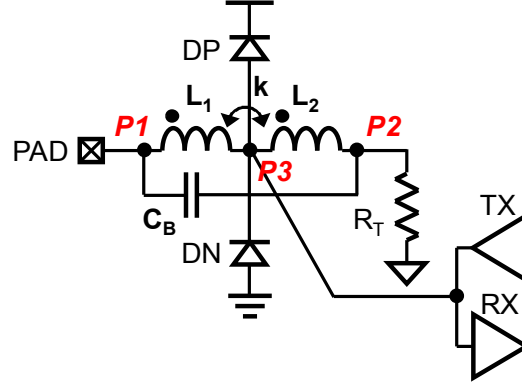


Figure 3.3. Example of a T-coil-assisted high-speed transceiver.

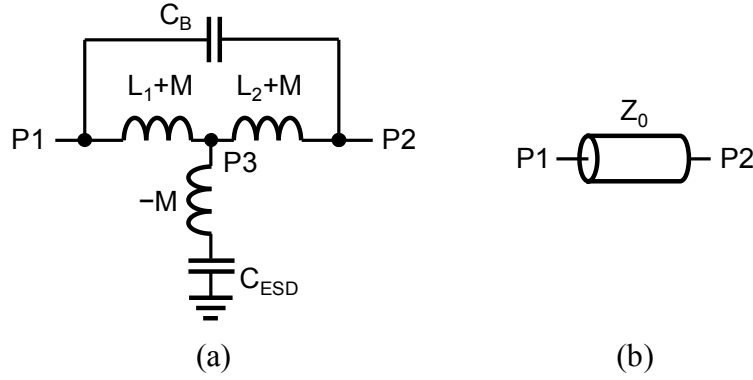


Figure 3.4. Transmission line analogy of T-coil network: (a) equivalent circuit of T-coil with decoupled inductors, and (b) transmission line with a characteristic impedance of Z_0 .

T-coils can be categorized into two ways, as shown in Figure 3.5, depending on the placement of the termination resistor relative to the transceiver circuit. In a conventional T-coil structure, denoted as T-coil1 [25], [29], [31], the termination resistance R_T and the receiver circuit or, more generally, the transceiver (TRX) are placed at different ports, as illustrated in Figure 3.5(a). The two versions of T-coil1 shown in the figure are equivalent, according to the reciprocity theorem [35]. In Figure 3.5(b), the TRX

and R_T are instead both placed at the same port; this structure is denoted as T-coil2 and was used in [30], [33]. Theoretically, T-coil2 is an all-pass system, which has a flat magnitude response in frequency domain [36], if the parasitic capacitance from the TRX is zero. However, some parasitic capacitance from the TRX can provide band-limitation and ripples on the pass-band in T-coil2, which needs careful parameter optimization. These two T-coil structures have different performance in terms of impedance matching and bandwidth extension, and should be selected according to the application requirement.

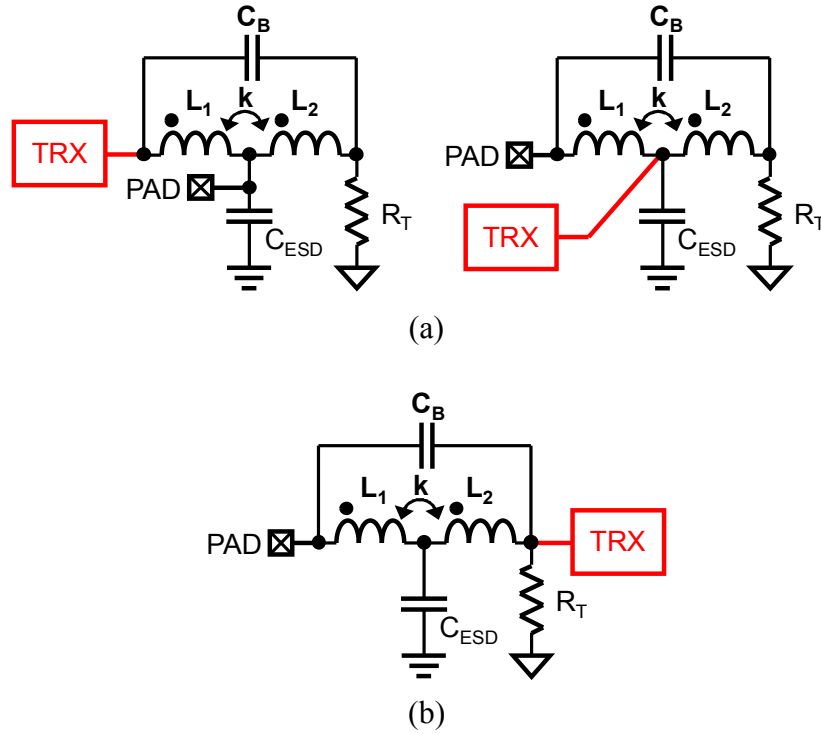


Figure 3.5. Various T-coil configurations. (a) T-coil1: circuit (TRX) and termination resistor (R_T) are at different ports, and (b) T-coil2: TRX and R_T are at the same port. Parasitic capacitance of primary ESD device is shown as C_{ESD} .

3.2.1 T-coil1 structure

The transfer function of the “bridged” T-coil1 structure in Figure 3.6 can be calculated by decoupling the two inductors, and then applying Y- Δ transform [37], [38]. Transimpedance is calculated for the input current (I_i) to the output voltage (V_o). The original transfer function has four poles and two zeros, but by pole-zero cancellation, the equation can be re-organized as shown in (3.2) [37],

$$Z_{\text{T-coil1,br}} = \frac{V_o}{I_i} = R_T \frac{1}{s^2 \frac{(C_{ESD} + C_{IN})^2 R_T^2}{4} \left(\frac{1-k}{1+k} \right) + s \frac{(C_{ESD} + C_{IN}) R_T}{2} + 1}. \quad (3.2)$$

The conditions for pole-zero cancellation is as follows (assuming symmetric T-coils, with $L_1 = L_2$) [37]:

$$C_B = \frac{(C_{ESD} + C_{IN})}{4} \frac{1-k}{1+k}, \quad L_1 = L_2 = \frac{R_T^2 (C_{ESD} + C_{IN})}{2(1+k)}, \quad k = \frac{4\zeta^2 - 1}{4\zeta^2 + 1}, \quad (3.3)$$

where ζ is a damping factor of the network. Four T-coil design parameters can be selected as given in (3.3), and dynamic response of the T-coil network can be determined by selecting proper ζ . For example, for the uniform group delay, $\zeta = \sqrt{3}/2$ (i.e., $k = 0.5$) can be chosen, and $\zeta = 1/\sqrt{2}$ (i.e., $k = 1/3$) leads to the maximally flat amplitude [25].

In case of the “unbridged” version of T-coil1 with $C_B = 0$, the transfer function can be expressed as follows:

$$Z_{\text{T-coil1,unbr}} = R_T \frac{1}{s^2 \frac{C_{ESD}^2 R_T^2}{2} \left(\frac{1}{1+k} \right) + s C_{ESD} R_T + 1}. \quad (3.4)$$

Comparing bandwidth between the bridged and unbridged versions of T-coil1, the bridged T-coil1 has a larger bandwidth.

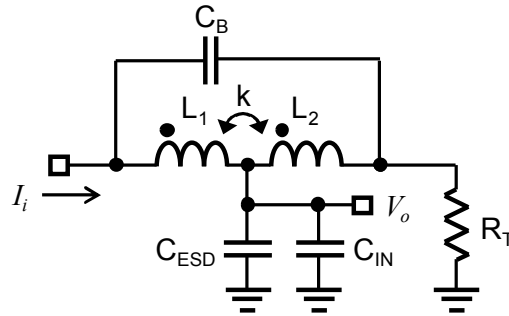


Figure 3.6. Equivalent circuit of the bridged T-coil1. The capacitance C_{IN} stands for the parasitic input capacitance of transceiver circuits.

3.2.2 T-coil2 structure

Figure 3.7 illustrates the bridged T-coil2 structure, and the transfer function is calculated with the same way used in Section 3.2.1:

$$Z_{\text{T-coil2,br}} = R_T \frac{s^4 B + s^2 [C_B (L_1 + L_2 + 2M) - MC_{ESD}] + 1}{s^5 A + s^4 B + s^3 C + s^2 D + sE + 1}, \quad (3.5)$$

$$\begin{cases} A = C_{ESD} C_B C_{IN} R_T (L_1 L_2 - M^2) \\ B = C_{ESD} C_B (L_1 L_2 - M^2) \\ C = R_T [C_{ESD} C_{IN} L_2 + C_B (L_1 + L_2 + 2M) \cdot (C_{ESD} + C_{IN})] \\ D = C_{ESD} L_2 + C_B (L_1 + L_2 + 2M) \\ E = R_T (C_{ESD} + C_{IN}) \end{cases} \quad (3.6)$$

The transfer function has four zeros and five poles, and unfortunately pole-zero cancellation is not possible, thus the T-coil design parameters should be first calculated from (3.3) to get the starting points of the design, and then minor tuning is needed to optimize the circuit performance.

The transfer function of the unbridged T-coil2 is calculated as follows:

$$Z_{\text{T-coil2,unbr}} = R_T \frac{-s^2 MC_{ESD} + 1}{s^3 C_{ESD} C_{IN} R_T L_2 + s^2 C_{ESD} L_2 + s R_T (C_{ESD} + C_{IN}) + 1}. \quad (3.7)$$

The zero locations of (3.5) and (3.7) are in the right-half plane (RHP); both are nonminimum-phase systems [39], which have initial undershoot and several ripples during signal transition in a step response. However, the unbridged T-coil2, (3.7) has only two RHP zeros, less than those of (3.5), which is expected to have better frequency- and time-domain responses, because more RHP zeroes cause phase distortion and ripples on the pulse transient. The AC magnitude and phase responses with various C_B are simulated and plotted in Figure 3.8. The simulation parameters are as follows; $L_1 = L_2 = 200$ pH, $k = 0.5$, $C_{ESD} = 200$ fF, $C_{IN} = 30$ fF, $R_T = 50$ Ω and C_B is a variable. Here, C_{IN} is only 30 fF, assuming receiver input capacitance. However, larger C_{IN} from transmitters can also be compensated by T-coils; C_{IN} of 600 fF was compensated by the T-coil2 structure in [30] for the source-series-terminated transmitters. As a baseline, “no T-coil” network is also simulated, which is composed of a termination resistor R_T and the

parasitic capacitors C_{ESD} and C_{IN} ; all frequencies are normalized to the 3-dB roll-off frequency of the no T-coil case. As shown in Figure 3.8(a), when C_B decreases, the ripples in the AC magnitude is reduced, and a larger 3-dB bandwidth is achieved. For $C_B = 0$ fF (unbridged T-coil2), the 3-dB bandwidth is lower than that of the case with $C_B = 10$ fF, but it guarantees better phase response as shown in Figure 3.8(b). Therefore, in T-coil2 case, the unbridged structure (or, minimal C_B) will be a better choice.

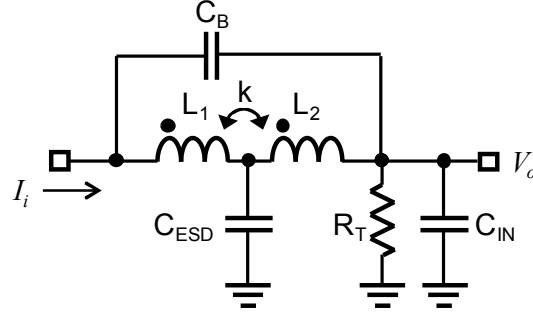


Figure 3.7. Equivalent circuit of the bridged T-coil2 for transfer function calculation.

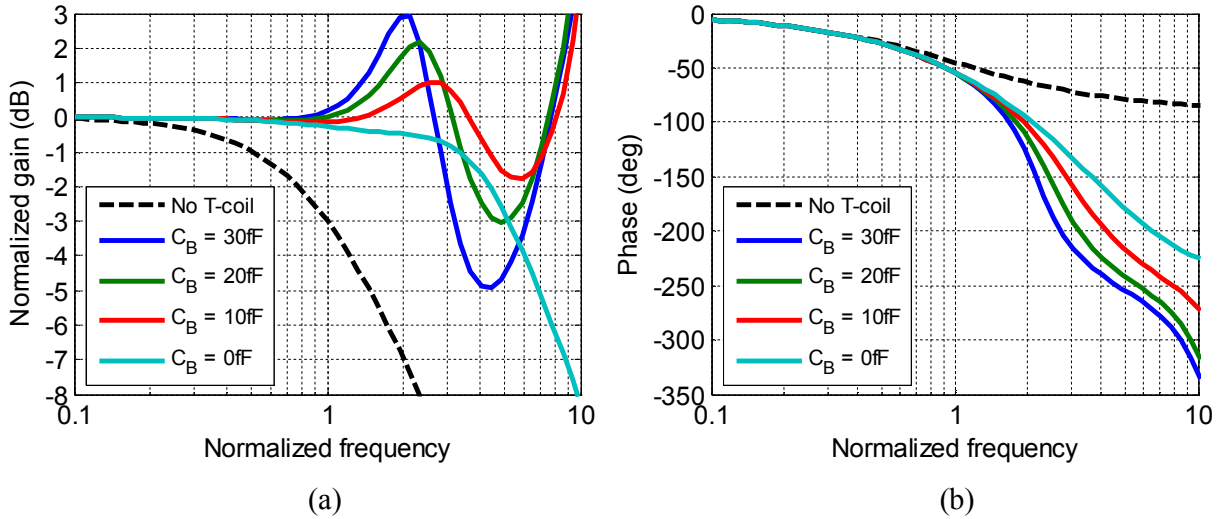


Figure 3.8. Comparison of magnitude and phase responses for various bridging capacitors (C_B) of the T-coil2 structure. (a) magnitude response, and (b) phase response. All frequencies are normalized to the 3-dB bandwidth of the no T-coil case.

3.2.3 Performance comparison

Using simulation, the two T-coil configurations are evaluated in terms of two performance metrics—bandwidth and impedance matching. The first is the 3-dB bandwidth as measured by the transimpedance

gain, and the latter is characterized in terms of return loss, S_{11} . Taking $C_{ESD} = 200$ fF, $C_{IN} = 30$ fF and $R_T = 50 \Omega$, each T-coil is designed to have the maximum bandwidth. The corresponding parameter values are $L_1 = L_2 = 200$ pH with a series resistance of 2Ω , and $k = 0.5$. The bridging capacitor C_B is 20 fF for T-coil1 and 0 fF for T-coil2. The simulation is performed with the equivalent circuits for T-coil1 (Figure 3.6), and T-coil2 (Figure 3.7). For benchmarking purposes, the front-end of a receiver without a T-coil (“No T-coil”) is also simulated. The results are plotted in Figure 3.9 in a normalized frequency to the 3-dB bandwidth of the “no T-coil” receiver. From Figure 3.9(a), it is observed that T-coil2 provides more bandwidth extension than T-coil1; T-coil2 extends bandwidth by $5.1\times$, while T-coil1 provides only $2.8\times$ bandwidth extension, which is, in fact, its theoretical limit [25], [31]. However, as seen in Figure 3.9(b), the S_{11} of T-coil1 remains below the -10 dB guideline over a wider frequency range than does that of T-coil2. This is not surprising; as shown in Figure 3.7, the parasitic capacitance of the TRX, C_{IN} , appears in parallel with R_T and results in degraded impedance matching. The S_{11} for those three circuits are plotted in the Smith chart as shown in Figure 3.10, with a frequency span from 1 GHz to 100 GHz. The trajectory of T-coil1 is always concentrated around the center of the circle (perfect matching point), and no T-coil shows the worst trajectory which moves from the center to the leftmost point (“short” point), as the frequency increases.

Therefore, the selection of a T-coil configuration should be made based on the target application. In the case of broadband wireline I/Os (e.g. backplane transceivers), maintaining a flat gain response up to at least one-half baud-rate is relatively more important than impedance matching, because a frequency-dependent gain will increase the intersymbol interference (ISI), which is a major concern for wireline communication links. Thus, the T-coil2 topology is better suited for wireline transceivers.

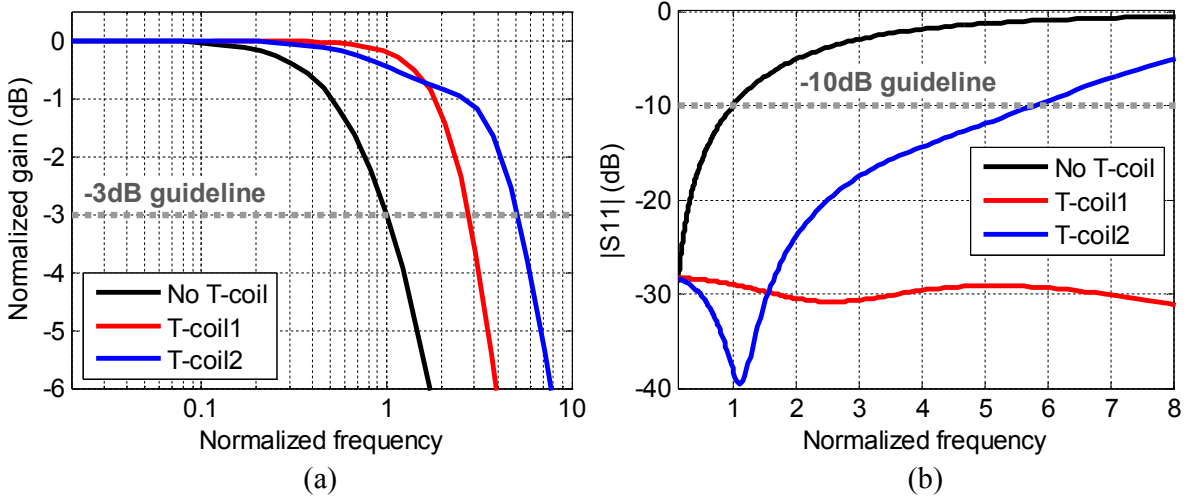


Figure 3.9. Performance comparison between T-coil1 and T-coil2 structures: (a) bandwidth from the normalized transimpedance gain, and (b) impedance matching (S_{11}). As a baseline, the characteristics of an input without a T-coil is also plotted. All frequencies are normalized to the no T-coil case.

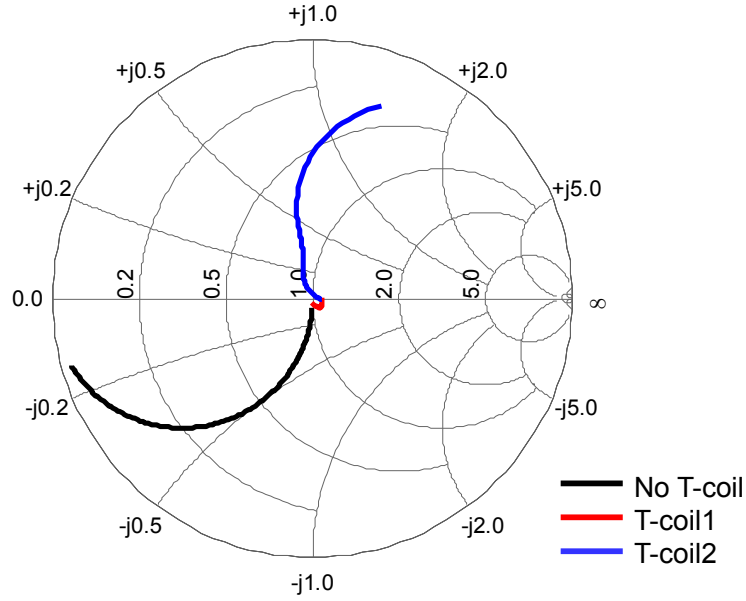


Figure 3.10. Simulated S_{11} on the Smith chart. The frequency span is from 1 GHz to 100 GHz.

3.3 CDM Hazard in T-Coil

T-coil2 has been reported to reduce CDM robustness in 90-nm CMOS process [40]. In [40], an equivalent CDM level is evaluated using a 2 ns VF-TLP pulse, and 3.8 A failure current was measured, which is roughly equivalent to 300 V CDM according to [40]. In a 90-nm CMOS process, the required

CDM level is 500 V [3], which manifests that T-coil2 may have a potential CDM hazard. The source of the problem must be identified and then mitigated by design.

Very-fast TLP (VF-TLP) is often used to assess CDM reliability. Of course, VF-TLP is a two-pin test [20], while CDM is a one-pin discharge [16]; thus the current flow and voltage build-up during the two tests cannot be exactly the same. However, large voltage transients during VF-TLP are indicative that excessive over-voltage stress may arise during CDM. Since the CDM current and the induced stress are determined by the IC package size, VF-TLP is better suited to a study of the intrinsic robustness of an I/O circuit. In order to study the CDM reliability in a receiver which uses T-coil2, a VF-TLP current pulse is applied at the input pad in simulation as shown in Figure 3.11. The pulse width (t_w) is 1 ns, the rise time (t_r) is 100 ps, and the current (I_{peak}) is 5 A. The current pulse amplitude was selected to be similar to the peak current obtained during 250 V CDM testing of an IC in a 1000 mm² BGA package [3]. Two cases are simulated: positive I/O pad stress with respect to VSS (PS mode), and negative I/O pad stress with respect to VDD (ND mode). The T-coil parameters and parasitics are the same as in Section 3.2.3.

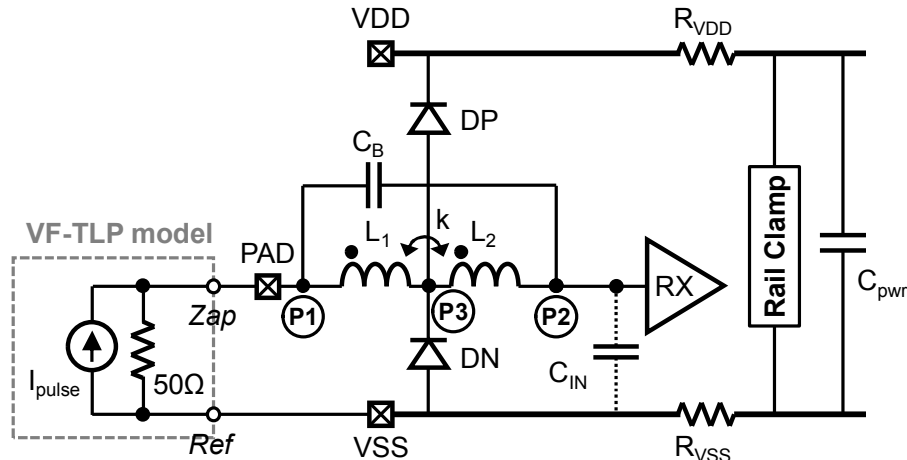


Figure 3.11. PS-mode VF-TLP simulation setup in a typical wireline receiver with a T-coil and an ESD protection network based on dual-diodes and rail clamps. In case of ND mode, reference node can be connected to VDD, with inverse current pulse polarity at PAD.

PS-mode VF-TLP waveforms are plotted at P1 (PAD), P3 (dual-diodes), and P2 (receiver) nodes, as shown in Figure 3.12. At P1, voltage overshoot occurs at the beginning of the incoming current pulse, and voltage undershoot is observed at the end of the current pulse. P2 node does not have any voltage peaking,

while P3 shows voltage peaks with opposite direction to those of P1. Voltage overshoot at P2 shown in Figure 3.12(c) can potentially compromise CDM reliability by causing gate-oxide breakdown at the receiver input transistors.

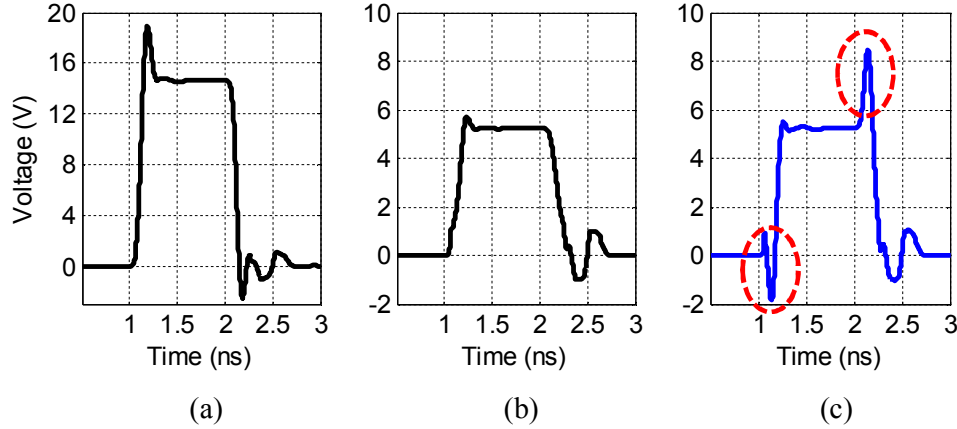


Figure 3.12. Voltage waveforms at (a) P1, (b) P3, and (c) P2, for PS-mode VF-TLP.

In order to further investigate this phenomenon, waveforms at the receiver input node (P2) are plotted with different coupling factors (k) and input current directions, as shown in Figure 3.13 and Figure 3.14. Figure 3.13 provides the waveforms for PS-mode VF-TLP and Figure 3.14 are for ND-mode; the parameter k is also varied in the simulations, as specified in the figure caption. Figure 3.13(a) is specific to the case of PS-mode VF-TLP and a positive magnetic coupling factor k , which is identical to Figure 3.12(c). For $k = 0$ (Figure 3.13(b) and Figure 3.14(b)), no voltage peaks are observed. Negative k ($k = -0.5$ in Figure 3.13(c) and Figure 3.14(c)) and positive k ($k = 0.5$ in Figure 3.13(a) and Figure 3.14(a)) produce opposite voltage peaking. Moreover, the waveforms in Figure 3.13 are identical to those in Figure 3.14, if flipped over the horizontal axis. These observations imply that magnetic coupling between the primary winding (L_1) and the secondary (L_2) induces voltage peaking at the secondary winding's output, with the peak being dependent on the amplitude and polarity of both the magnetic coupling factor and the input current pulse. Indeed, this is the expected behavior of a transformer, as will be demonstrated next.

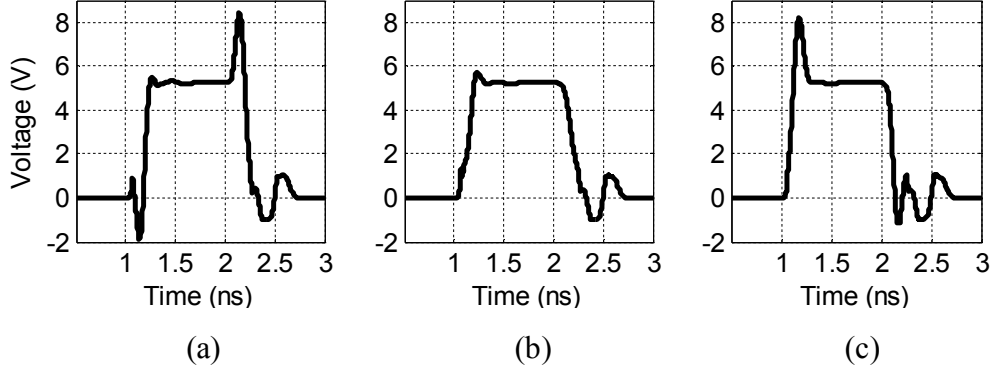


Figure 3.13. Voltage waveforms at P2 (receiver input) for PS-mode VF-TLP: (a) $k = 0.5$, (b) $k = 0$, and (c) $k = -0.5$.

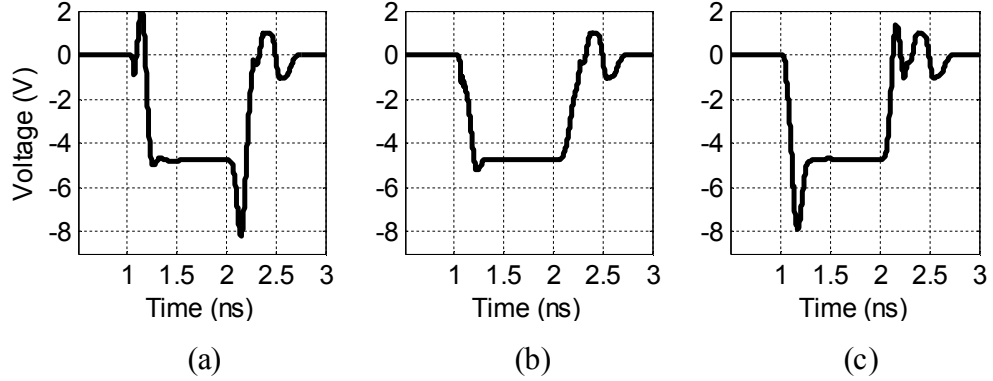


Figure 3.14. Voltage waveforms at P2 (receiver input) for ND-mode VF-TLP: (a) $k = 0.5$, (b) $k = 0$, and (c) $k = -0.5$.

Equivalent circuit of T-coil during ESD is redrawn in Figure 3.15. At normal operation, dual-diodes are turned off, thus P3 is in high impedance, whereas during ESD one of dual-diodes turns on to make P3 node low impedance (see the ground symbol at P3 in Figure 3.15(b)). The VF-TLP current pulse can be decomposed into an AC current during the pulse transition, and a quasi-static (DC) current at the pulse plateau. In a transformer, the voltage at the secondary side is induced only by a changing magnetic field, which is generated by AC current. The induced AC voltages at node P1 (V_1) and node P2 (V_2) are functions of the independent AC currents into L1 (I_1) and L2 (I_2). The relation can be expressed as follows in a matrix form:

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = s \begin{bmatrix} L_1 & -M_{12} \\ -M_{21} & L_2 \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \end{bmatrix}. \quad (3.8)$$

In (3.8), M_{12} and M_{21} are mutual inductances; in a passive transformer, these are given by

$$M_{12} = M_{21} = k\sqrt{L_1 L_2}. \quad (3.9)$$

The preceding analysis assumes an ideal transformer without series resistance. During an ESD event, independent AC current I_2 is 0, because the ESD current is injected only at the primary side, i.e., I_1 . Solving (3.8) with $I_2 = 0$ and then taking the inverse Laplace transform, the AC voltage at P2 ($v_2(t)$) is found to be

$$V_2|_{I_2=0} = -sM_{21}I_1 \xrightarrow{\text{Inverse Laplace transform}} v_2(t) = -M_{21} \frac{di_1(t)}{dt}. \quad (3.10)$$

The negative sign on the right-hand side of (3.10) is consistent with the simulation results. For example, see the dashed circles in Figure 3.12(c): for a positive k , i.e., positive M_{21} , undershoot occurs at the rising input edge ($di_1(t)/dt > 0$) and overshoot occurs at the falling edge ($di_1(t)/dt < 0$).

Figure 3.16 illustrates current flow during PS-mode ESD, and the corresponding voltages induced by magnetic coupling, assuming $k > 0$. Quasi-static current flows from PAD to the top diode DP via L_1 . AC current flows from L_1 to L_2 and the polarity of the induced voltage is marked in Figure 3.16(a) using + and – signs. Superimposing the AC voltage ($V_{AC,P2}$) waveform on the quasi-static voltage ($V_{QS,P2}$) yields the composite voltage waveform shown in Figure 3.16(b), which is very similar to that in Figure 3.12(c) (or Figure 3.13(a)).

From (3.10), it may be concluded that in order to reduce the peak voltage overshoot, the mutual inductance (M_{21}) must be reduced, because $di_1(t)/dt$ has a fixed value for a given CDM event. Usually in a passive transformer, the mutual inductance is not tunable, because it is determined by geometry of the transformer. However, in this dissertation, a new mutual inductance tuning technique for an on-chip passive transformer is proposed.

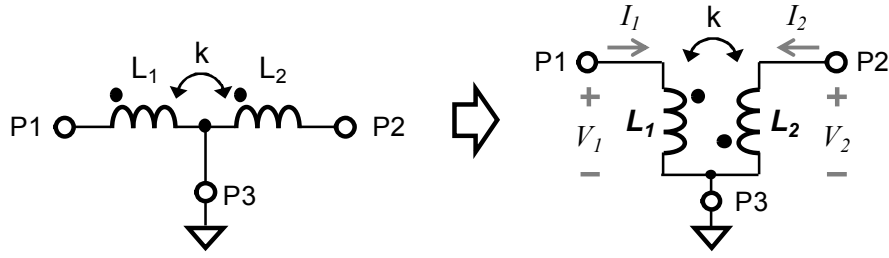


Figure 3.15. T-coil equivalent circuit for transformer equation during ESD.

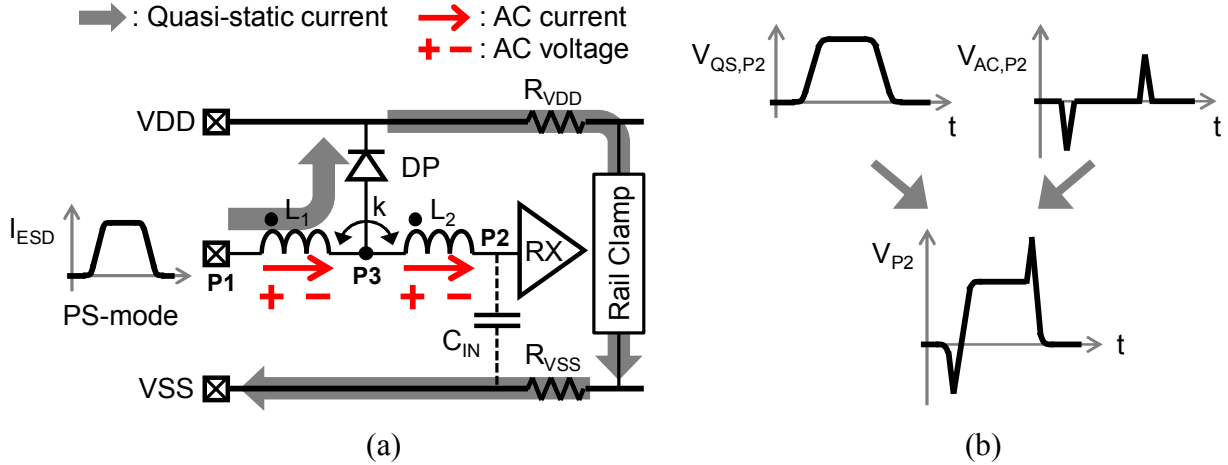


Figure 3.16. Illustration of current flow and induced voltage at T-coil during PS-mode ESD. Unnecessary components such as the bottom diode (DN) and termination resistor (R_T) are not shown for simplicity. (a) Quasi-static current and AC current, and (b) composite ESD-induced voltage at the receiver input (V_{P2}) which is composed of quasi-static voltage ($V_{QS,P2}$) and peak AC voltage ($V_{AC,P2}$).

3.4 CDM-Reliable T-Coil

3.4.1 Inductance-halving

To reduce the likelihood of receiver gate dielectric breakdown during CDM, the T-coil's mutual inductance should be reduced during the ESD event. Mutual inductance reduction can be accomplished by self-inductance or coupling factor reduction. Self-inductance tuning was used previously in LC-tuned voltage-controlled oscillators (VCO), where it can increase the frequency tuning range [41], [42]. Direct inductor switching was introduced in [41]; it provides a wide inductance tuning range by short-circuiting intermediate nodes of a spiral inductor through switches, although the parasitic capacitance from the switches would be a concern for a T-coil application. This dissertation uses direct inductor switching;

here, the switches are actuated by an electrostatic discharge, and the switches are designed to minimize bandwidth degradation. For completeness, it is noted that in [42], inductance is tuned by turning on extra coupling inductors to change the magnetic coupling between main and extra inductors. However, the achievable inductance tuning range is limited by a bandwidth performance. This will be more discussed in Section 3.4.4.

Therefore, the “inductance-halving” T-coil is proposed to utilize direct switching, as conceptually illustrated in Figure 3.17. During ESD, it reduces the self-inductance on the primary and secondary sides by partially short-circuiting L_1 and L_2 with the switches $S1$ and $S2$. Furthermore, it decreases the coupling factor k during ESD by increasing the physical separation between the residual L_1 and L_2 . Both of these effects serve to reduce the mutual inductance, as indicated by (3.9). The best self-inductance reduction may be achieved by connecting $S1$ between $P1$ and $P3$, which may lead to “inductance-zeroing”. However, the increased parasitic capacitance on $P1$ by the $S1$ switch will limit bandwidth of the T-coil network because the parasitic capacitance on $P1$ cannot be canceled by T-coils. Moreover, a huge voltage drop between $P1$ and $P3$ during ESD may break $S1$ if the switch device is not properly designed; larger $S1$ will be needed, which can further decrease bandwidth of T-coils.

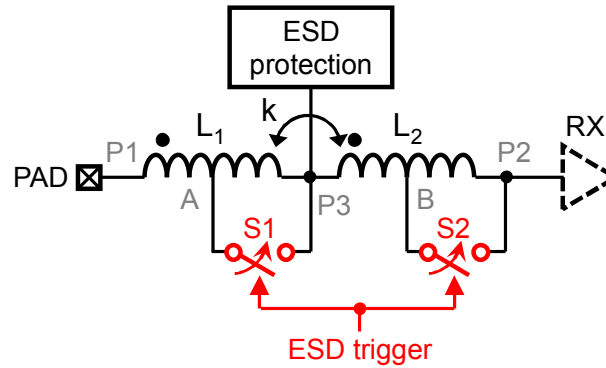


Figure 3.17. The concept of inductance-halving T-coil. Two switches ($S1$ and $S2$) turn on during ESD to short out part of the inductors on the primary and secondary sides to reduce mutual inductance. RT is not shown because it is enabled only during normal operation, and has no effect during power-off ESD.

The actual implementation of the T-coil with inductance halving is shown in Figure 3.18. In the figure, intermediate points of the primary inductor ($L_{11}+L_{12}$) and the secondary inductor ($L_{21}+L_{22}$) are

marked as A and B, respectively. The ideal switches in Figure 3.17, S1 and S2, are replaced by a P+/N-well diode (D1) and a PMOS switch (M1), respectively; these turn on only during an ESD event. S1 is implemented as a diode because it lies on the main ESD current path and must be able to handle a large current. S2 is implemented as a PMOS transistor because nodes B and P2 will be larger than VDD during PS-mode ESD. The circuit is designed specifically to mitigate voltage overshoot during PS-mode ESD, which is similar to negative CDM in regards to the polarity of the ESD current at the pin. High-speed receiver front-end amplifiers are generally designed with only NMOS transistors and resistive loads. The PAD-to-VSS potential difference during PS-mode ESD is much higher than during ND-mode, thereby stressing the gate of the input NMOS more severely. During ND-mode ESD, the proposed circuit does not create more voltage stress than does the normal T-coil.

During PS-mode ESD, the voltage at node A is sufficiently larger than at P3 and thus D1 turns on. M1 also turns on as it is triggered by the RC-timer-based trigger circuit shown in Figure 3.18(b). As noted earlier, turning on these switches reduces L_1 , L_2 and k , all of which contribute to lowering the mutual inductance; refer to (3.9).

The RC time-constant of the circuit in Figure 3.18(b) needs to only cover the short CDM time-scale (a few nano-seconds), so the RC timer layout size is small enough to be included in T-coil network. During PS-mode ESD, the voltage at node B can be larger than VDD; to bias the N-well of M1 properly, the N-well bias generator illustrated in Figure 3.18(c) is used to select the larger of VDD and V_B . The resistor between VDD and the N-well, R2, sets V_{NW} to VDD during normal operation and power-up; this resistor is made large (about 80k Ω in this design) so as not to interfere with M2's operation during ESD. Figure 3.19 provides cross-sectional views of D1 and M1 with its N-well bias generator.

The sizing of M1 and D1 involves a trade-off between voltage overshoot and bandwidth. A larger switch will more effectively reduce the mutual inductance, but it may restrict bandwidth by adding more parasitic capacitance. The optimal sizes are found with the aid of simulation; the full design space was explored by sweeping the sizes of D1 and M1 iteratively. All simulations are performed using a process

design kit (PDK) for a 65-nm CMOS technology. Figure 3.20 shows the simulated values of peak voltage and bandwidth when the size of D1 or M1 is varied, and the other is held near its optimal value. In Figure 3.20(a), the D1 perimeter is varied while the width of M1 has a fixed value of 300 μm , and Figure 3.20(b) shows the results obtained when the M1 width is varied and the D1 perimeter is fixed at 25 μm . The M1 channel length was set to the minimum value. Based on these results, D1 is sized at 25 μm perimeter, and M1 is sized at 300 μm width.

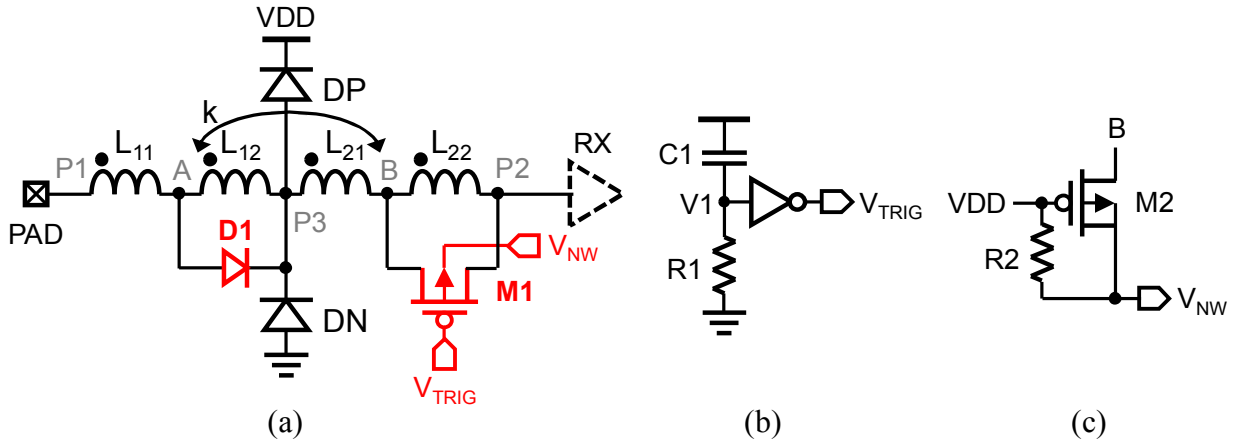


Figure 3.18. Schematic of the proposed T-coil with inductance-halving: (a) T-coil with a diode (D1) and a switch (M1), (b) trigger voltage generator, and (c) well-bias generator for M1.

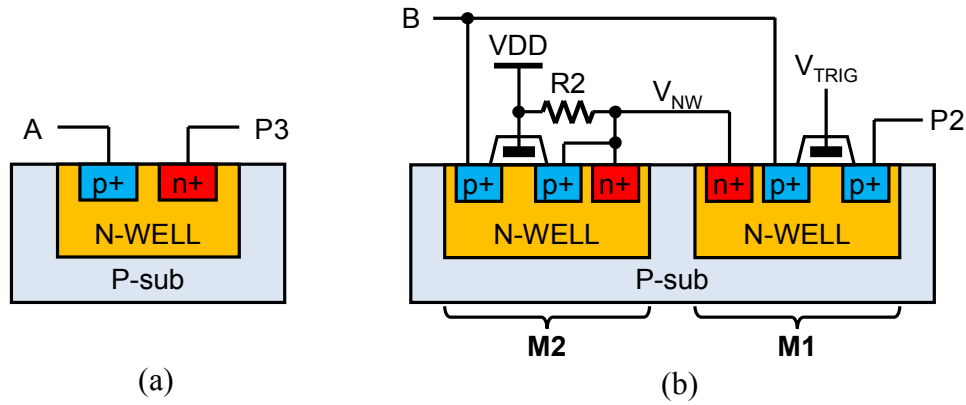


Figure 3.19. Cross-sections of (a) P+/N-well diode (D1), and (b) PMOS switch (M1) and N-well bias generating transistor (M2).

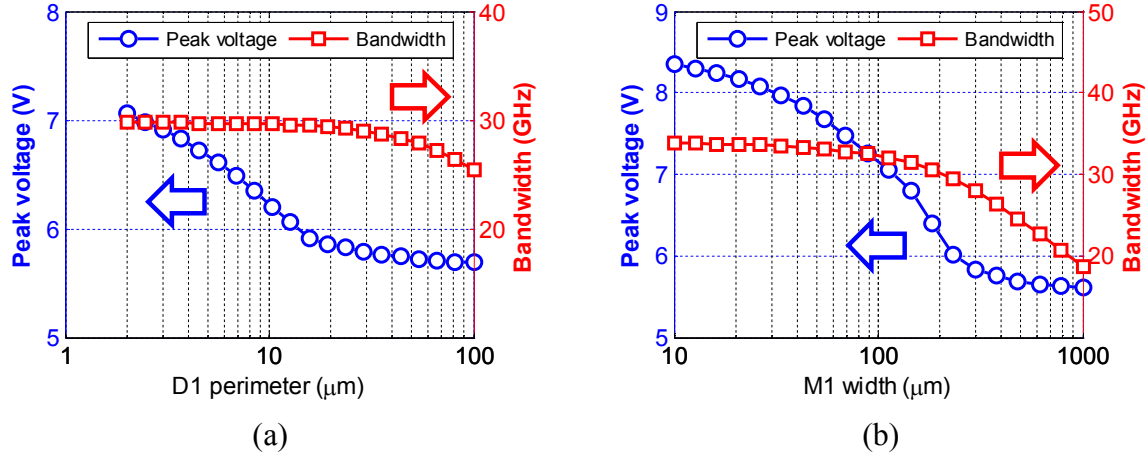


Figure 3.20. Simulated peak voltage (circle symbol) and bandwidth (square symbol) with respect to (a) D1 perimeter (M1 has a fixed width of 300 μm), and (b) M1 width (D1 perimeter is fixed at 25 μm).

Figure 3.21 presents the simulated current waveforms in each component of the inductance-halving T-coil when the PAD pin is subjected to 5 A VF-TLP with 1 ns pulse width and 100 ps rise-time. In Figure 3.21(a), the ESD current in L_{11} is observed to branch out into D1 and L_{12} , and the two branch currents merge back together in the primary ESD protection device DP. Over most of the pulse-width, the current through the 25 μm -perimeter D1 is 2.6 A. Based on information given in the PDK, the failure current (I_{f2}) of a 25 μm P+/NW diode is about 1.1 A when measured with 100ns TLP; for 1 ns VF-TLP, the I_{f2} should be more than double, and thus a 5 A stress at the pin is expected to be sustainable. Figure 3.21(b) and (c) show that only AC current flows through the secondary winding to the receiver. At the receiver (Figure 3.21(c)), the current is highly attenuated, which indicates that the voltage overshoot will be greatly mitigated.

3.4.2 T-coil design and layout optimization

The desired T-coil parameters are obtained using the optimization program in Advanced Design System (ADS) [43]. The ESD parasitic capacitance (C_{ESD}) to be compensated by the T-coil is 280 fF from the 300 μm dual-diodes in 65-nm CMOS PDK; each diode has a 300- μm perimeter. The termination resistance (R_T) is 50 Ω , and input capacitance of the receiver (C_{IN}) is 30 fF (see the equivalent circuit in Figure 3.7).

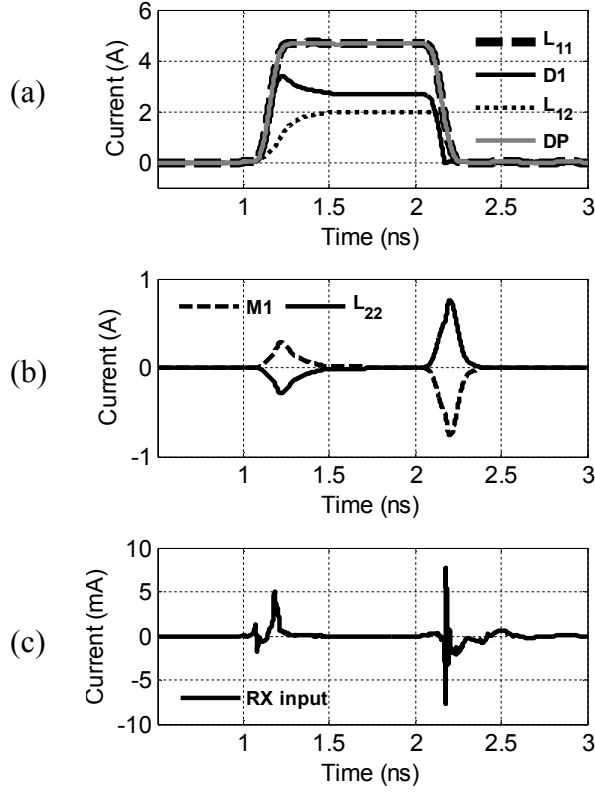


Figure 3.21. Current waveforms for 5A VF-TLP pulse with $t_w = 1$ ns and $t_r = 100$ ps. (a) Currents on the primary side, (b) currents on the secondary side, (c) current entering the receiver.

In this dissertation, an asymmetric T-coil is used, in which the primary and secondary inductances are different ($L_1 \neq L_2$) [24], [40]. The greater design freedom provided by an asymmetric T-coil makes it easier to optimize the AC gain response, i.e., adjust the pole and zero frequencies to minimize ripples in the passband and linearize the phase, as is needed to not exacerbate the ISI. The goals for the optimization by ADS are set up as sufficient bandwidth, the minimal peaking on the passband in the AC magnitude response, and fast transient response. By several iterations, the most optimal T-coil parameters are $L_1 = 430$ pH, $L_2 = 288$ pH, $k = 0.52$, and $C_B = 5$ fF, which are summarized in Table 3.1.

Table 3.1. The desired T-coil parameters for compensating $C_{ESD} = 280$ fF and $C_{IN} = 30$ fF.

L_1	L_2	k	C_B
430 pH	288 pH	0.52	5 fF

From the target parameters of the base T-coil, the layout of the inductance-halving T-coil should be optimized using combined electromagnetic (EM) and circuit simulations to determine the proper spiral inductor geometry and tapping points (A and B). The design objectives are sufficient mutual inductance reduction during ESD, minimal bandwidth degradation during normal operation, convenient routing to the receiver circuit, and small area consumption. The T-coil is based on a center-tapped symmetric inductor in an octagonal shape with an 80 μm outer dimension (OD), 3 μm spiral width (W) and 2.4 μm turn-by-turn spacing (SP). The number of turns (N) and the center-tap point are modified, resulting in the asymmetric T-coil. To shield the on-chip spiral inductors from the lossy silicon substrate, a typical method is to use the “patterned ground shield” [44], which is a hatched shield pattern in lower-metal-layers (usually in metal1 or gate-poly) with a perpendicular direction to the current flow of the main spiral inductor, to minimize eddy current generation [45] on the metal shield. With the patterned ground shield, the quality factor (Q-factor) of the spiral inductor can be improved, but with a cost of the reduced self-resonant frequency (SRF) because of the increased vertical parasitic capacitance [44]. When a spiral inductor is used as a bandwidth extension purpose, extremely high Q-factor is not necessary. Hence, in this design, a “moat” layer [46] is instead placed under the inductor to increase the substrate resistance; even though this is a less effective shielding method than the patterned ground shield, this does not lower the SRF, which is critical for wideband operation.

There are additional layout considerations when an on-chip inductor is incorporated into the on-chip protection network. The primary inductor is a conduit for ESD current. Failure may occur at metal layer transition points at an underpass [33], [47]. In the 65-nm CMOS process technology for this design, the top three metal layers (pad metal, metal7 and metal6) are used to form the T-coil. The primary inductor is laid out with only a top metal layer (metal7) and lower-metal underpass is avoided. Very thick pad metal and large pad-connecting vias are used to tap out ports on the primary side (P1 and A in Figure 3.18(a)).

The resistance of the P3 tap should be minimized to avoid additional voltage build-up when the ESD current is shunted through DP, because this voltage will appear at the receiver input (P2) as an elevated

quasi-static voltage. The spiral inductor layout of Figure 3.22(b) was developed to reduce the resistance between the center-tap and P3. For comparative purposes, an ordinary spiral inductor directly generated from the PDK is shown in Figure 3.22(a); its winding direction from P1 to P3 spirals from the outside to the inside of the structure. For this layout, the parasitic resistance from the center-tap point (marked by a white circle) to node P3 will increase as the dimension of the spiral inductor increases. However, if the spiraling direction is reversed as shown in Figure 3.22(b) (an “inside-out” style), the distance from the center-tap to P3 is shortened, and the quasi-static voltage at the receiver input P2 will be reduced. Another benefit for the proposed inside-out layout is freedom to utilize higher-level metal routing. In the conventional inductor in Figure 3.22(a) (an “outside-in” style layout), P3 is laid out by only metal5. However, in the proposed layout in Figure 3.22(b), P3 is connected through metal6 and metal7 in parallel to further decrease the resistance of the P3 tap. From EM simulations, the tap resistance from the center-tap to P3 of the proposed layout is only $0.195\ \Omega$, which is $6\times$ smaller than that of the original PDK inductor, $1.18\ \Omega$. Figure 3.23 shows the 3 A VF-TLP waveforms at P2 with $t_w = 1\ \text{ns}$ and $t_r = 100\ \text{ps}$ from simulation. Evidently, the decreased tap resistance of the inside-out layout is effective to reduce the quasi-static voltage at P2 by 2.5 V, compared to the conventional outside-in layout. The calculation method of the tap-resistance is explained in Appendix B in more details.

From these design considerations, the most optimal T-coils are designed and laid out, as shown in Figure 3.24(a) for the three-port normal T-coil, and Figure 3.24(b) for the five-port inductance-halving T-coil. As P1 and P3 nodes will be connected to the bonding pad and dual-diodes in the chip exterior, intermediate point in the primary inductor (A) should also be pulled out to the same direction. Nodes in the secondary inductor, P2 and B, should be faced to the chip core. The dimension of the T-coil is also indicated in Figure 3.24(a): $OD = 80\ \mu\text{m}$, $W = 3\ \mu\text{m}$, $SP = 2.4\ \mu\text{m}$, $N_1 = 2\ 1/8$ turns and $N_2 = 1\ 3/8$ turns; N_1 and N_2 are the number of turns for the primary and secondary side, respectively. Figure 3.25 shows the 3D-rendered layout of the optimized inductance-halving T-coil. As seen in the figure, all ports including

additional tapping points (A and B) are located at the preferable positions for easier routing to the corresponding switch devices, D1 and M1.

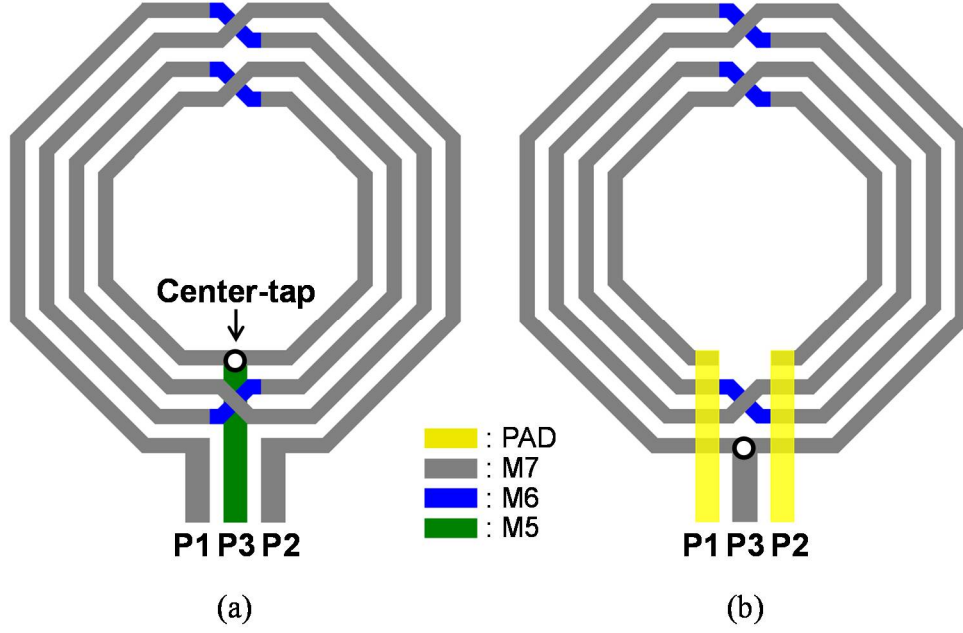


Figure 3.22. Two types of the center-tapped spiral inductor layout: (a) a normal 4-turn inductor layout generated by PDK, and (b) a new “inside-out” layout, which reverses the spiraling direction to shorten the routing length of P3.

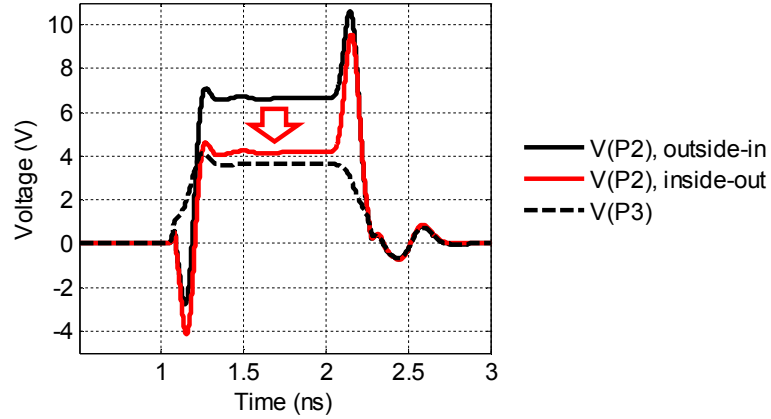


Figure 3.23. Simulated waveforms from the two different T-coil layout styles in Figure 3.22 for 3 A VF-TLP with $t_w = 1$ ns, $t_r = 100$ ps. By the proposed inside-out layout, the quasi-static voltage level at P2 is decreased by 2.5 V.

From this layout, multi-port S-parameters are extracted using EM simulations with Momentum [48], and then the T-coil parameters are calculated using the equations derived in Appendix B. The designed T-

coil parameters are $L_1 = 449$ pH, $L_2 = 320$ pH, and $k = 0.47$ at 10 GHz, as plotted in Figure 3.26. No explicit bridging capacitor (C_B) is added, but intrinsic parasitic capacitance is indeed included, as confirmed by the S-parameters obtained from EM simulations of the T-coil; C_B is calculated as 11 fF from the SRF. The designed T-coil parameters are summarized in Table 3.2; the actual parameter values are very similar to the desired parameters as given in Table 3.1.

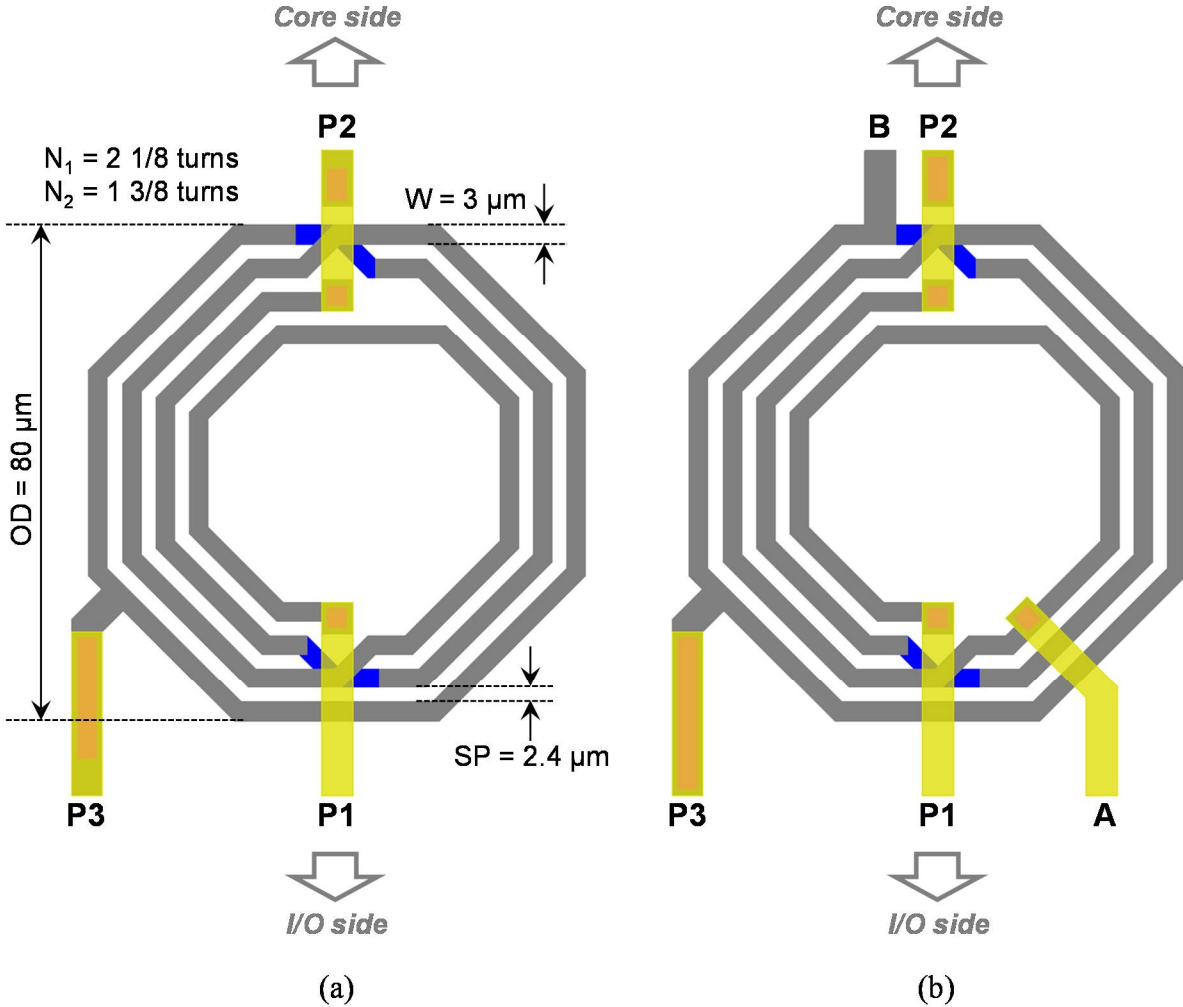


Figure 3.24. T-coil layout: (a) 3-port normal T-coil with geometry parameters (OD for the outer dimension, W for the width, SP for the turn-to-turn spacing, and N_1 for the total number of turns of the primary side, and N_2 for the turns of the secondary side), and (b) 5-port inductance-halving T-coil (yellow: pad metal, gray: $4\times$ top metal, blue: $1\times$ second top metal). Chip outer side is indicated as “I/O side”, where I/O pad cells with ESD devices are located, and “Core side” represents chip core area where receiver circuitry is located.

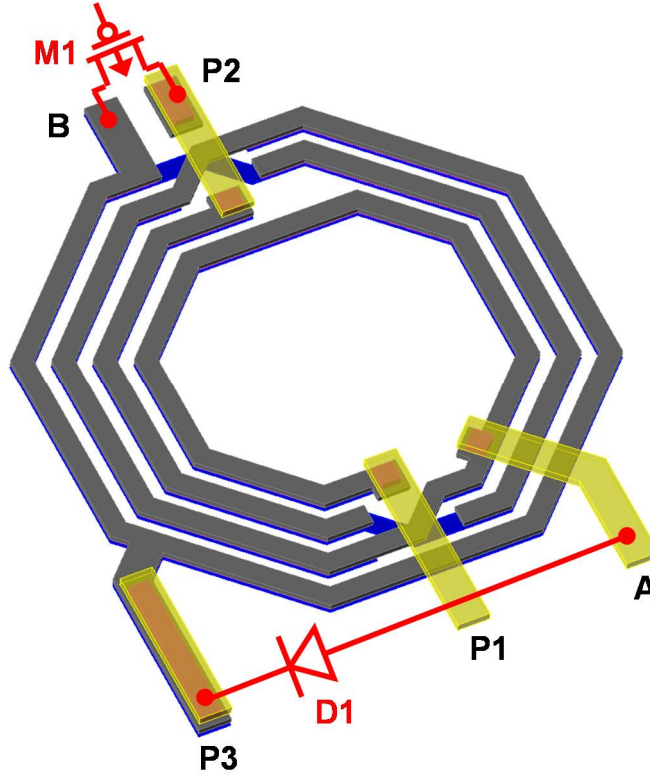


Figure 3.25. 3-D rendering of the proposed inductance-halving T-coil (yellow: pad metal, gray: 4× top metal, blue: 1× second top metal).

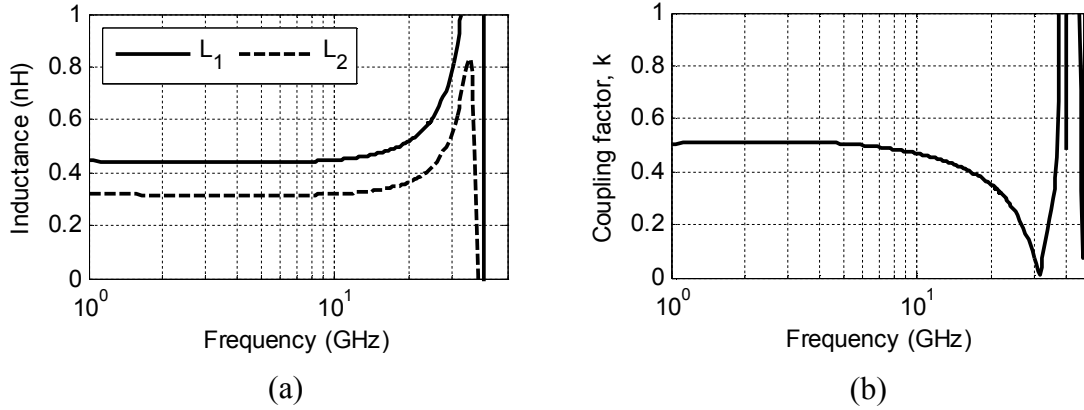


Figure 3.26. Simulation results for the T-coil of Figure 3.24(a). (a) Inductance of primary (L_1) and secondary (L_2) inductors, and (b) coupling factor. These are frequency-dependent parameters.

Table 3.2. The designed T-coil parameters, extracted from the layout in Figure 3.24(a).

L_1	L_2	k	C_B
449 pH	320 pH	0.47	11 fF

The SRF is 45 GHz, which is sufficient for the target 25-Gb/s data-rate, but not over-specified. Usually, the required SRF should be much higher than the operating frequency, because the behaviors of inductors are degraded at much smaller frequency than the SRF [49]. If a higher data-rate were required, the SRF could be increased by changing the geometry of the spiral inductor, which usually tends to lower the inductance [49]. This lower inductance would decrease the ability of parasitic capacitance compensation by the T-coil, which may require size reduction of dual-diodes to maintain I/O bandwidth; it could compromise the ESD protection level.

3.4.3 Secondary ESD protection

Two-stage ESD protection using secondary ESD device is one of the conventional ways to protect input circuitry especially in CDM event, by clamping voltage overshoot [33]. Figure 3.27 shows a T-coil with a secondary ESD protection device, composed of a series resistor, R_{ser} , and small dual-diodes (DP2 and DN2) with a perimeter of P_{dio} . The secondary dual-diodes can be sized less than primary dual-diodes, to suppress voltage overshoot at the receiver less than the gate-oxide breakdown voltage. In this dissertation, the secondary protection devices are sized as follows: $R_{ser} = 50 \Omega$, and $P_{dio} = 40 \mu m$. The perimeter of $40 \mu m$ is the minimum dual-diode size guaranteed from the 65-nm PDK, and R_{ser} is optimized considering peak voltage suppression and bandwidth reduction.

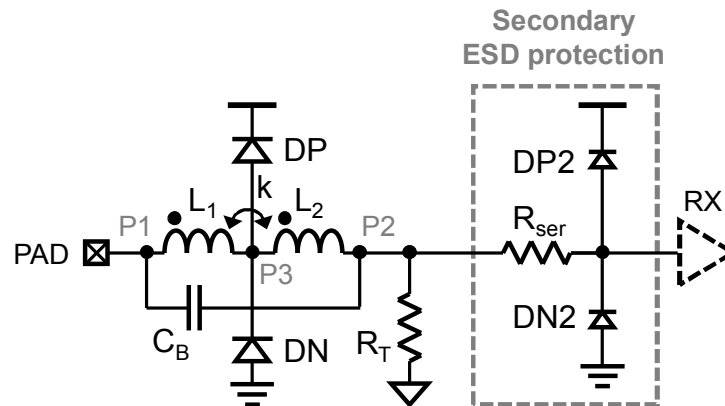


Figure 3.27. T-coil with a secondary ESD protection composed of a series resistance (R_{ser}) and small dual-diodes (D_T and D_B).

3.4.4 Shadow inductor

A third approach for voltage overshoot mitigation in T-coils is to maximize the eddy current effect for self-inductance reduction by utilizing a “shadow inductor.” This idea was motivated by the experiment results in [44] (see Figure 3.28), to study different shielding patterns for Q-factor improvement in the on-chip spiral inductors. In Figure 3.28, “SGS” means a solid ground shield, designed by solid metal planes underneath the spiral inductor. With an SGS, the inductance is reduced significantly, and the series resistance increases, as shown in Figure 3.28(a) and (b), respectively. This is because of the eddy current losses by SGS. Figure 3.29 illustrates the eddy current generation. When an AC current flows in a metal line, a time-varying magnetic field is generated. If a conductive plane is placed under the metal line, a current is induced by the time-varying magnetic field in the conducting plane. This induced current is called as the eddy current, which has an opposite direction to the original source current on the metal line. The eddy current generates a counter-magnetic field with the opposite direction to the original magnetic field, which decreases the magnetic field intensity. The adverse effect of the eddy current is to reduce inductance of the original metal line, because inductance is proportional to the magnetic field intensity; this effect already had been proved by the experiment results in Figure 3.28.

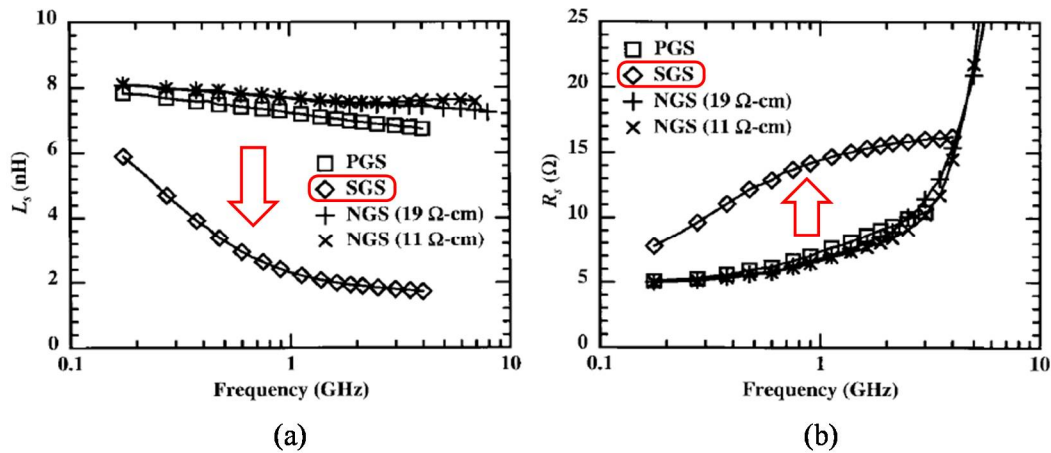


Figure 3.28. Spiral inductor measurement results with various shielding patterns from [44]. (a) Self-inductance, and (b) series resistance of the spiral inductors (PGS: patterned ground shield, SGS: solid ground shield, NGS: no ground shield with different silicon substrate resistivity).

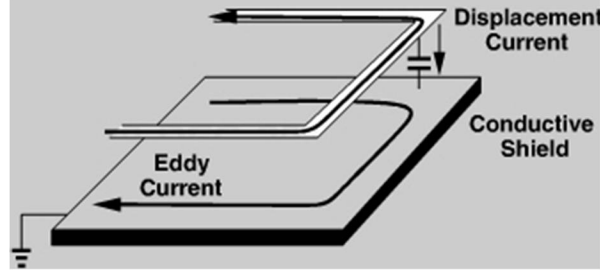


Figure 3.29. Conceptual illustration for eddy current generation in the conductive plane [50].

This inductance reduction mechanism by the eddy current may be applied to mutual inductance reduction during ESD to reduce voltage peaking. Figure 3.30 illustrates the concept of inductance control using the “shadow inductor”. The shadow inductor loop (L_{sh}) underneath the main inductor (L_m) is connected to a switch (S0) that is triggered by an ESD event. During normal operation, the switch (S0) is off, which may not make any harmful effect on the main inductor. However, during ESD, the S0 turns on to make a closed loop in the shadow inductor; the eddy current can flow in the closed loop to generate a counter-magnetic field against the main inductor, which reduces the main inductance, L_m .

To estimate the amount of inductance reduction with the shadow inductor, the eddy current effect is modeled as a transformer relationship between L_m as a primary winding and L_{sh} as a secondary winding in a similar way used in [50]. This is shown in the equivalent circuit in Figure 3.31. The coupling between L_m and L_{sh} is modeled as a magnetic coupling factor, k . Series resistance (R_s) on the shadow inductor is mainly contributed by the on-resistance of the switch, S0. Hence, the input impedance at the main inductor side can be calculated as follows:

$$\frac{V_{in}}{I_{in}} = \frac{k^2 \omega^2 L_m L_{sh} R_s}{R_s^2 + \omega^2 L_s^2} + j\omega L_m \left(1 - \frac{k^2 \omega^2 L_s^2}{R_s^2 + \omega^2 L_s^2} \right). \quad (3.11)$$

The imaginary part of (3.11) is an equivalent inductance modulated by the eddy current effect. Theoretically, if $k = 1$ and $R_s = 0$, the equivalent inductance in the main inductor becomes zero, which is most desirable for overshoot reduction. When S0 is off ($R_s \sim \infty$), the right-hand side of (3.11) reduces to $j\omega L_m$, demonstrating that the original inductance is unaffected. However, these situations are ideal, and cannot be achieved in practice.

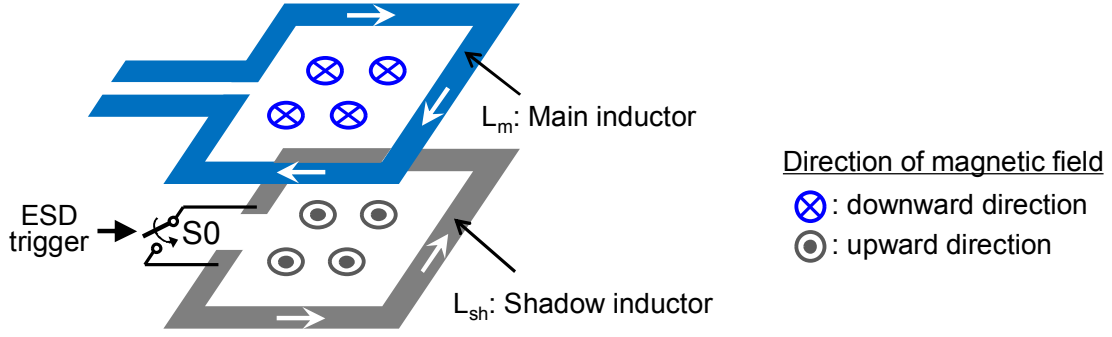


Figure 3.30. Concept of the indirect inductance control by the shadow inductor.

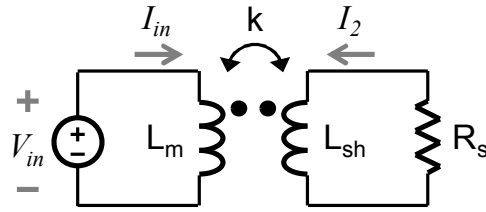


Figure 3.31. Equivalent circuit of Figure 3.30.

To check the feasibility of this technique with practical component values, parameters in (3.11) are varied to observe practically attainable inductance reduction. In (3.11), there are three design parameters (R_s , k , and L_{sh}), and two given parameters (L_m and ω). In Figure 3.32, the effective inductance of the main inductor ($L_{m,eff}$), is calculated by taking the imaginary part of (3.11) with variable design parameters (R_s , k , and L_{sh}) and a frequency (ω). The design default values are $L_m = 500$ pH, $L_{sh} = 500$ pH, $k = 0.9$, and frequency of concern is 3.5 GHz. This 3.5 GHz is a frequency component of the CDM pulse with a 100 ps rise-time, which is calculated by the general relationship between rise time and bandwidth, $bandwidth = 0.35 / rise\ time$ [51]. The target inductance reduction ratio is 70% as an initial goal, to lead to $L_{m,eff}$ of 150 pH from $L_m = 500$ pH. In Figure 3.32(a), R_s is varied with three different k values of 0.8, 0.9 and 1. For $L_{m,eff}$ less than 150 pH, R_s should be smaller than 4 Ω at $k = 0.9$. For $k = 0.8$, 70% reduction cannot be achieved. For $k = 1$, maximum R_s can be relieved to 7 Ω , but $k = 1$ is not obtainable in practical on-chip transformers. Figure 3.32(b) shows $L_{m,eff}$ with different k 's in three R_s cases: 5, 10 and 20 Ω . The similar results to Figure 3.32(a) are given; R_s should be less than 5 Ω in reasonable k values to achieve more than 70% effective inductance reduction. In submicron CMOS technologies, achieving 5- Ω on-

resistance using thin-gate oxide transistors is not impossible, but it needs an increased transistor size which may increase parasitic capacitance. Shadow inductance (L_{sh}) is varied in Figure 3.32(c) to relieve the requirements of k and R_s . By increasing L_{sh} to 1 nH or more (more than twice), the maximum R_s requirement can be relieved to 10 Ω . Figure 3.32(d) presents the trend of effective inductance with the frequency component of the CDM pulse. With a higher frequency, i.e. shorter rise-time of the CDM pulse, the efficiency of the shadow inductor is improved. In other words, if the rise time of the incoming CDM pulse is slow, the shadow inductor will not be as effective as expected.

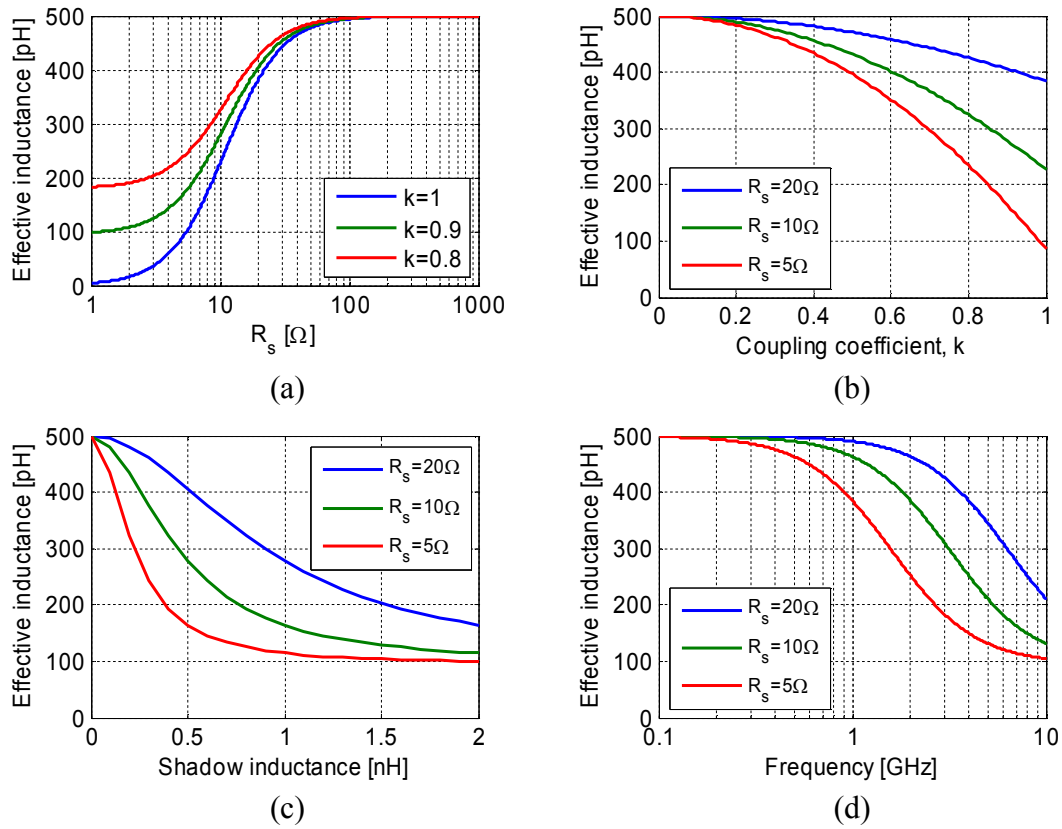


Figure 3.32. Simulation results of effective inductance of the main inductor for $L_m = 500$ pH, with the varied parameters: (a) R_s varied with $k = 0.8, 0.9$ and 1 , (b) k varied with $R_s = 5, 10$ and 20Ω , (c) L_{sh} with $R_s = 5, 10$ and 20Ω , and (d) frequency with $R_s = 5, 10$ and 20Ω . Other parameters, if not specified in each plot, are at the default values: $L_{sh} = 500$ pH, $k = 0.9$ and frequency = 3.5 GHz.

From the observations in Figure 3.32, L_{sh} should be at least two times larger than L_m at a 100 ps rise-time CDM event. The overall design space is explored as shown in Figure 3.33. The two major design parameters, R_s and k , are varied on the x- and y-axis, respectively. The color bar shown in the right

indicates the effective inductance ($L_{m,eff}$). For 70% reduction ($L_{m,eff} = 150$ pH), k should be larger than 0.85 and R_s should be smaller than 10Ω from the contour plot in Figure 3.33. If more than 70% is needed, the design space becomes much smaller. For example, if 90% reduction is needed, the minimum k is 0.95 and R_s should be much less than 5Ω , which are not in practical ranges.

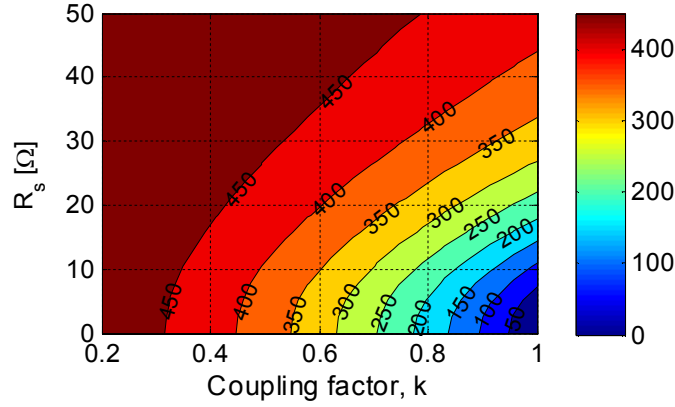


Figure 3.33. Overall design space for the shadow inductor at $L_m = 500$ pH, $L_{sh} = 1$ nH, and frequency of 3.5 GHz.

In order to solve this issue with more practical ranges of the component values, multiple shadow inductors can be used. The equation (3.11) can be expanded to the multiple shadow inductor cases, and the effective main inductance ($L_{m,eff}$) can be calculated as follows:

$$L_{m,eff} = L_m \left(1 - \sum_i \frac{k_i^2 \omega^2 L_{s,i}^2}{R_{s,i}^2 + \omega^2 L_{s,i}^2} \right), \quad (3.12)$$

where i is an index for the i -th shadow inductor, and all the parameters with the subscript, i , denote the parameters for the i -th shadow inductor. Thus, each shadow inductor does not need to have tight design constraints as in Figure 3.33, and has more freedom to choose geometry of the shadow inductors and switch sizes. After several tedious iterations using EM simulations, the T-coil with the most optimal shadow inductors is shown in Figure 3.34. Three shadow inductors are used in addition to the main T-coil in Figure 3.24(a). The shadow inductor 1 and 2 are the outer and inner rims, respectively; both are coplanar to the main T-coil (laid out with metal7 and metal6, and also metal5 to further decrease series resistance). The shadow inductor 3 is located underneath the main T-coil in metal5, which covers all the

bottom areas that the main T-coil occupies. The simulation results are given in Table 3.3 for the inductance and coupling factor changes in the main T-coil, by turning on and off the shadow inductors. In this simulation, no actual switches are added, and shadow inductors are turned on by simply connecting the inductor loops with the metal lines. The mutual inductance of the original T-coil is 178 pH, but is reduced to as low as 16 pH by turning on the shadow inductors, which creates 91% reduction. This remarkable result will be beneficial for voltage overshoot reduction. However, the intrinsic bandwidth of this T-coil is significantly reduced; it is only 23 GHz, while the inductance-halving T-coil can achieve as high as 32 GHz and the secondary protection has 26 GHz bandwidth, all from pre-layout simulation. This reduced bandwidth is caused by the increased vertical parasitic capacitance through the shadow inductor 3. If the shadow inductor 3 is removed, the bandwidth can be improved, but the mutual inductance reduction becomes insufficient: only 40-50% reduction. Therefore, the shadow inductor technique is theoretically plausible for mutual inductance control, but its efficiency is practically limited by bandwidth performance. Thus, this technique was not included in the test chip.

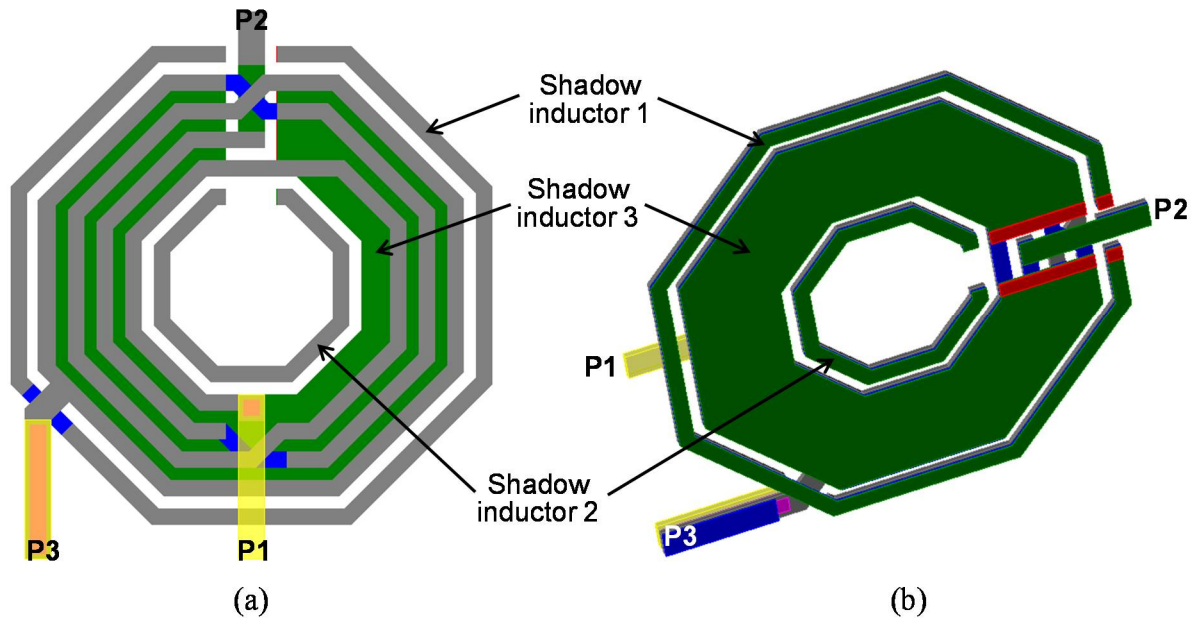


Figure 3.34. Layout of the proposed T-coil with three shadow inductors: (a) top view, and (b) 3-D rendered view at the bottom (yellow: pad metal, gray: metal7, blue: metal6, green: metal5, red: metal4).

Table 3.3. Simulation results of the T-coil with the shadow inductors.

	L_1	L_2	k	M_{12}
Shadow inductors: OFF	438 pH	314 pH	0.48	178 pH
Shadow inductors: ON	204 pH	187 pH	0.08	16 pH

3.5 Receiver Circuit Implementation

To compare the performance of different T-coil schemes in the context of actual I/O circuits, 25-Gb/s wireline receivers will be implemented. One of the target applications of this high data-rate is 100-Gb/s backplane Ethernet (e.g. IEEE P802.3bj) based on four lanes at 25 Gb/s per lane [52], [53]. The circuits are implemented in a 65-nm CMOS technology with seven metal layers. Supply voltage for thin-gate oxide transistors is 1.2 V.

Figure 3.35 illustrates a full chip block diagram of the test chip. Three receivers, denoted by HSRX1-3 are implemented; HSRX1 and HSRX2 have T-coils with inductance halving and secondary ESD protection, respectively. HSRX3 has the same T-coil as in HSRX2 with secondary ESD protection, but it has a different receiver front-end circuit, which will be discussed in Section 3.5.8. The basic receiver architecture for the three splits is as follows: (1) T-coil for compensating parasitic from primary ESD diode, (2) the input termination network (R_{term}), (3) the continuous-time linear equalizer (CTLE) with resistive and capacitive degeneration and shunt peaking [26], (4) the CTLE buffer to reduce loading capacitance seen at the output of the CTLE, (5) the internal buffer to interface output of each individual receiver to the output driver, and (6) the 50- Ω output driver. Because of the limited number of bond-pads and chip size, only a single output driver is employed, and each receiver is connected through long on-chip interconnection lines to the output driver. The test chip is configured by control registers via three-wire interface signals (LE, SDI and SCK). Detailed circuit design will be given in the following sections.

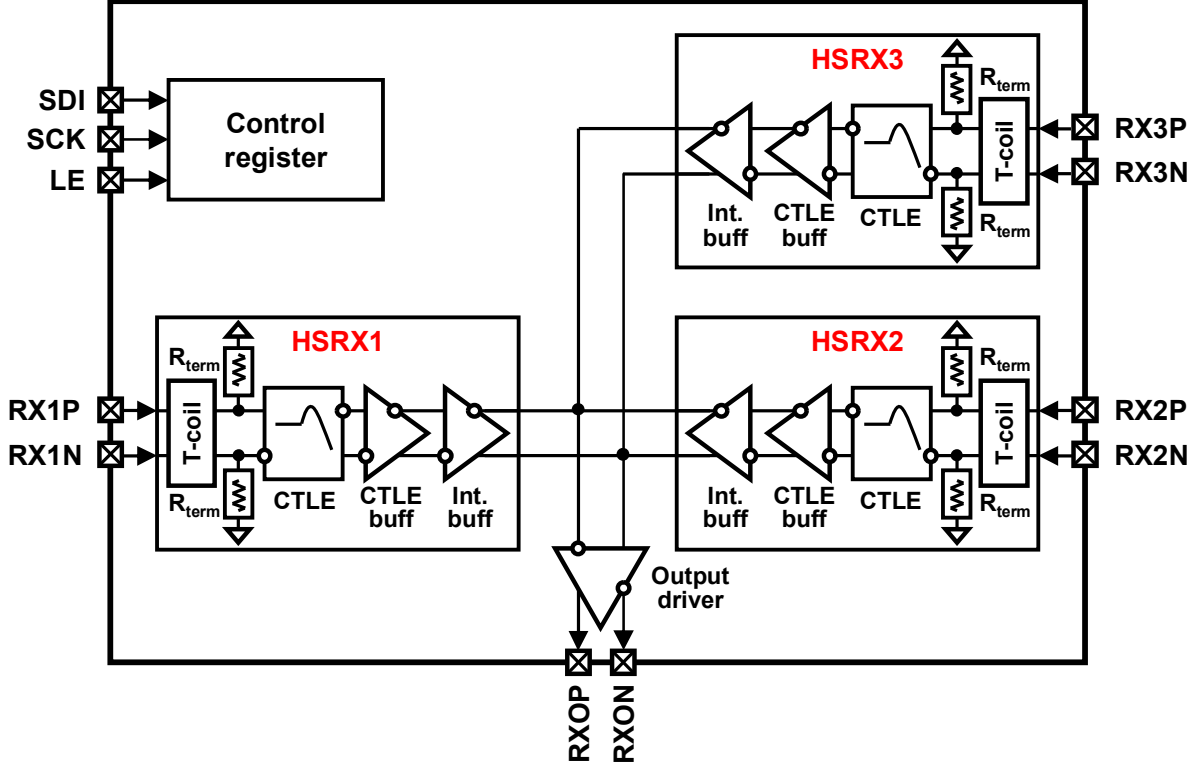


Figure 3.35. Full-chip block diagram of the test chip.

3.5.1 Continuous-time linear equalizer

The role of an equalizer in communication links is to flatten low-pass channel response using an “inverse filter” implemented by a high-pass filter (or a band-pass filter in actual implementation), up to the frequency of concern. Generally, composite frequency response of channel and equalizer should be flat up to one-half baud-rate to preserve input signal information in an acceptable amount. The equalization function can be implemented at the transmitter by de-emphasis, or at the receiver by continuous-time equalizer or by decision-feedback equalizer (DFE) [54]. Figure 3.36 shows the behavior of an ideal equalizer as a high-pass filter to compensate for a lossy channel with a 3-dB bandwidth, ω_{ch} . Equalizers have four important parameters: “DC gain” at low frequency, “max. gain” at ω_0 (“peak frequency”), and “peaking” which is the difference between max. gain and DC gain; the amount of peaking should be sufficient to compensate for channel loss. As shown in Figure 3.36, the response of channel plus equalizer is “equalized” up to ω_0 , by the equalizer.

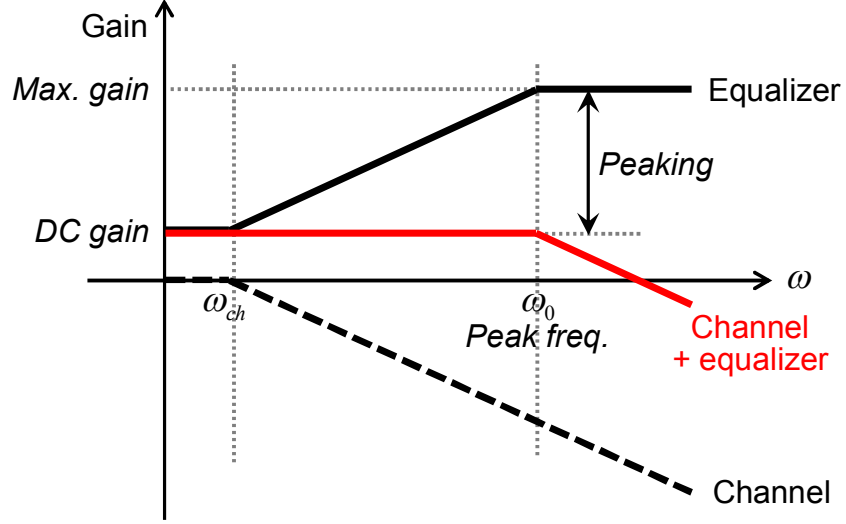


Figure 3.36. Concept of ideal equalizer in high-pass filter with a pole frequency at ω_0 .

In this study, the CTLE with an active amplifier will be implemented, as illustrated in Figure 3.37. The basic concept of the CTLE is to use resistive and capacitive degeneration at the source of the differential amplifier to attenuate DC gain and relatively boost high-frequency gain [54]. The transfer function of this CTLE is expressed as follows:

$$H(s) = \frac{V_o}{V_i} = -\frac{g_m R_2}{1 + g_m R_1} \frac{1 + sC_1 R_1}{\left(1 + s \frac{C_1 R_1}{1 + g_m R_1}\right)(1 + sC_L R_2)}, \quad (3.13)$$

where g_m , R_1 , C_1 , R_2 and C_L denote transconductance of the input NMOS transistor, source-degeneration resistor, source-degeneration capacitor, load resistor, and load capacitor, respectively. Tail current (I_0) flows in each rail to determine transconductance, g_m . The transfer function in (3.13) has one zero and two poles as below:

$$\omega_{z1} = \frac{1}{C_1 R_1} \quad (3.14)$$

$$\omega_{p1} = \frac{1 + g_m R_1}{C_1 R_1}. \quad (3.15)$$

$$\omega_{p2} = \frac{1}{C_L R_2} \quad (3.16)$$

and DC gain is expressed as follows:

$$A_0 = \frac{g_m R_2}{1 + g_m R_1}. \quad (3.17)$$

The Bode plot of (3.13) is plotted in Figure 3.38. The four design parameters, R_1 , R_2 , C_1 and g_m should be optimized to achieve the maximum peaking at the desired peak frequency. The zero frequency (ω_{z1}) first should be matched with the channel bandwidth (ω_{ch} in Figure 3.36). Typically, ω_{p1} is designed to be smaller than ω_{p2} to control the peak frequency (ω_0 in Figure 3.36), because ω_{p2} is not fully controllable due to the load capacitor term, C_L . The max. gain is calculated as $A_0 \omega_{p1} / \omega_{z1}$ by plugging $s = j\omega_{p1}$ into (3.13). This shows that the peaking is expressed as:

$$\text{Peaking} = \frac{\omega_{p1}}{\omega_{z1}} = 1 + g_m R_1. \quad (3.18)$$

This peaking is also the same as the distance between ω_{p1} and ω_{z1} , which is the distance between one-half baud-rate and channel bandwidth. The term in (3.18) is also related to the DC gain in (3.17). Hence, to increase peaking ($1 + g_m R_1$), DC gain should be sacrificed; the overall output amplitude of the CTLE is attenuated, which gives more burden in the following stages to have larger gain to compensate for DC gain attenuation. In order not to affect DC gain while increasing peaking, only g_m should be increased, because $A_0 = R_2 / R_1$ for large g_m . However, it needs significant power consumption because $g_m \propto \sqrt{I_0}$. This issue becomes more critical in higher data-rate links on heavily lossy channels, where the separation between ω_{p1} and ω_{z1} is larger.

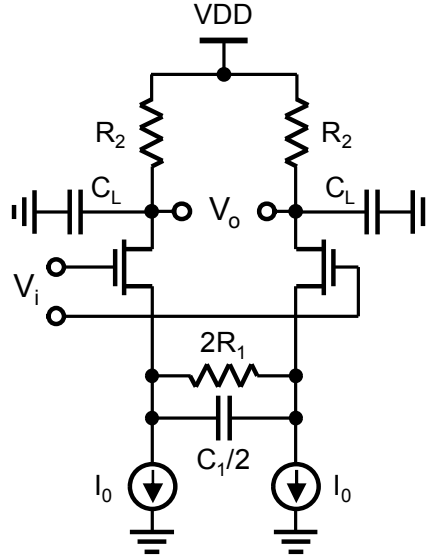


Figure 3.37. CTLE with source-degeneration resistor and capacitor.

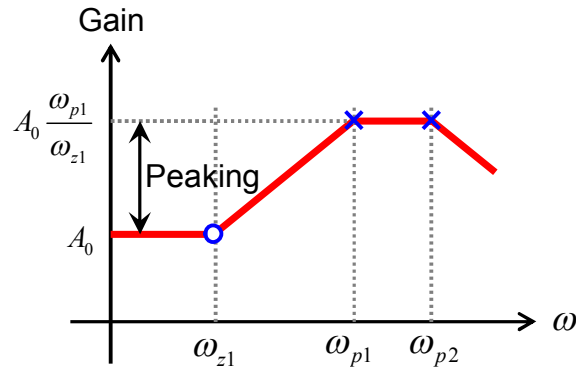


Figure 3.38. Bode plot for Figure 3.37.

To solve this problem, shunt-peaking is used in addition to source degeneration [28], [55] (see Figure 3.39). The basic idea for this scheme is to decouple peaking from DC gain, and widen the distance between the peak frequency and the zero frequency using the pole-zero cancellation (or, pole-zero doublets) [56]. The CTLE in Figure 3.39 has additional load inductor, L_3 . The transfer function of Figure 3.39 can be expressed as follows:

$$H(s) = \frac{V_o}{V_i} = -\frac{g_m R_2}{1 + g_m R_1} \frac{(1 + s C_1 R_1) \left(1 + s \frac{L_3}{R_2}\right)}{\left(1 + s \frac{C_1 R_1}{1 + g_m R_1}\right) (1 + s C_L R_2 + s^2 C_L L_3)}. \quad (3.19)$$

The DC gain (A_0) is exactly the same as that of Figure 3.37, which is given in (3.17). This transfer function, (3.19), has two zeros and three poles, and the denominator has the second-order polynomial.

The pole and zero frequencies, and additional parameters relative to peaking are determined as below:

$$\omega_{z1} = \frac{1}{C_1 R_1} \quad (3.20)$$

$$\omega_{p1} = \frac{1 + g_m R_1}{C_1 R_1} \quad (3.21)$$

$$\omega_{z2} = \frac{R_2}{L_3} \quad (3.22)$$

$$\omega_n = \frac{1}{\sqrt{C_L L_3}} \quad (3.23)$$

$$\zeta = \frac{R_2}{2} \sqrt{\frac{C_L}{L_3}}. \quad (3.24)$$

The first zero (ω_{z1}) and the first pole (ω_{p1}) are identical to those in (3.13). This transfer function has one more zero (ω_{z2}) and a natural frequency (ω_n), instead of ω_{p2} in (3.14). From the damping factor (ζ) in (3.24), the amount of peaking and the peak frequency can be determined. According to the frequency response of the second-order systems, the under-damping condition for ζ is

$$0 \leq \zeta \leq 1/\sqrt{2} \quad (3.25)$$

to have peaking around ω_n , and the actual peak frequency (ω_{pk}) is derived as follows [57]:

$$\omega_{pk} = \omega_n \sqrt{1 - 2\zeta^2} = \sqrt{\frac{1}{C_L L_3} - \frac{1}{2} \frac{R_2^2}{L_3^2}}. \quad (3.26)$$

At the peak frequency, the max. gain can be calculated as follows:

$$|H(j\omega_{pk})| = A_0 \frac{\omega_{p1}\omega_{pk}}{\omega_{z1}\omega_{z2}} \frac{1}{2\zeta\sqrt{1-\zeta^2}} = \frac{g_m L_3}{R_2 C_L} \sqrt{\frac{1-2\zeta^2}{1-\zeta^2}}. \quad (3.27)$$

This CTLE can extend the peak frequency and the amount of peaking by pole-zero cancelation between ω_{p1} and the additional zero (ω_{z2}) [55]. Figure 3.40 shows the Bode plot of Figure 3.39; the peak frequency is extended up to ω_{pk} , by $\omega_{p1} = \omega_{z2}$. For the target specifications, five design parameters, R_1 , R_2 , C_1 , L_3 and g_m can be determined in a given C_L , as follows:

$$A_0 = \frac{g_m R_2}{1 + g_m R_1} \geq \text{min. DC gain} \quad (3.28)$$

$$\zeta = \frac{R_2}{2} \sqrt{\frac{C_L}{L_3}} \leq 1/\sqrt{2}, \text{ for peaking condition} \quad (3.29)$$

$$\omega_{z1} = \frac{1}{C_1 R_1} = \omega_{ch}, \text{ for channel bandwidth} \quad (3.30)$$

$$\omega_{p1} = \frac{1 + g_m R_1}{C_1 R_1} = \frac{R_2}{L_3} = \omega_{z1}, \text{ for the pole-zero doublets} \quad (3.31)$$

$$\omega_{pk} = \sqrt{\frac{1}{C_L L_3} - \frac{1}{2} \frac{R_2^2}{L_3^2}} = 2\pi \times \frac{f_{baud}}{2}, \text{ for peaking frequency.} \quad (3.32)$$

However, the above equations are derived from ideal circuit behavior; in actual circuits for very high data-rate applications, finite output resistance of input transistors and non-dominant poles can change DC and max. gains, and peak frequency. Moreover, some parameters such as L_3 and its series resistance (R_s) are frequency-dependent. Thus, starting from the hand-calculated parameters, several iterations using circuit simulators are needed to optimize the design parameters.

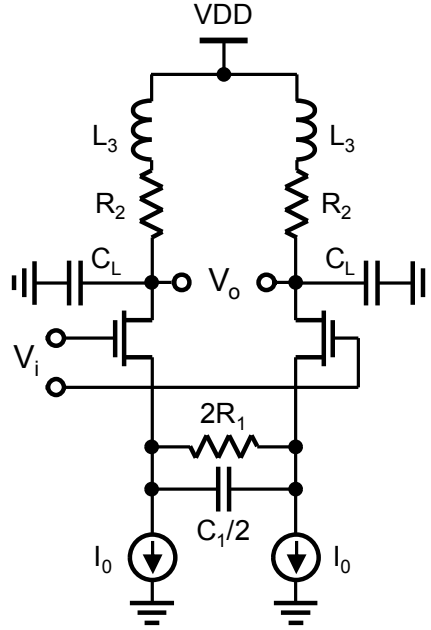


Figure 3.39. CTLE with source-degeneration, and shunt-peaking at the load.

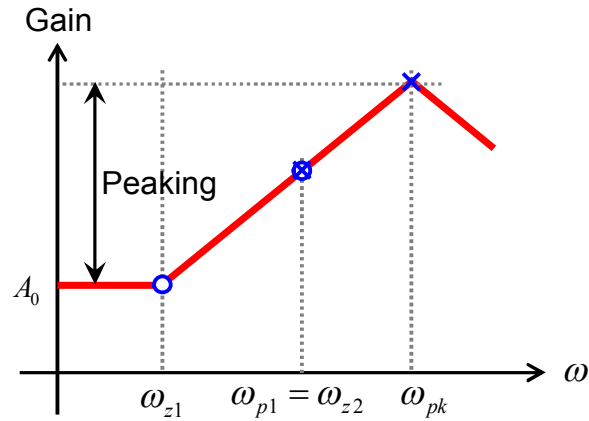


Figure 3.40. Bode plot for Figure 3.39.

The actual CTLE based on Figure 3.39 is implemented as shown in Figure 3.41. Source degeneration resistors (R_1) and capacitors (C_1) are implemented by resistor banks controlled by 2-bit register, “RES1[1:0],” and capacitor banks controlled by 2-bit register, “CAP1[1:0],” respectively. The tuning ranges of R_1 and C_1 are as follows:

$$R_1 = (210 + 140 \times \text{RES1}[1:0]) / 2 \text{ } (\Omega) \quad (3.33)$$

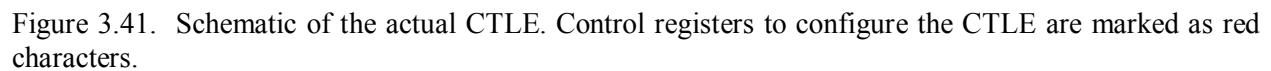
$$C_1 = (60 + 60 \times \text{CAP1}[1:0]) \times 2 \text{ (fF)}. \quad (3.34)$$

The default register values are $\text{RES1}[1:0] = 01$ for $R_1 = 175 \Omega$, and $\text{CAP1}[1:0] = 01$ for $C_1 = 240 \text{ fF}$. The load resistor (R_2) is also tunable by 1-bit signal “R2_SEL”; $R_{21} = 300 \Omega$ and $R_{22} = 750 \Omega$, and the default value of R2_SEL is 0 for $R_2 = 250 \Omega$ including on-resistance of MP1 or MP2 $\approx 800 \Omega$, but excluding series resistance from L_3 . The load capacitor (C_L) is assumed to be 30 fF, including input capacitance of the following amplifier stage and layout parasitics (not shown in Figure 3.41). An equalization turn-off option is also implemented by a control register “EQ_OFF”; this option is to test the intrinsic characteristic of the receiver amplifier without peaking. Because three different receivers share a single output driver, each receiver stage should be shut down by disabling tail currents. This is implemented by another control register (not shown in Figure 3.41 for simplicity).

The tail current (I_0) determines output DC common-mode voltage as $V_{DD} - I_0 R_2$, which should be in the range to drive the next-stage amplifiers. In such high-speed applications above 10 Gb/s, only NMOS input transistors are used; at the 1.2-V supply voltage, the output common-mode voltage is usually set to be larger than one-half V_{DD} to turn on NMOS transistors with sufficient margin. The input transistor, MN1 (or MN2) is sized as $40\mu\text{m}/80\text{nm}$ considering the required g_m ; non-minimum length is used to avoid DC gain reduction by small output resistance of the input transistor.

The required inductance for shunt-peaking (L_3) is around 3 nH, which generally needs a large silicon area. Typically, to save area, two load inductors could be implemented by a single center-tapped symmetric inductor. However, symmetric inductors may make signal routing difficult in the cascading stages, and the distance between adjacent blocks may be increased, which adds more parasitic capacitance [58]. In this case, L_3 will be better implemented by two individual “series inductors” as shown in Figure 3.42, which makes a series connection between the metal7 inductor and the metal6 inductor to effectively increase total length within a limited area. Figure 3.42(a) shows the top view of the designed L_3 ; total inductance of 3.1 nH is obtained in such a small outer dimension (OD) as $50 \mu\text{m}$, by the effective number

As the base amplifier of Figure 3.41 is a differential amplifier, the circuit layout should be carefully drawn symmetrically, including all transistor placements and signal routings. Unexpected parasitics are minimized by several iterations through post-layout simulation.



R_1	R_2	C_1	L_3	$R_{s,L3}$	C_L	I_0	$(W/L)_{MN1,2}$
175 Ω	250 Ω	240 fF	3.1 nH	63 Ω	30 fF	1 mA	40 μ m/80nm

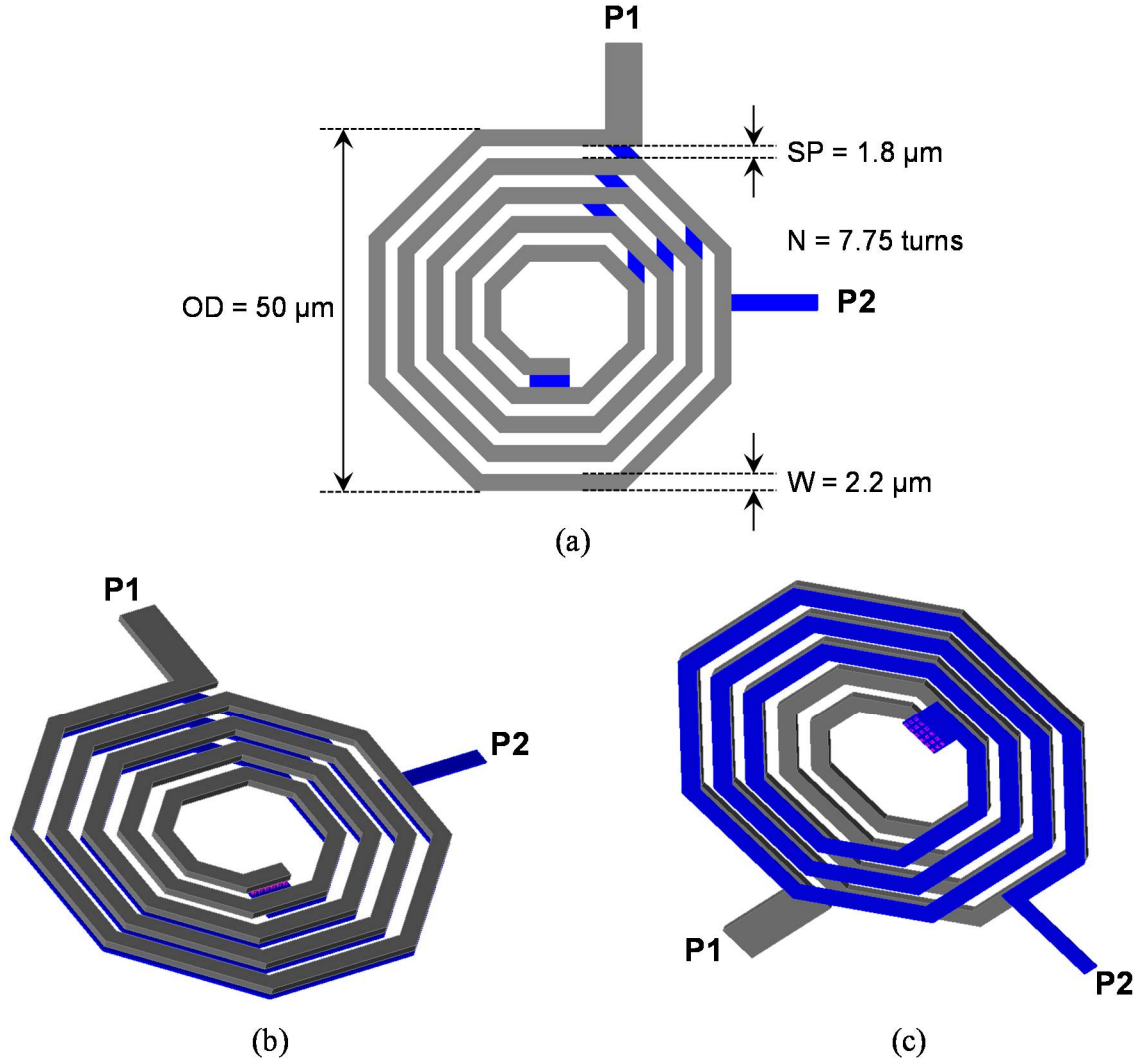


Figure 3.42. Layout of shunt-peaking inductor, L_3 . (a) top view with geometry parameters (OD for the outer dimension, W for the width, SP for the turn-to-turn spacing, and N for the total number of turns), (b) 3-D rendered view above the spiral inductor, and (c) 3-D rendered view below the spiral inductor (gray: 4 \times top metal, blue: 1 \times second top metal).

The AC simulation result is presented in Figure 3.43 in post-layout simulation. It shows total peaking gain of 9.51 dB with DC gain of -1.89 dB and max. gain of 7.62 dB at 9.52 GHz at the default register settings. Figure 3.44 shows tunability of the CTLE response with various R_1 and C_1 values; R_1 is varied in Figure 3.44(a) with a fixed C_1 of 240 fF, and C_1 is varied in Figure 3.44(b) with $R_1 = 175 \Omega$. It covers 5.2-dB DC gain tuning range by R_1 control and 1.2-dB max. gain tuning by C_1 control.

The performance of the CTLE is summarized in Table 3.5. Total power consumption of the CTLE is 2.3 mW at the 1.2-V power supply.

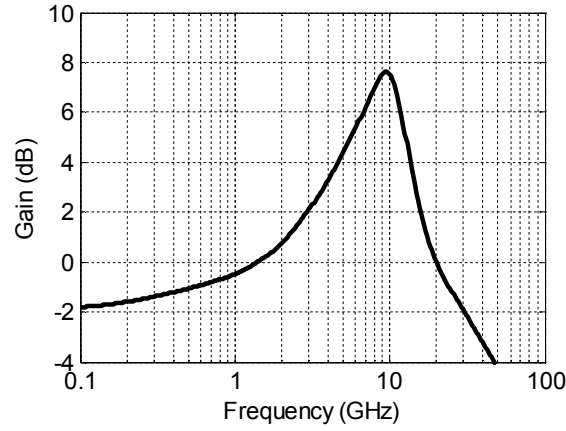


Figure 3.43. Simulation results of AC magnitude response of the CTLE at default register values (RES1[1:0] = 01, CAP1[1:0] = 01, and R2_SEL = 0).

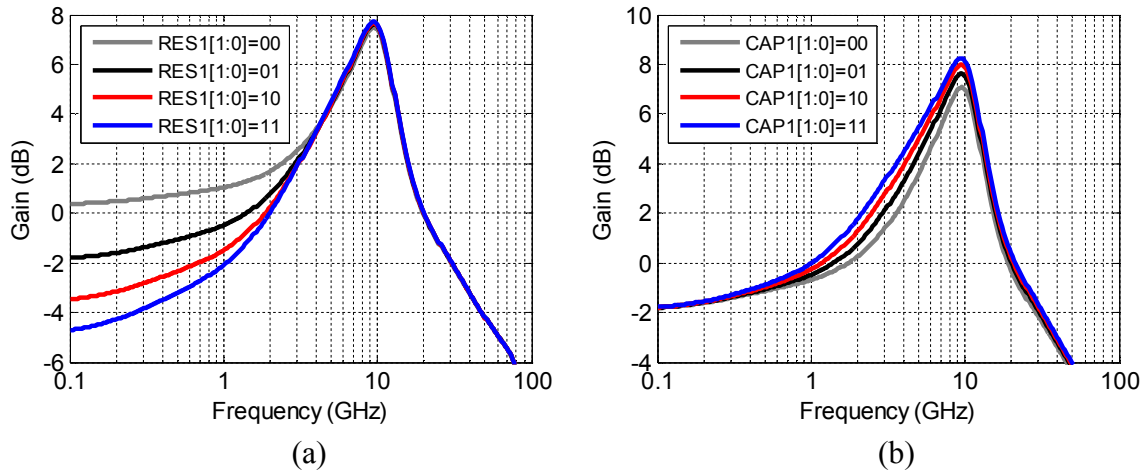


Figure 3.44. AC simulation results to show tunability of the CTLE: (a) variable RES1[1:0] at default C_1 of 240 fF, and (b) variable CAP1[1:0] at default R_1 of 175 Ω .

Table 3.5. Performance summary of the post-layout simulation results for the CTLE.

Items	Results		
	Min.	Max.	Default
DC gain	-4.91 dB	0.33 dB	-1.89 dB
Max. gain	7.07 dB	8.24 dB	7.62 dB
Peaking	6.60 dB	13.21 dB	9.51 dB
Peak frequency	9.52 GHz		
Power consumption	2.3 mW		

3.5.2 CTLE output monitoring

Monitoring the output of the CTLE is not trivial. As the CTLE output is commonly used by the other internal circuitry, the CTLE generally does not have sufficient driving capability to drive heavy external loads such as bondwire and board trace, and its output impedance is not properly matched to 50 Ω . Hence, in this design, several amplifier stages will be cascaded at the output of the CTLE for monitoring with external oscilloscopes. This is a similar configuration to the limiting amplifiers in optical receiver front-end [27], [59], but the DC gain requirement is much relieved—ideally, 0 dB, because the CTLE output amplitude is already sufficiently high (~a few hundreds of mV) and the actual output waveforms should be preserved.

Three-amplifier cascading stages will follow as shown in Figure 3.35, which is composed of the CTLE buffer, the internal buffer and the output driver. Theoretically, the overall bandwidth (ω_{tot}) of N identical stages with a bandwidth of ω_0 is expressed [59] by

$$\omega_{tot} = \omega_0 \sqrt{\sqrt[N]{2} - 1}. \quad (3.35)$$

If $N = 3$ and $\omega_{tot} = 12.5$ GHz, the required bandwidth of each stage (ω_0) is 24.5 GHz, which is as high as the data-rate frequency! Considering the purpose of this study to compare the relative performance of the receivers with different T-coil schemes, the requirement in (3.35) is not a necessary condition, but one of the sufficient conditions. Each block design will be presented in the next sections.

3.5.3 CTLE buffer

Figure 3.45 shows the CTLE buffer, the next stage of the CTLE. To extend the bandwidth of the CTLE buffer, a “capacitive degeneration” scheme is used [59] by degeneration resistor (R_S) and capacitor (C_S) as shown in Figure 3.45, which is conceptually the same as the conventional CTLE in Figure 3.37. The transfer function of Figure 3.45 is calculated as

$$H(s) = -\frac{g_{m3}(R_{L1} \parallel R_{L2}/2)}{1 + g_{m3}R_S/2} \frac{1 + sC_S R_S/2}{\left(1 + s\frac{C_S R_S/2}{1 + g_{m3}R_S/2}\right)(1 + sC_{L1}[R_{L1} \parallel (R_{L2}/2)])}, \quad (3.36)$$

where g_{m3} , R_S , C_S , R_{L1} , R_{L2} and C_{L1} denote transconductance of the input NMOS transistor, source-degeneration resistor, source-degeneration capacitor, load resistor, output shunt resistor, and load capacitor, respectively. This transfer function has one zero and two poles as follows:

$$\omega_{z11} = \frac{1}{C_S R_S} \quad (3.37)$$

$$\omega_{p11} = \frac{1 + g_{m3}R_S}{C_S R_S} \quad (3.38)$$

$$\omega_{p22} = \frac{1}{C_{L1}(R_{L1} \parallel (R_{L2}/2))}. \quad (3.39)$$

By locating ω_{z11} at the 3-dB frequency of the intrinsic amplifier, the final bandwidth can be extended to ω_{p11} by a factor of $1 + g_{m3}R_S$ [59]. Another design point is to keep ω_{p22} far from ω_{p11} so that the 3-dB bandwidth at ω_{p11} is not further decreased. In a given load capacitance C_{L1} of 50 fF from the next stage of the CTLE buffer (the internal buffer), R_{L1} should be selected sufficiently small. However, it can truncate the output-high swing because of the elevated output common-mode voltage, expressed by $V_{DD} - I_1 R_{L1}$. To achieve a smaller output resistance not affecting output common-mode voltage, shunt resistor, R_{L2} , is located between two output terminals as shown in Figure 3.45. As the outputs are differential signals, the center of R_{L2} is virtually grounded; in the small-signal equivalent circuit, this gives a shunt resistance of

$R_{L2}/2$ in parallel with R_{L1} , but in the large-signal model, it does not change any quiescent points, including output common-mode voltage because no current flows across R_{L2} . Thus, output load resistance can be effectively decreased to $R_{L1} // (R_{L2}/2)$ and output common-mode voltage is the same as $VDD - I_1 R_{L1}$, by adding an output shunt resistor, R_{L2} .

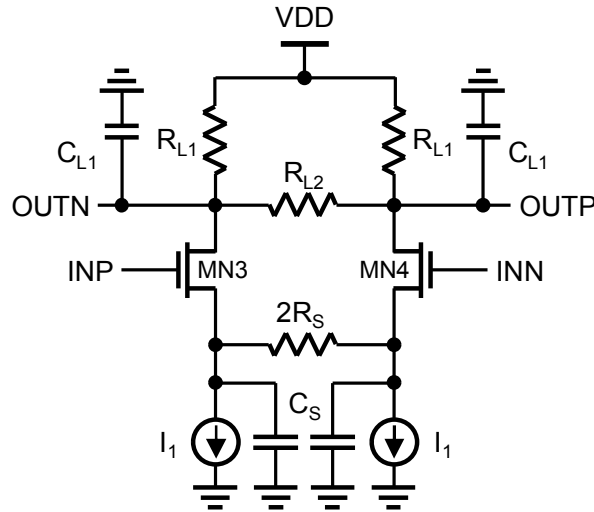


Figure 3.45. Schematic of the CTLE buffer.

The designed parameters are summarized in Table 3.6. From these parameters, AC simulation is performed with post-layout-extracted circuit netlists. Figure 3.46 shows AC magnitude response of the CTLE buffer. DC gain is 0.19 dB and 3-dB bandwidth is 20 GHz. To monitor attenuation behavior, the frequency point to have 1-dB gain attenuation from DC gain is calculated as 11 GHz. The performance of the designed CTLE buffer is summarized in Table 3.7.

Table 3.6. Designed parameters for Figure 3.45.

R_S	C_S	R_{L1}	R_{L2}	I_1	$(W/L)_{MN3,4}$
57.7 Ω	430 fF	375 Ω	1.5 k Ω	1 mA	12 μ m/60nm

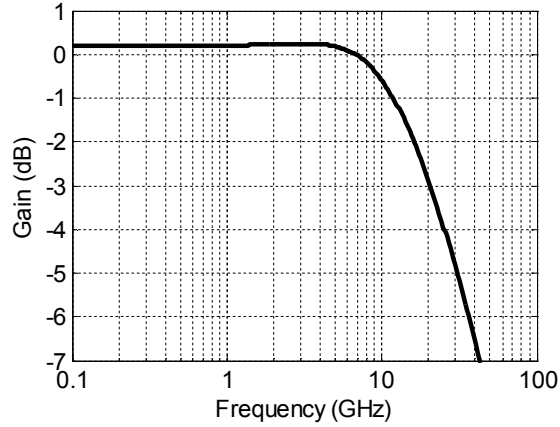


Figure 3.46. AC magnitude response of the CTLE buffer.

Table 3.7. Performance summary of the post-layout simulation results for the CTLE buffer.

Items	Results
DC gain	0.19 dB
1dB bandwidth	11.0 GHz
3dB bandwidth	20.0 GHz
Power consumption	2.3 mW

3.5.4 Internal buffer

As the test chip has three receiver splits in a limited die size ($1\text{ mm} \times 1\text{ mm}$) and limited pad count (at most, 40 pads in a given die area), the outputs of three receiver circuits are multiplexed into a single output driver. Output multiplexing is implemented by a “wired-OR” connection, which merges output metal lines of the amplifiers and turns off the deactivated amplifiers by eliminating tail current paths. The actual layout of on-chip routing between the internal buffers in each split and the output driver is shown in Figure 3.47. HSRX1 and HSRX2 have the same routing length of $100\text{ }\mu\text{m}$, but HSRX3 has a longer metal line of $280\text{ }\mu\text{m}$. The interconnection lines are designed by the differential microstrip lines, with metal6 as a differential signal line and metal1 as a bottom shield. The internal buffers are placed at the last stage of each receiver chain to drive these long interconnection lines.

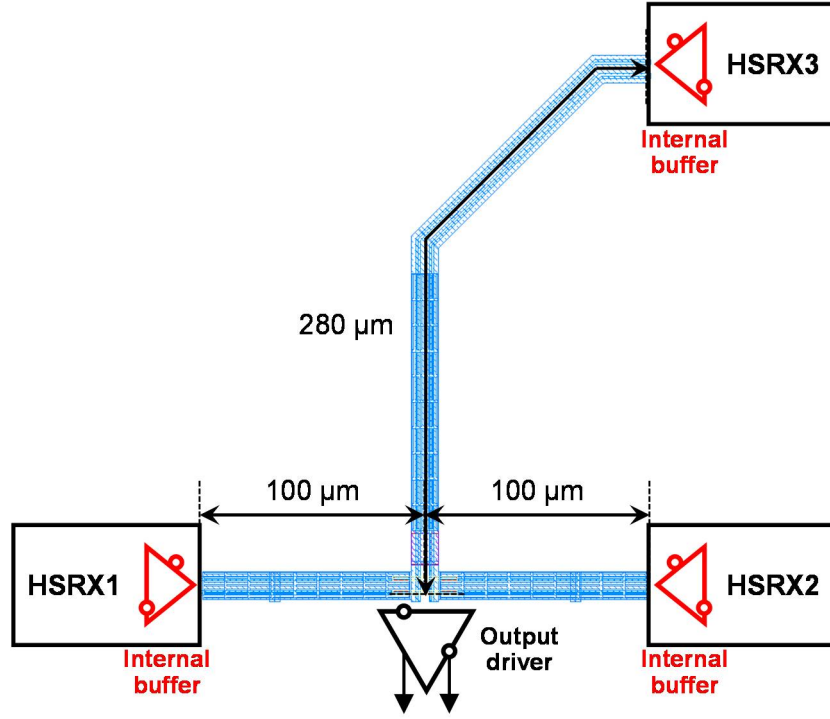


Figure 3.47. Actual layout of internal routing from each receiver circuit to the output driver: 100 μm for HSRX1 and HSRX2 and 280 μm for HSRX3. The metal routing is implemented by differential microstrip with metal6 for differential signals and metal1 for a bottom shield.

Figure 3.48 shows the internal buffer schematic based on a general differential amplifier; it has an output shunt resistor, R_{L4} , like Figure 3.45, to increase the dominant pole frequency by decreasing load resistance but not affecting output common-mode voltage. Two NMOS transistors, MN7 and MN8, are added to completely shut down unnecessary current flow through two NMOS transistors by pulling down the gates of NMOS input transistors, MN5 and MN6. To equalize the parasitic effect of the different routing lengths and to overcome bandwidth degradation from such long on-chip interconnects including parasitic capacitance from the other buffers, the internal buffer is configured to have “distributed” load resistors, as shown in Figure 3.49. This is a similar scheme to a “pseudo-open-drain” (POD) driver [60]. The original POD scheme is to achieve higher data transfer speed in a given power consumption [61], compared to conventional SSTL I/Os [12], and becomes a standard I/O scheme for DDR4 DRAM [62]. In this design, pull-up loads are distributed into all three internal buffers, to assist pull-up operation in each local site, thus to increase transition speed and bandwidth. As shown in Figure 3.49, three internal buffers

are connected in wired-OR connection; two disabled buffers only provide distributed load resistors (gray-colored resistors). With the similar concept, the load resistors can be merged at the input of the output driver, which makes a conventional “open-drain” buffer. From pre-layout simulation, the open-drain configuration can perform similarly to the proposed distributed-load configuration. However, the bandwidth difference between the near-end and far-end buffers is slightly smaller for the distributed-load configuration: 5.7% for the distributed-load vs. 7.5% for the open-drain configuration.

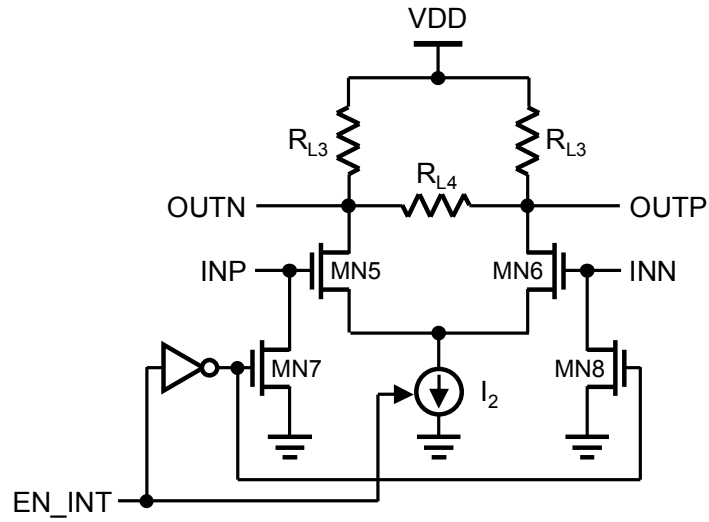


Figure 3.48. Schematic of the internal buffer.

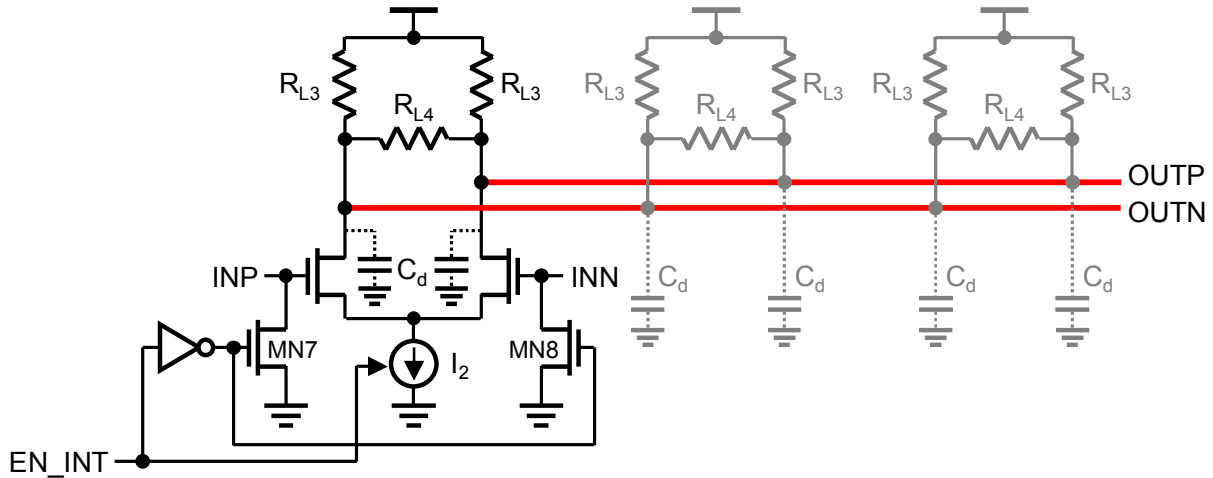


Figure 3.49. Distributed load configuration by an enabled buffer and the other two disabled buffer stages.

The designed parameters are given in Table 3.8. From these resistor values, the equivalent output resistance is $55.5\ \Omega$. The AC simulation results are plotted in Figure 3.50 for the near-end internal buffers (at HSRX1 and HSRX2) and the far-end internal buffer (at HSRX3), all from post-layout simulation. The near-end internal buffer has larger DC gain, but slightly smaller 3-dB bandwidth. The results are summarized in Table 3.9. The 3-dB bandwidth is approximately 17 GHz for both near-end and far-end cases, which demonstrates the distributed load configuration effectively mitigates the performance difference between the near-end and far-end buffers. Power consumption is 7.7 mW at the 1.2-V supply.

Table 3.8. Designed parameters for Figure 3.48.

R_{L3}	R_{L4}	I_2	$(W/L)_{MN5,6}$
375 Ω	600 Ω	5.2 mA	30 μ m/60nm

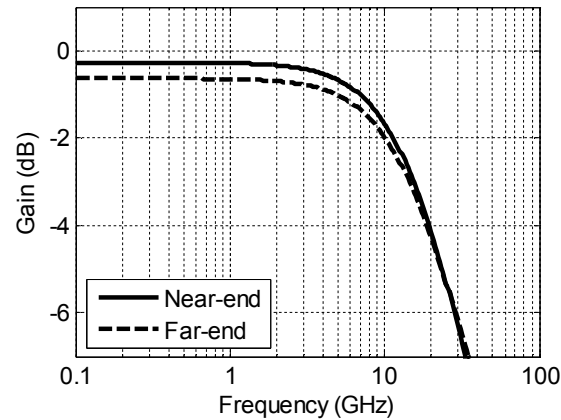


Figure 3.50. AC magnitude response of the internal buffer.

Table 3.9. Performance summary of the post-layout simulation results for the internal buffer.

Items	Results	
	Near-end	Far-end
DC gain	-0.27 dB	-0.63 dB
1dB bandwidth	8.2 GHz	8.4 GHz
3dB bandwidth	16.5 GHz	17.2 GHz
Power consumption	7.7 mW	

3.5.5 Output driver

The output driver should drive $50\ \Omega$ for impedance matching. In addition, there are heavy loads such as capacitance and inductance from bond-pads, bondwires, package leadframe and board traces, which should be driven by the output driver. The output driver schematic is given in Figure 3.51. It utilizes shunt-peaking by inductor L_4 to compensate for bandwidth reduction from output parasitics. Shunt-peaking inductor L_4 is implemented by center-tapped symmetrical inductor with OD = $75\ \mu\text{m}$, W = $2.4\ \mu\text{m}$, SP = $1.8\ \mu\text{m}$, and N = 6 for $L_4 = 1.1\ \text{nH}$ and series resistance $R_{s,L4} = 10.4\ \Omega$. The output driver has an option to select load resistance as either $50\ \Omega$ or $70\ \Omega$, by the control register, “EN_50,” as shown in Figure 3.51. Equivalent load resistance is $50\ \Omega$ at EN_50 = 1, and $70\ \Omega$ at EN_50 = 0, combining on-resistance of MP3 (or MP4) and series resistance of L_4 . According to the EN_50 option, the tail current I_3 is also scaled to maintain the same output DC level. The designed parameters are summarized in Table 3.10; I_3 is the nominal value for EN_50 = 0.

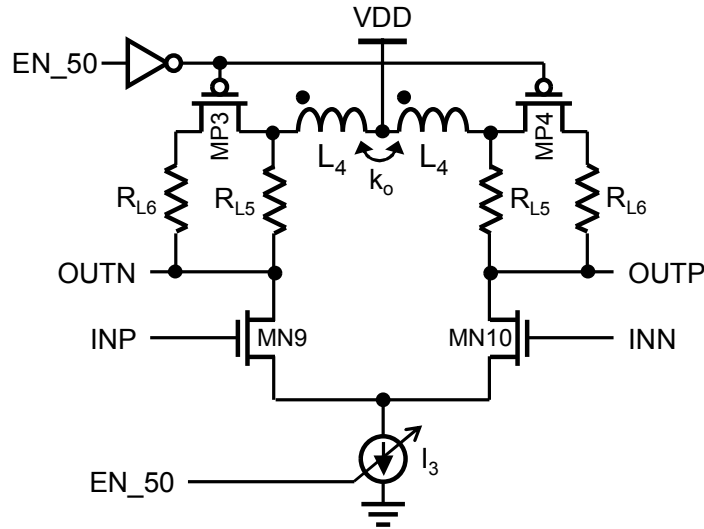


Figure 3.51. Schematic of the output driver.

Table 3.10. Designed parameters for Figure 3.51 at EN_50 = 0.

R_{L5}	R_{L6}	L_4	$R_{s,L4}$	k_0	I_3	$(W/L)_{MN9,10}$
$58\ \Omega$	$58\ \Omega$	$1.1\ \text{nH}$	$10.4\ \Omega$	0.61	$14.7\ \text{mA}$	$80\ \mu\text{m}/60\text{nm}$

Figure 3.52 shows AC simulation results with the different EN_50 values. The circuit netlist is obtained from post-layout extraction, and includes bond-pad model and all external parasitics from 1.5-mm bondwire with double-bonding, leadframe of QFN package and 1.1-inch PCB trace; the end of the PCB trace is terminated to 50 Ω . Because of the shunt-peaking inductor, the magnitude has a peaking around 10 GHz, and the attenuation rate is steep after the peak frequency. This peaking will be helpful to compensate for the magnitude attenuation at the internal buffer and becomes larger for EN_50 = 1 because of reduced load resistance. As shown in Table 3.11 for performance summary, if EN_50 = 1, 3-dB bandwidth is 16.1 GHz, but 1-dB bandwidth is 15.4 GHz, which is quite close to 3-dB bandwidth. Power consumptions are calculated as 17.2 mW for EN_50 = 0 and 21.0 mW for EN_50 = 1, both in the 1.2-V supply.

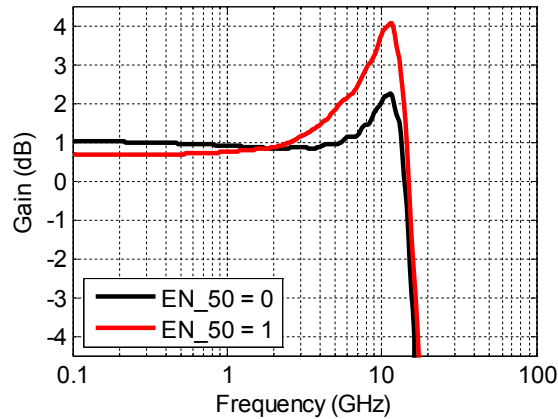


Figure 3.52. AC magnitude response of the output driver for both cases of the “EN_50” register.

Table 3.11. Performance summary of the post-layout simulation results for the output driver.

Items	Results	
	EN_50 = 0	EN_50 = 1
DC gain	1.05 dB	0.69 dB
1dB bandwidth	14.5 GHz	15.0 GHz
3dB bandwidth	15.4 GHz	16.1 GHz
Power consumption	17.2 mW	21.0 mW

3.5.6 Input termination network

Receiver inputs should be properly terminated to the system characteristic impedance: typically 50 Ω . Figure 3.53 shows the receiver termination scheme used in this design. The channel output is AC-coupled to the input of the receiver, as shown in Figure 3.53(a), to block DC voltage of the transmitter and set input DC common-mode voltage as preferable to the receiver operation. Receiver input termination is implemented by “Thevenin termination” with pull-up (R_U) and pull-down (R_D) resistors [63], which determines equivalent termination resistance (R_T) and the Thevenin equivalent termination voltage (V_T) as follows:

$$R_T = \left(\frac{1}{R_U} + \frac{1}{R_D} \right)^{-1} \quad (3.40)$$

$$V_T = V_{DD} \frac{R_D}{R_U + R_D}. \quad (3.41)$$

Thus, R_T is 50 Ω from $R_U = 75 \Omega$ and $R_D = 150 \Omega$, and V_T is 0.8 V in the 1.2-V supply voltage. Because the input stage of the receiver (CTLE) is implemented by NMOS transistors, the input termination voltage is larger than one-half VDD.

One of the commonly used matching schemes for the differential input is the “differential termination,” which places a 100- Ω resistor between two input terminals, with the center node of 100 Ω connected to the internal (or external) DC voltage. However, the differential matching scheme needs an extra on-chip bias voltage generator, which demands significant design efforts, or one more pad to use external bias voltage. Hence, in this project, a “single-ended” Thevenin termination scheme is adopted for easy and simple implementation, as shown in Figure 3.53(b), even though this scheme may consume more current. The pull-up and pull-down resistors of the termination circuit are configured by 5-bit registers (“INP_RTERM[4:0]” for positive input and “INN_RTERM[4:0]” for negative input), to tune DC input voltage in each input rail independently. The benefit of single-ended termination with 5-bit resistance control is to compensate for input-referred DC offset in the receiver chain. Figure 3.54 shows

the concept of input-referred offset cancellation. If the differential input-referred offset is V_{OFF} , which corresponds to $\pm V_{\text{OFF}}/2$ single-ended offset, input termination network can be configured to set V_T to $V_T - V_{\text{OFF}}/2$ for positive input and $V_T + V_{\text{OFF}}/2$ for negative input, to cancel out the input DC offset.

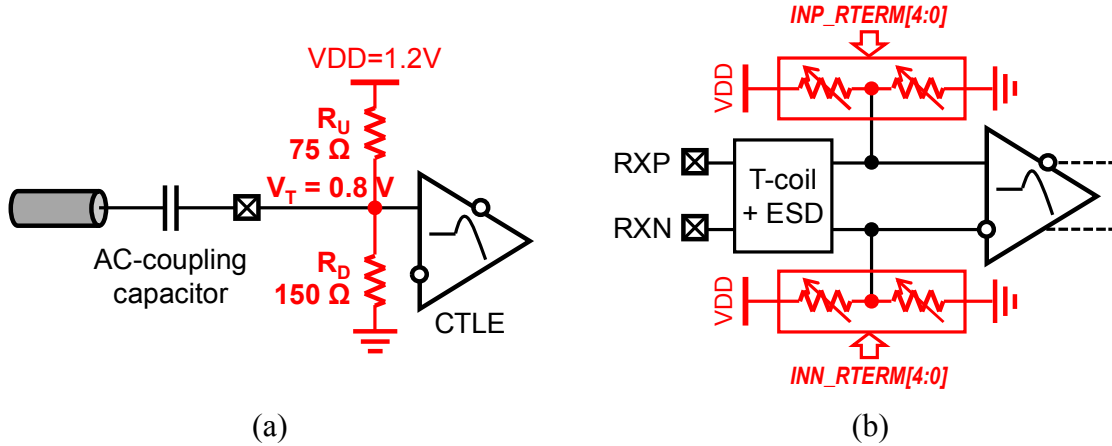


Figure 3.53. Receiver input termination using Thevenin termination. (a) Receiver input connected to the output of a channel through AC-coupled capacitor, and terminated by $75\ \Omega$ pull-up (R_U) and $150\ \Omega$ pull-down (R_D) resistors for equivalent $50\ \Omega$ and DC termination voltage (V_T) of $0.8\ \text{V}$. (b) Actual implementation of differential input termination by single-ended termination: positive and negative rail impedance to be controlled independently by 5-bit control registers, $\text{INP_RTERM}[4:0]$ and $\text{INN_RTERM}[4:0]$.

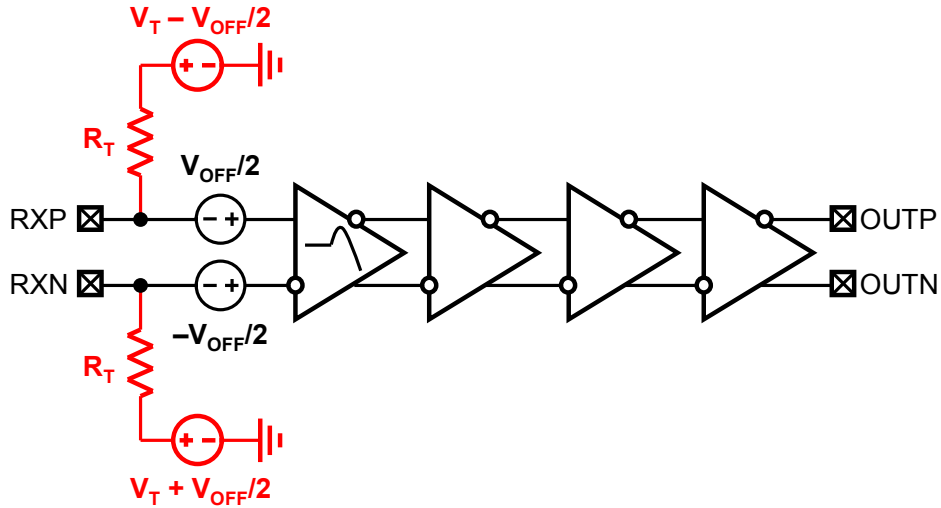


Figure 3.54. Input-referred offset cancellation by input termination network. Differential input offset of V_{OFF} ($\pm V_{\text{OFF}}/2$ in single-ended) can be canceled out by changing input termination voltage (V_T) to $V_T - V_{\text{OFF}}/2$ for positive input and $V_T + V_{\text{OFF}}/2$ for negative input.

The input termination network is an array of pull-up and pull-down resistors, as illustrated in Figure 3.55. The resistor array is controlled by the 5-bit switches (SW[5:0]), which will be connected to the control registers, INP_RTERM[4:0] and INN_RTERM[4:0], for positive and negative inputs, respectively. By changing SW[4:0] from 0 to 31 in decimal number, the pull-up resistor keeps increasing and the pull-down resistor keeps decreasing, which will decrease common-mode voltage monotonically. The termination network can be disabled by the control register, “TERM_EN,” to eliminate any unnecessary current flow when the receiver is not in use.

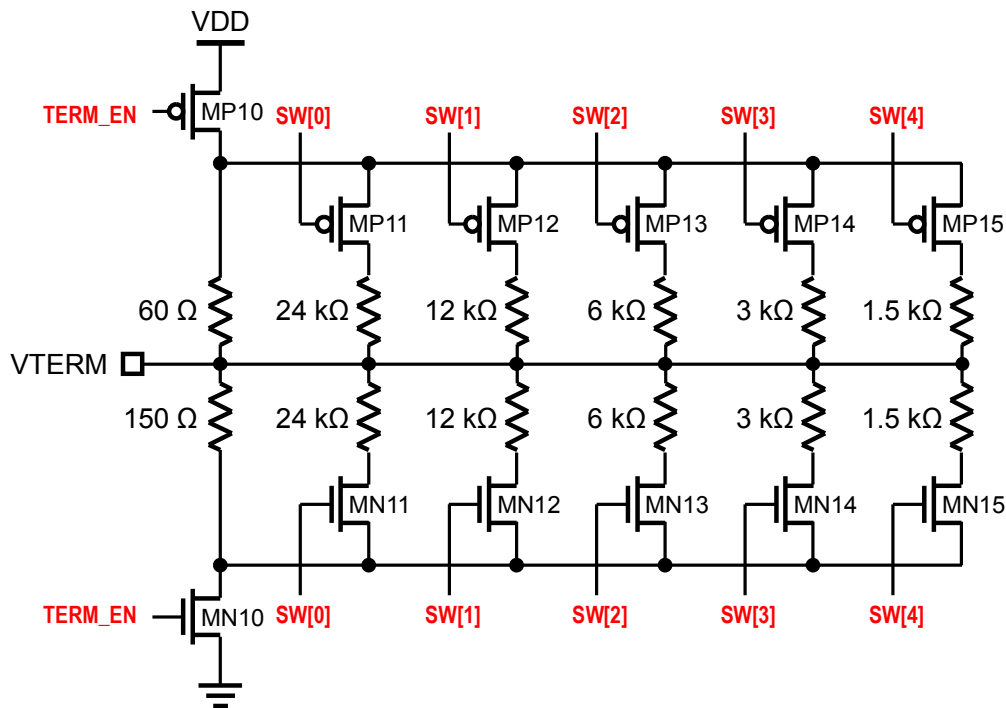


Figure 3.55. Schematic of the input termination network.

The post-layout simulation results for termination impedance and common-mode voltage are given in Figure 3.56. Transient simulation is performed to sweep the 5-bit control register from 0 to 31. Binary code for SW[4:0] has been increased by 1 every 10 ns, which may show some peaking in the output waveforms during code transition, but this does not occur in real chip operation because this control register is programmed only once at the beginning of chip operation and not reprogrammed during receiver operation. Effective termination resistance stays in the range of 51.1 – 51.9 Ω for all control

register variation, as shown in Figure 3.56(a), which demonstrates stable termination resistance in any register setting. Input common-mode voltage is controlled in the range of 770.3 – 824.6 mV by 1.7 mV step, as shown in Figure 3.56(b). When the input termination network operates in differential mode, the total input offset coverage is 108.6 mV with 3.4 mV step, which shows sufficiently fine resolution and wide coverage for any worst-case mismatches.

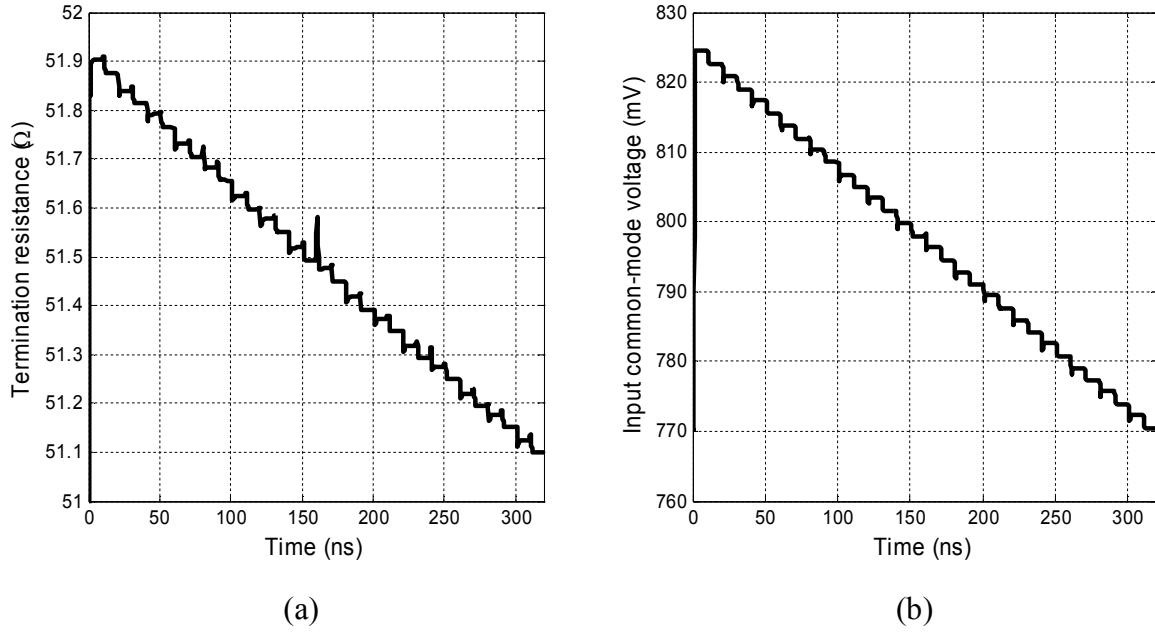


Figure 3.56. Simulation results to show (a) termination resistance variation, and (b) input common-mode voltage (termination voltage) tuning: 770.3 – 824.6 mV by 1.7 mV step. This simulation is performed by increasing 5-bit control register every 10 ns in post-layout transient simulation.

3.5.7 Overall receiver chain simulation

AC simulation for the overall receiver chain composed of the CTLE, the CTLE buffer, the internal buffer and the output driver is performed in post-layout-extracted netlists. The AC magnitude responses at the output of each stage are plotted in Figure 3.57, and summarized in Table 3.12. Because of the finite bandwidth of the subsequent stages after the CTLE, the peak frequency and the amount of peaking of the CTLE are degraded. However, such degraded performance can be recovered by $EN_{50} = 1$ at the output of the output driver. As shown in Table 3.12, the amount of peaking and peak frequency at the output driver are very similar to those at the CTLE.

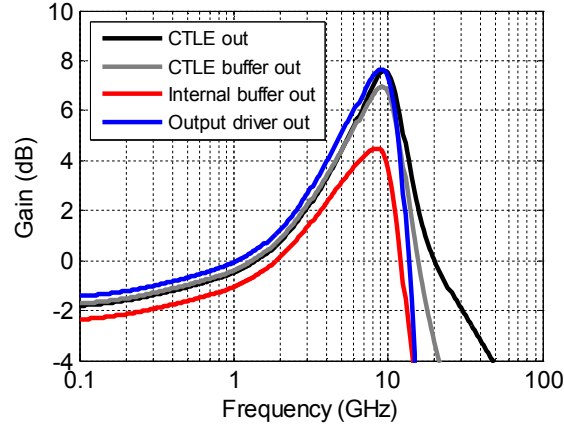


Figure 3.57. Overall AC simulation results.

Table 3.12. Performance summary of the receiver chain.

Measure Point	Peaking	DC gain	Peak freq.
CTLE	9.50 dB	-1.89 dB	9.5 GHz
CTLE buffer	8.76 dB	-1.81 dB	9.2 GHz
Internal buffer	6.93 dB	-2.44 dB	8.5 GHz
Output driver	9.15 dB	-1.49 dB	9.1 GHz

3.5.8 Receiver utilizing inductive neutralization

This section describes the third receiver split, HSRX3 in Figure 3.35. The basic receiver chain for HSRX3 is identical to that of the other two receivers: a CTLE in the front-end and the following three-stage buffers (the CTLE buffer, the internal buffer, and the output driver). However, in this CTLE, a new broadband technique, “inductive neutralization,” is used. This technique is motivated by the classic neutralization technique [25], which was invented in the 1920s, and is still used in some high-speed applications [26], [59], [64]. The schematic for a differential amplifier utilizing neutralization is illustrated in Figure 3.58. In high-speed amplifiers, the input transistors are sized large to increase current driving capability. However, large transistor size also increases the gate-drain capacitor, denoted by C_{gd} in Figure 3.58, which is known as “Miller capacitance,” to degrade the amplifier’s bandwidth by the Miller effect [65]. To overcome this Miller effect, the neutralizing capacitor (C_N) with the same size of C_{gd} is

added in a cross-coupled manner, as shown in Figure 3.58; C_N cancels the Miller effect by opposite capacitive coupling to that of C_{gd} .

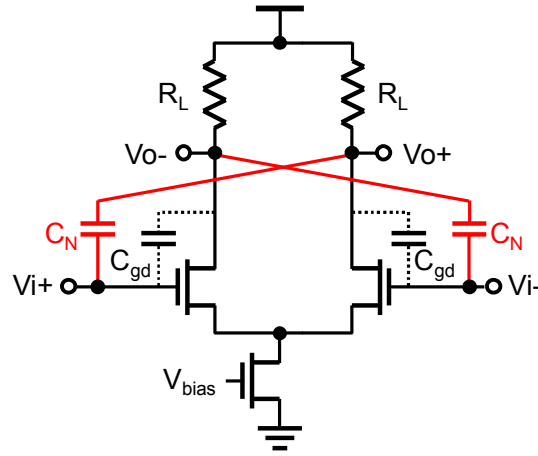


Figure 3.58. Differential amplifier with a conventional neutralization technique with neutralizing capacitors (C_N).

The basic strategy in the conventional neutralization technique with capacitive coupling is adopted in the proposed “inductive neutralization” to utilize magnetic coupling. Figure 3.59(a) illustrates the concept of inductive neutralization in a differential amplifier utilizing T-coil (L_1+L_2) at the input and a load inductor (L_{33}) for shunt-peaking; only half-circuit is shown for simplicity. When the input current flows in the incoming direction from PAD to the node V_i to increase the gate voltage of the input transistor, output voltage (V_o) should be decreased by the load current on L_{33} with the current direction as indicated in Figure 3.59(a). If magnetic coupling between the main T-coil and the load inductor (k_{13} between L_1 and L_{33} , and k_{23} between L_2 and L_{33}) occurs in the desirable direction to assist the load current flow, it can boost the switching speed of the transistor (i.e., bandwidth of the amplifier). Figure 3.59(b) depicts the construction of the proposed inductive neutralization in a T-coil. To achieve the maximum magnetic coupling (k_{13} and k_{23}), L_{33} should be placed immediately underneath the T-coil, to form a stacked transformer, which generally has the highest magnetic coupling factor in on-chip spiral transformers [66]. From Figure 3.59(b), another advantage of the inductive neutralization can be inferred; it can save layout area significantly by hiding L_{33} underneath the T-coil. Layout size comparison will be given in Section 3.6.

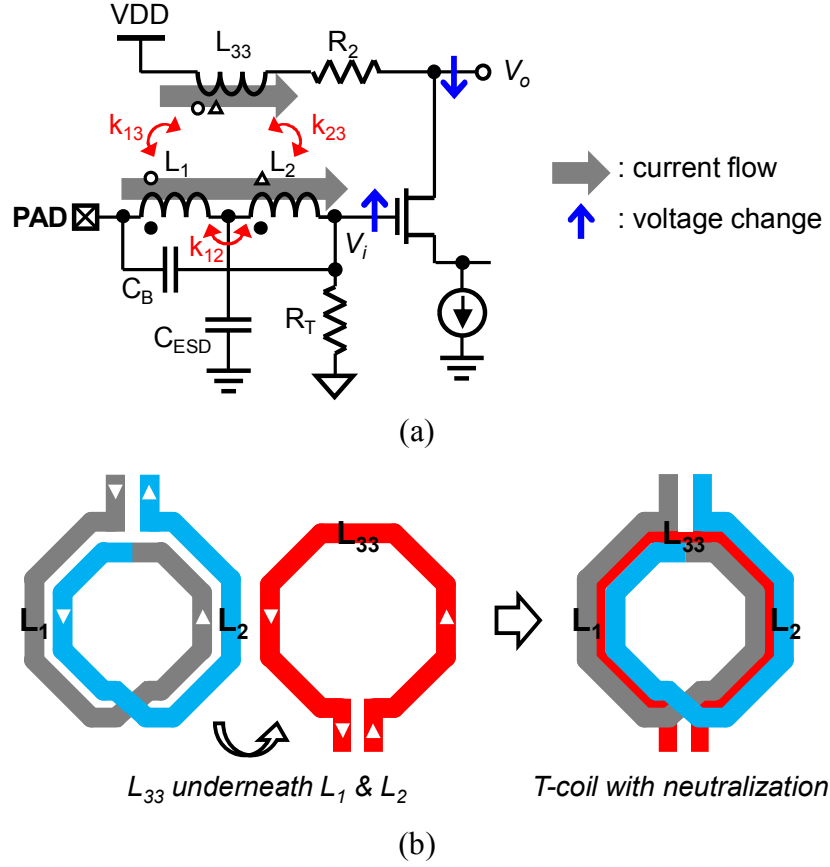
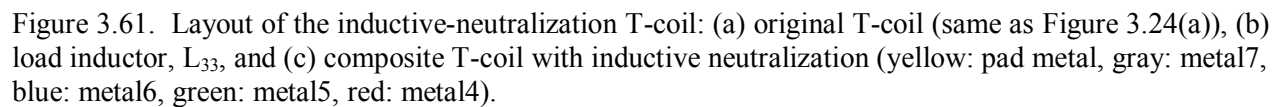


Figure 3.59. Concept of inductive neutralization: (a) magnetic coupling between the input T-coil and load inductor to assist current flow, and (b) construction of a T-coil for inductive neutralization, which place L_{33} underneath the T-coil with L_1 and L_2 .

The overall schematic of the CTLE for HSRX3, which utilizes the proposed inductive neutralization, is shown in Figure 3.60. The core circuit of the CLTE is identical to that of HSRX1 and HSRX2, but the load inductor L_{33} is magnetically coupled to each inductor leg of the T-coil, L_1 and L_2 , with a couple factor of k_{13} and k_{23} , respectively. The designed L_{33} is 1.48 nH, considering k_{13} and k_{23} of 0.81 each, and the other circuit parameters are the same as in Table 3.4..

Figure 3.61 shows the layout of the designed T-coil with inductive neutralization. The original T-coil (Figure 3.24(a)) is redrawn in Figure 3.61(a). The load inductor (L_{33}) in Figure 3.61(b) is designed by a series inductor similar to Figure 3.42, using metal5 and metal4, with dimensions of OD = 80 μm , W = 2.4 μm , SP = 1.8 μm , and N = 2.75. The composite inductive-neutralization T-coil is shown in Figure 3.61(c) and the 3-D rendered view is illustrated in Figure 3.62.



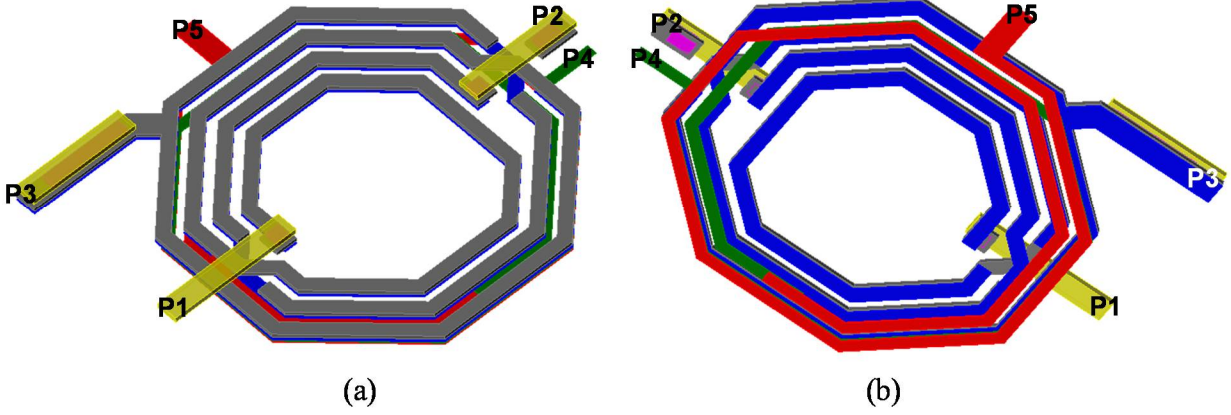


Figure 3.62. 3-D rendered view of the T-coil in Figure 3.61(c): (a) view above the T-coil, and (b) view below the T-coil. The color codes and the port names are matched to those in Figure 3.61.

3.6 IC Layout and Chip Microphotograph

The test chip is laid out in a 65-nm low-power CMOS process with seven-metal layers. The partial layouts of the two receivers, HSRX1 and HSRX2, highlighting the T-coils with ESD protection devices, are shown in Figure 3.63. HSRX1 in Figure 3.63(a) uses the inductance halving technique, while HSRX2 in Figure 3.63(b) has secondary ESD protection; the receiver circuits are identical for HSRX1 and HSRX2. The additional ESD protection circuits for each scheme are outlined by solid boxes. Only $690 \mu\text{m}^2$ of area is added for inductance halving, which includes two copies of D1 and M1 (x2) and a single trigger circuit; the secondary protection has a larger area of $1,280 \mu\text{m}^2$.

The layout of the three receiver splits (HSRX1-3) is shown in Figure 3.64. The sub-blocks in the receiver chain are marked as white boxes. HSRX3 occupies the smallest die area of $22,552 \mu\text{m}^2$ by overlapping L_3 with the main T-coil, whereas HSRX1 and HSRX2 have relatively larger areas of $29,162 \mu\text{m}^2$ and $29,752 \mu\text{m}^2$, respectively. HSRX3 reduces the die area by 24%, compared to HSRX2.

The full chip layout is shown in Figure 3.65; the entire die area is $1 \text{ mm} \times 1 \text{ mm}$. This test chip is shared with another research project for system-level ESD study [67], which is marked as gray boxes at the top side of the chip layout, and will not be covered in this dissertation. Two types of bond-pads are used: one in a smaller square shape of $50 \mu\text{m} \times 50 \mu\text{m}$, and the other in a larger rectangular shape of 85

$\mu\text{m} \times 50 \mu\text{m}$, which is intentionally made larger for Kelvin probing during on-die VF-TLP testing. At the center of the bottom side of Figure 3.65, differential outputs of the output driver (RXOP and RXON) have two smaller pads per output signal, for double-bonding to reduce bond-wire inductance. To filter out power supply noise, a total of 163-pF decoupling capacitors are inserted in vacant areas. The chip microphotograph is shown in Figure 3.66.

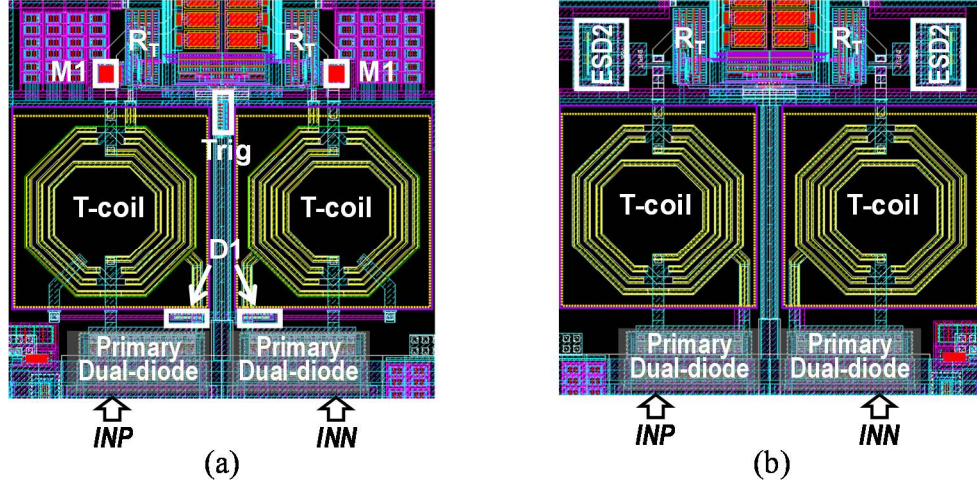


Figure 3.63. Layout of the T-coils and ESD protection devices for (a) HSRX1: T-coil with inductance halving (labeled as D1, M1 and Trig) and (b) HSRX2: T-coil with secondary ESD protection (labeled as ESD2).

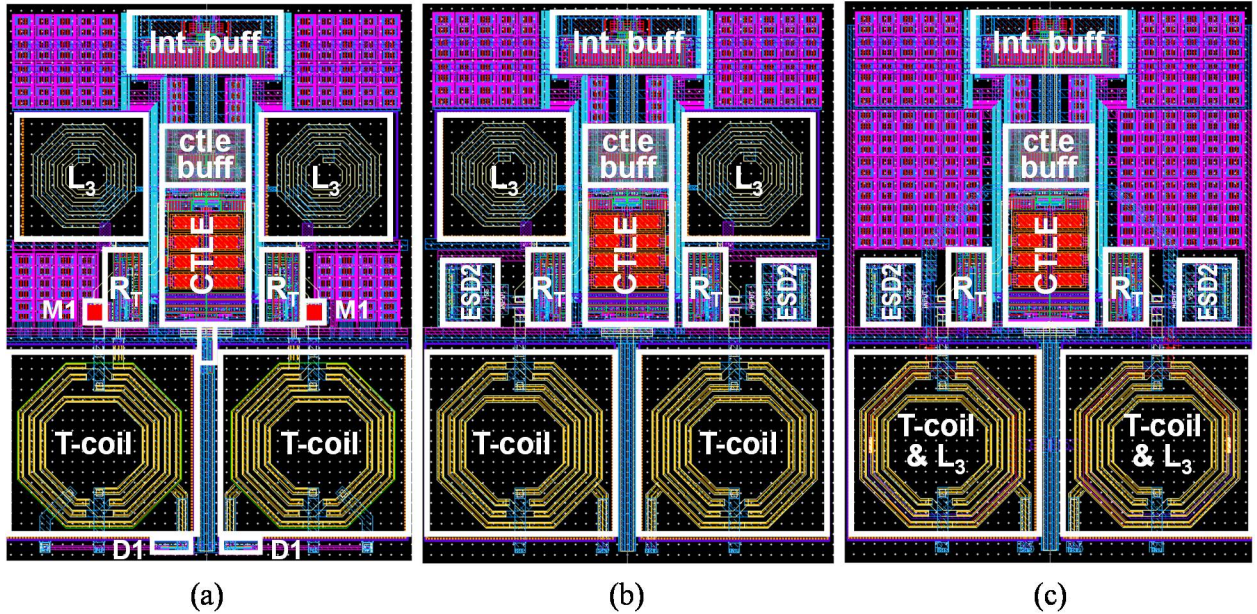


Figure 3.64. Layout of the receiver splits: (a) HSRX1, (b) HSRX2 and (c) HSRX3.

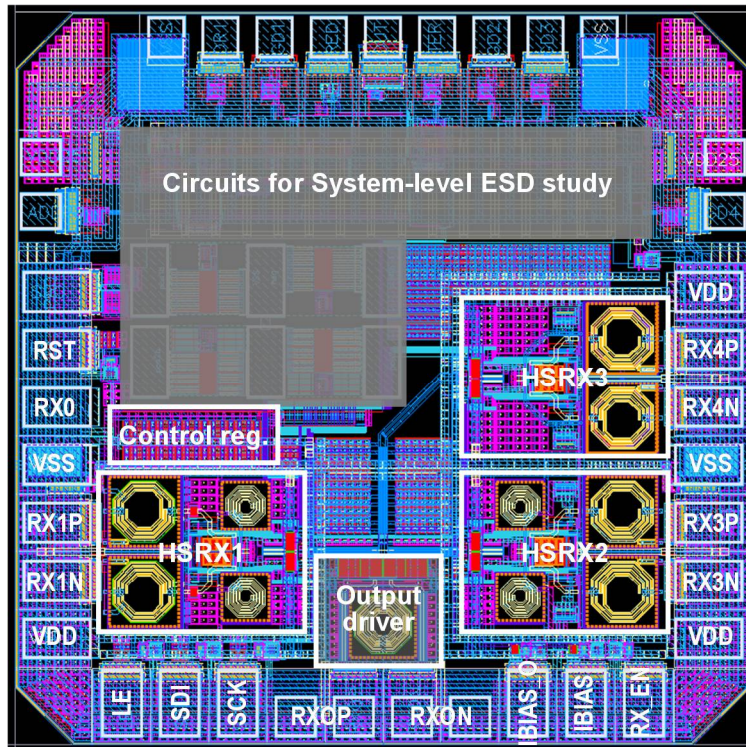


Figure 3.65. Full-chip layout. High-speed receivers with T-coils are placed at the bottom side, and another project related to system-level ESD study is also implemented at the top of the chip.

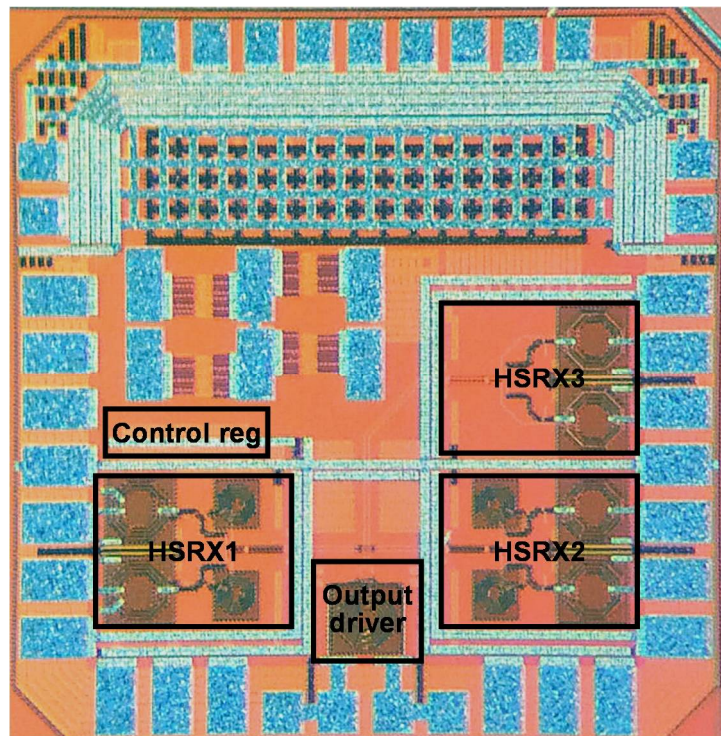


Figure 3.66 Chip microphotograph.

Table 3.13. T-coil circuit split.

	Primary ESD	Peak cancellation
Inductance halving (HSRX1)	Yes	Yes (ind. halv)
Secondary protection (HSRX2)	Yes	Yes (secondary prot.)
Uncompensated T-coil	Yes	No
No T-coil	Yes	No

Figure 3.68 shows the simulated voltage waveforms at the gate of the NMOS input transistor (i.e., node X1 in Figure 3.67). The inductance-halving T-coil suppresses the CDM voltage overshoot that is seen in the uncompensated T-coil, which shows that the proposed scheme works as intended. The secondary-protection scheme provides the best CDM performance because it contains a voltage divider that reduces the quasi-static voltage at X1; also, the low-pass filter formed by the series resistor and the capacitance of the secondary diodes filters out the high-frequency peak signal. The peak gate-source voltage ($V_{GS,peak}$, the peak voltage difference between X1[X2] and Y1[Y2] in Figure 3.67) for the receiver input NMOS (MN1 or MN2 in Figure 3.67) are found to be 3.55 V and 2.82 V, for T-coils with inductance-halving and secondary protection, respectively. The inductance-halving T-coil suppresses the peak voltage by 2 V (36%) relative to the uncompensated T-coil (5.55 V). In case of no T-coil, as shown in Figure 3.68, no overshoot is observed at the end of pulse transition, but there exists the initial peak, which is usually generated by the fast transient ESD pulse at the ESD protection devices [7], [68]. The simulated $V_{GS,peak}$ is 3.98 V, which is 0.43 V larger than that of the T-coil with inductance halving.

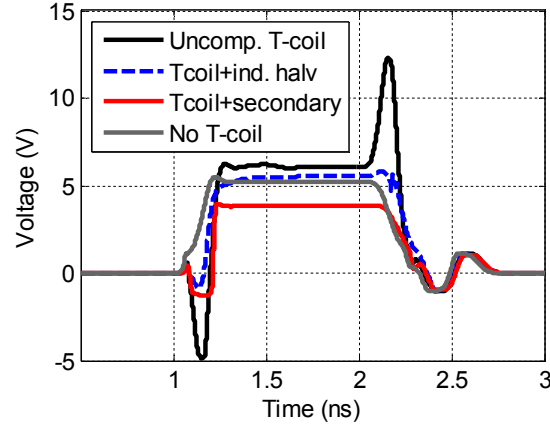


Figure 3.68. VF-TLP voltage waveforms at the receiver input (X1).

The simulated AC gain response of the normalized transimpedance is plotted in Figure 3.69. The bandwidth of the receiver with inductance-halving is 26 GHz, 33% larger than that of the receiver with secondary protection (19.6 GHz), although it is reduced by 16% relative to that of the receiver with an uncompensated T-coil (30.9 GHz). Nevertheless, the T-coil with inductance-halving still achieves $3.9\times$ larger bandwidth than a receiver with no T-coil (6.6 GHz), and its 26 GHz bandwidth is large enough to support data-rates in excess of 25 Gb/s without compromising ESD reliability.

Figure 3.69(b) shows the simulation results for the magnitude of S_{11} , which corresponds to the return loss but with an opposite sign, to compare the impedance matching performance. As a measure of comparison, a frequency point that has -10 dB return loss is defined as “ f_{RL10} .” In Figure 3.69(b), both inductance-halving and uncompensated T-coils show very similar f_{RL10} : $f_{RL10} = 18.5$ GHz for inductance halving and 19.2 GHz for the uncompensated T-coil, which are almost $4\times$ larger than that of no T-coil (4.5 GHz). However, the T-coil with secondary protection has only $f_{RL10} = 14.7$ GHz, which is 21% worse than the inductance halving case.

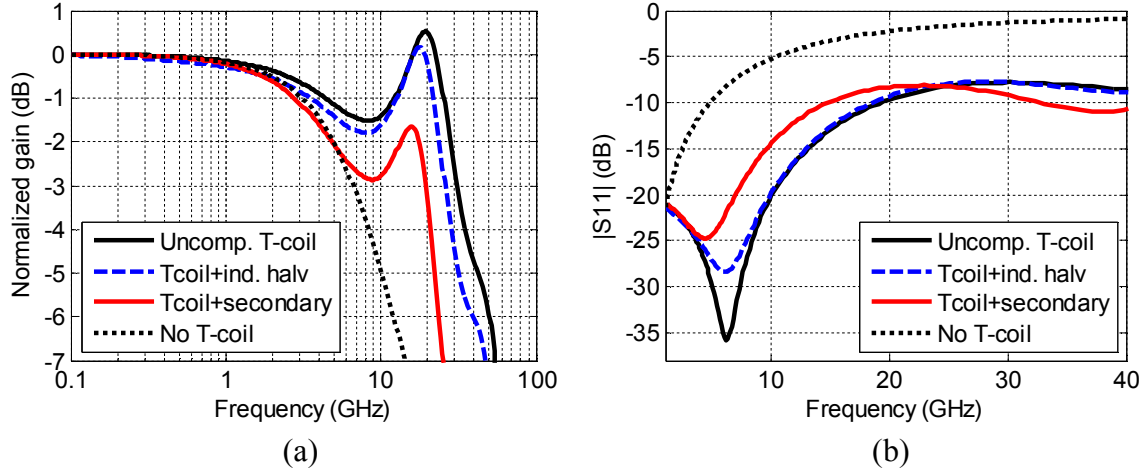


Figure 3.69. (a) AC magnitude responses at the receiver input (X1), and (b) impedance matching by the magnitude of S11.

By now, the simulations are performed by monitoring at the input of the receiver circuits to compare the performance of each individual T-coil. However, it will be beneficial to know how the different T-coil splits behave in the context of the actual receiver circuits with the equalization function. For this purpose, the transient simulations are performed with the entire receiver chain in Figure 3.35.

Figure 3.70 shows the channel loss used in the eye diagram simulations. It is composed of (1) 16-inch FR4 channel in RLGC model, (2) the parasitics of a 40-pin QFN package leadframe and a 1.8-mm bondwire in S-parameters extracted by EM simulations, (3) bond-pad parasitic, and (4) 300- μm dual-diodes by PDK model. As shown in Figure 3.70(a), the channel loss is -21 dB at 12.5 GHz, one-half baud-rate for 25 Gb/s. The transmit data of 2^7-1 PRBS in 400-mV_{ppd} amplitude is transmitted at 25 Gb/s through the channel modeled above, and the eye diagram in Figure 3.70(b) is monitored at the end of the channel. Because of the heavy channel loss, the incoming data eye is completely closed as shown in Figure 3.70(b).

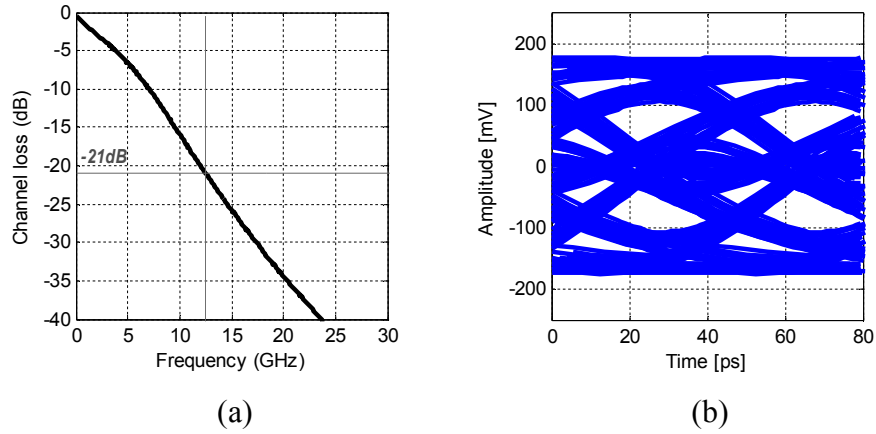


Figure 3.70. Simulation results for the 16-inch channel used in the eye diagram simulations: (a) channel loss of -21 dB at 12.5 GHz, and (b) eye diagram at the channel output. The channel includes 16-inch FR4 trace, 1.8-mm bondwire, leadframe parasitic of 40-pin QFN package, bond-pad and primary ESD protection diodes.

Figure 3.71 shows the eye diagrams from four different receivers that use the T-coils (or no T-coil), as given in Table 3.13. From the leftmost column in Figure 3.71, each column shows the eye diagrams of HSRX1, HSRX2, the receiver with the uncompensated T-coil, and the receiver without T-coil, respectively. The eye diagrams are measured at the two output points: the CTLE, which are shown in the first-row plots in Figure 3.71(a)-(d) and the output driver, which are shown in the second-row plots in Figure 3.71(e)-(h). The calculated eye height and eye width from the simulated eye diagrams are summarized in Table 3.14. The eye diagram at the CTLE with no T-coil in Figure 3.71(d) shows fairly improved eye diagrams compared to the channel output eye in Figure 3.70(b), with eye height of 80.6 mV and eye width of 31.4 ps. The receiver with the uncompensated T-coils in Figure 3.71(c) improves the eye diagram significantly to 124 mV eye height and 35.9 ps eye width. This improvement will be degraded to some extent, when the peaking suppression techniques are adopted in HSRX1 and HSRX2. As expected from the bandwidth performance in Figure 3.69(a), HSRX1 can outperform HSRX2 with the wider eye height of 101 mV, which is a 31% improvement relative to 77.0 mV for HSRX2. Eye widths are improved with the same trends as eye heights; HSRX1 shows a wider eye width than HSRX2, but a narrower eye than the uncompensated T-coil. Interestingly, the eye performance of HSRX2 is slightly

worse than that of no T-coil; more optimization may be needed for HSRX2, which will be discussed in Section 3.7.2.

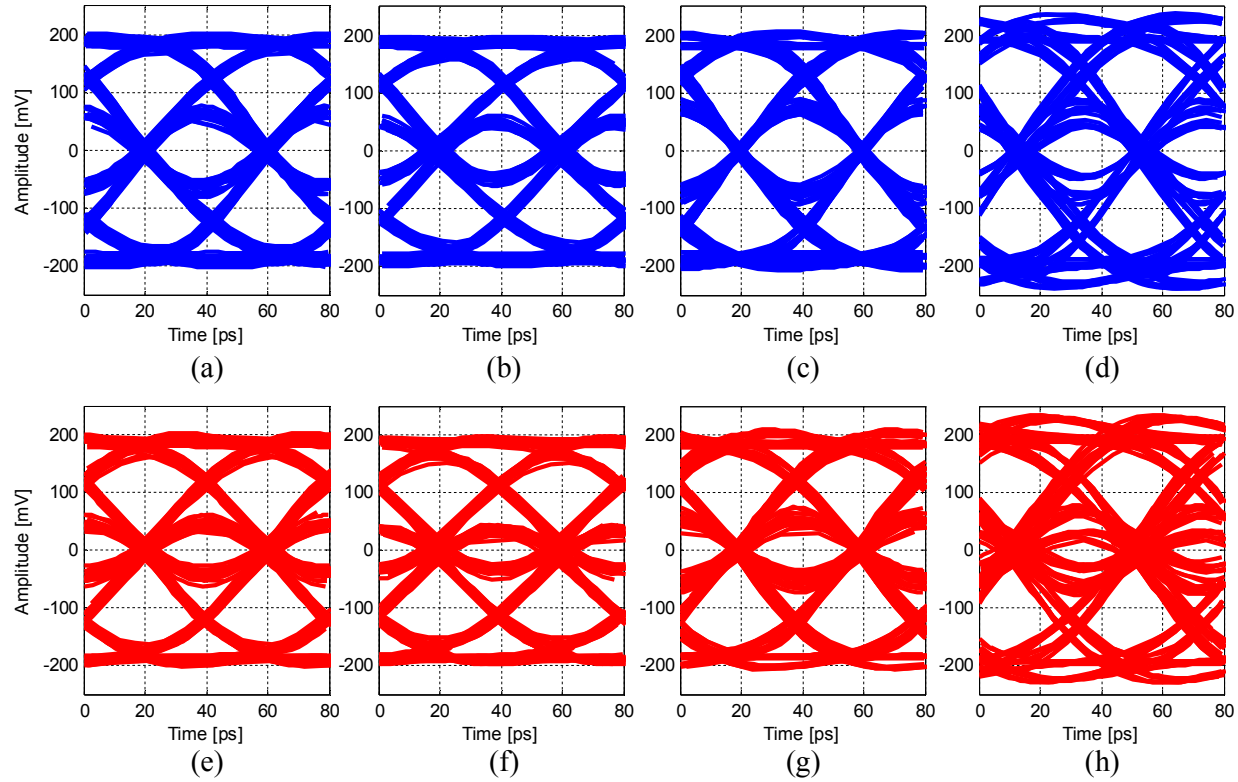


Figure 3.71. Eye diagram simulation results at 25 Gb/s on the 16-inch channel in Figure 3.70, at the CTLE outputs for (a) HSRX1, (b) HSRX2, (c) uncompensated T-coil and (d) no T-coil, and eye diagrams measured at the output driver for (e) HSRX1, (f) HSRX2, (g) uncompensated T-coil and (h) no T-coil.

Table 3.14. Eye diagram simulation results at 16-inch channel with -21 dB loss.

Items	@ CTLE		@ Output driver	
	Eye height	Eye width	Eye height	Eye width
Inductance halving (HSRX1)	101 mV	33.3 ps	58.2 mV	31.1 ps
Secondary protection (HSRX2)	77.0 mV	30.8 ps	39.1 mV	27.8 ps
Uncompensated T-coil	124 mV	35.9 ps	63.0 mV	32.2 ps
No T-coil	80.6 mV	31.4 ps	42.2 mV	23.0 ps

The single-bit response (SBR) is a good measure of equalizer performance to show the amount of ISI before and after the current transmit bit (cursor), which are called as pre-cursor ISI and post-cursor ISI, respectively. Figure 3.72 shows the SBR plots for the four receivers with the same order as in Figure 3.71;

the amplitudes of the pre- and post-cursor ISI are normalized to that of the cursor at the CTLE output (the blue lines). The SBR measured at the channel output shows at least three post-cursor ISI for the receivers with T-coils (Figure 3.72(a)-(c)), and an additional pre-cursor ISI for the receiver with no T-coil (Figure 3.72(d)). After the CTLE, most of the ISI terms are removed with only the first pre- and post-cursor ISI left in Figure 3.72(a)-(c). The ISI reduction in no T-coil case (Figure 3.72(d)) is relatively less effective compared to the other three cases. Moreover, it has negative second post-cursor ISI, which shows over-equalization. Generally, this negative ISI can deteriorate the pulse transition as shown in Figure 3.71(d), but sometimes it may be helpful to reduce the other positive ISI. This is why the eye diagram of the no T-coil receiver shows slightly better eye opening than that of HSRX2.

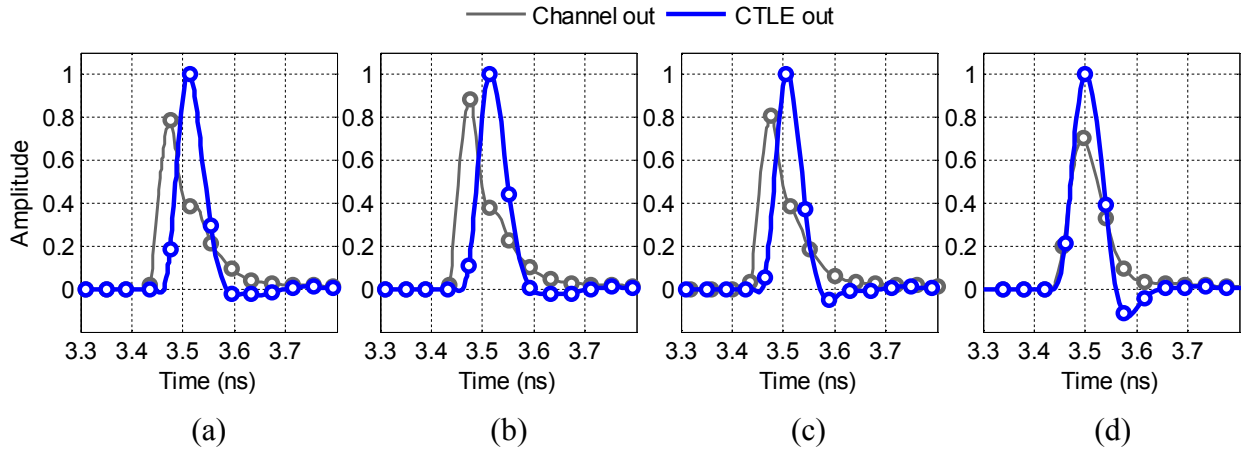


Figure 3.72. Single-bit responses (SBR) for (a) HSRX1, (b) HSRX2, (c) uncompensated T-coil and (d) no T-coil (gray: SBR at the channel out, blue: SBR at the CTLE out).

The same eye diagram analysis has been performed for the 10-inch channel in Figure 3.73; the other input channel parasitics are identical to those in the 16-inch channel in Figure 3.70, except for the FR4 channel length. The channel loss at 12.5 GHz is -15.8 dB, which shrinks the incoming eye with 48 mV eye height and 26 ps eye width, as shown in Figure 3.73 (b). Figure 3.74 shows the eye diagrams on the channel in Figure 3.73, with the same setup as Figure 3.71. The eye diagram performances are summarized in Table 3.15. The same performance trends are maintained in this -15.8 dB loss channel case.

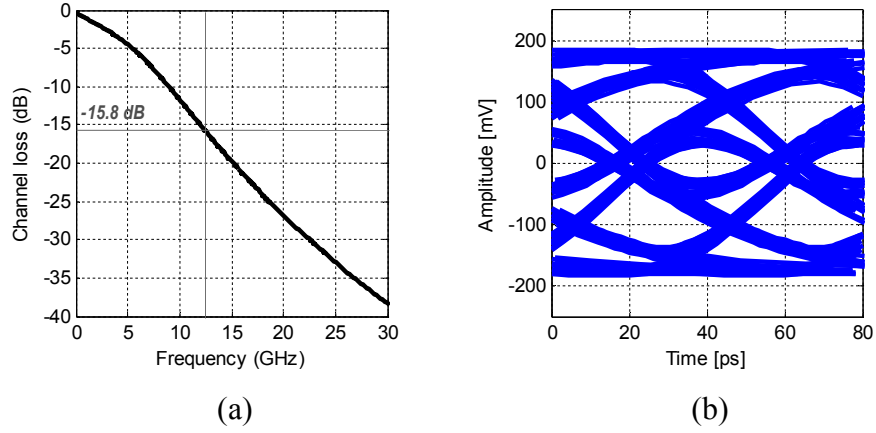


Figure 3.73. Simulation results for the 10-inch channel used in the eye diagram simulations: (a) channel loss, and (b) eye diagram at the channel output. The same parasitic components as Figure 3.70 are included except for the channel length.

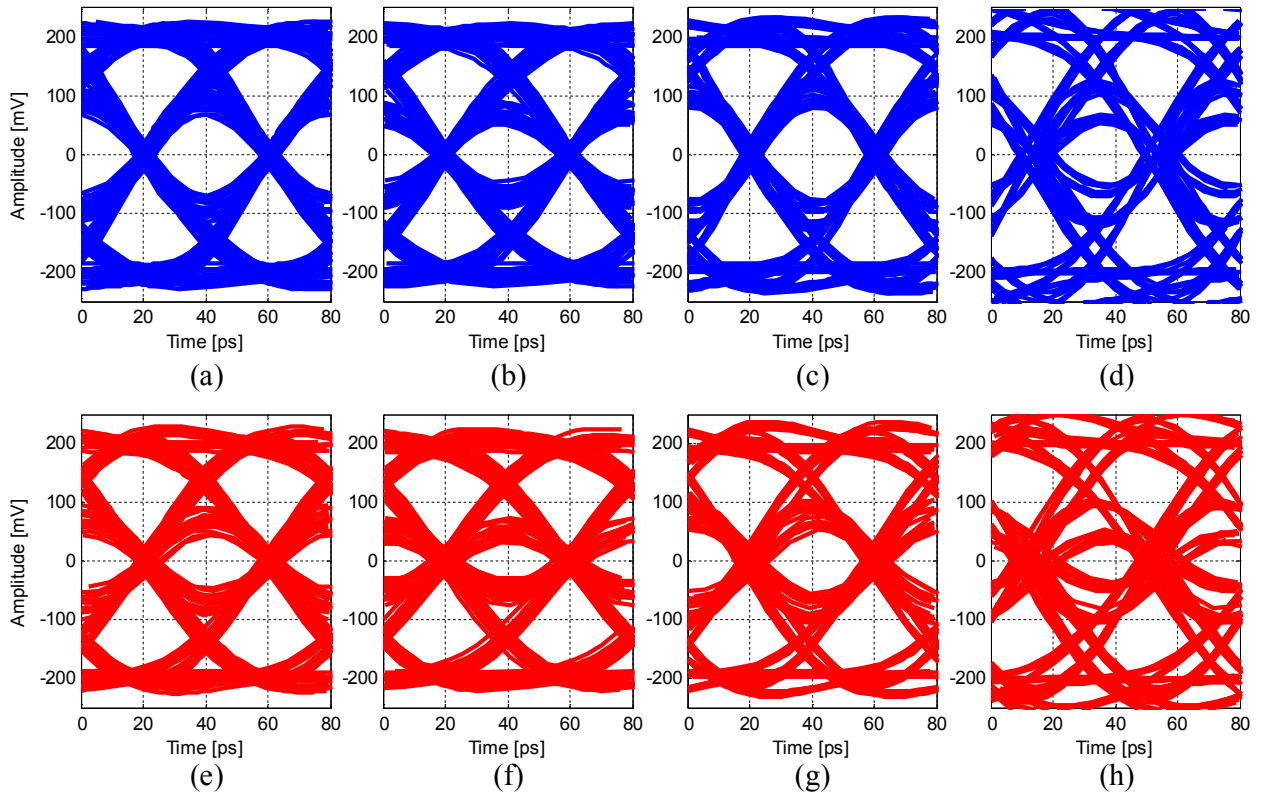


Figure 3.74. Eye diagram simulation results at 25 Gb/s on the 10-inch channel in Figure 3.73 at the CTLE output for (a) HSRX1, (b) HSRX2, (c) uncompensated T-coil and (d) no T-coil, and eye diagrams measured at the output driver for (e) HSRX1, (f) HSRX2, (g) uncompensated T-coil and (h) no T-coil.

Table 3.15. Eye diagram simulation results at 10-inch channel with -15.8 dB loss.

Items \ Circuit splits	@ CTLE		@ Output driver	
	Eye height	Eye width	Eye height	Eye width
Inductance halving (HSRX1)	137 mV	34.8 ps	88.1 mV	30.7 ps
Secondary protection (HSRX2)	96.5 mV	34.1 ps	60.0 mV	28.9 ps
Uncompensated T-coil	155 mV	33.1 ps	106 mV	30.3 ps
No T-coil	103 mV	27.9 ps	67 mV	23.6 ps

In summary, inductance halving removes voltage overshoot due to magnetic coupling in the T-coil and, relative to secondary protection, it performs better in signal transmission, both in frequency- and time-domain, and occupies less die area (46%); thus, it may be the better choice for high-speed applications. The overall performance for all four receiver splits is summarized in Table 3.16, and plotted in the dual y-axis plot in Figure 3.75.

Table 3.16. Performance summary for all four split cases.

Items \ Circuit splits	$V_{GS,peak}$	f_{-3dB}	f_{RL10}	Eye diagram (height /width)	Area overhead
Uncompensated T-coil	5.55 V	30.9 GHz	19.2 GHz	124 mV / 35.9 ps	-
Inductance halving (HSRX1)	3.55 V	26.0 GHz	18.5 GHz	101 mV / 33.3 ps	690 μm^2
Secondary protection (HSRX2)	2.82 V	19.6 GHz	14.7 GHz	77.0 mV / 30.8 ps	1280 μm^2
No T-coil	3.98 V	6.6 GHz	4.8 GHz	80.6 mV / 31.4 ps	-

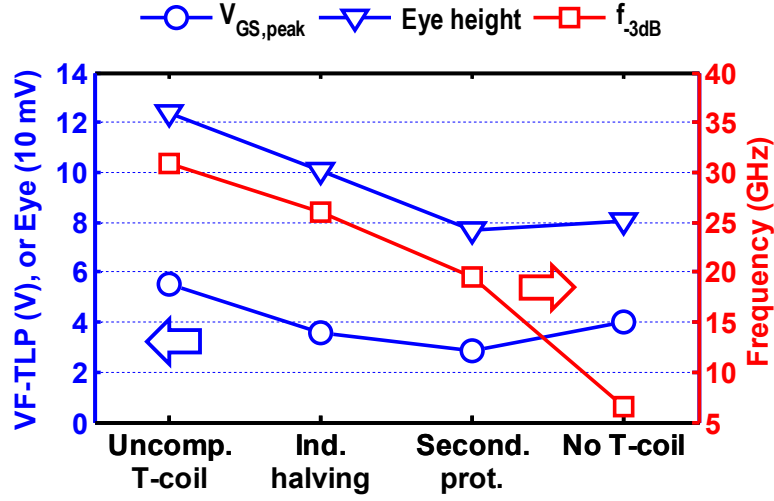


Figure 3.75. Performance summary plot for all four split cases. The VF-TLP ($V_{GS,peak}$) and eye height results are plotted on the left y-axis, and bandwidth is plotted on the right y-axis.

3.7.2 Optimization of secondary ESD protection

Based on the designed parameters for secondary ESD protection in the previous sections, further optimization study will be performed to achieve the best performance. The two design parameters for secondary ESD protection are a series resistor, R_{ser} , and a perimeter of secondary dual-diodes, P_{dio} (refer to Figure 3.27), which will be decreased for a better bandwidth and impedance matching performance. The split cases are summarized in Table 3.17. From the default design (Case0), only P_{dio} is decreased to 25 μm in Case1, and only R_{ser} is decreased to 20 Ω in Case2, with the other parameter kept the same. Case3 has both parameters decreased with the same amount as Case1 and Case2, and Case4 has no series resistance with the 25 μm P_{dio} .

Table 3.17. Split cases for optimization of secondary ESD protection.

	R_{ser}	P_{dio}
Case0	50 Ω	40 μm
Case1	50 Ω	25 μm
Case2	20 Ω	40 μm
Case3	20 Ω	25 μm
Case4	0 Ω	25 μm

From the split conditions, the same simulations as Section 3.7.1 are performed: VF-TLP, bandwidth and return loss. Figure 3.76(a) and (b) show the VF-TLP waveforms at the gate of the input transistor, and the calculated V_{GS} of the input transistor, respectively. As the size of secondary ESD protection decreases from Case0 to Case4, the quasi-static voltage level increases at the gate of the input NMOS, as shown in Figure 3.76(a). The same trend is observed in Figure 3.76(b) at the end of the pulses. However, the peak gate-source voltage ($V_{GS,peak}$) occurs at the beginning of the waveforms in Figure 3.76(b), which is similar for all the split cases. The minimum $V_{GS,peak}$ is 2.82 V for Case0 and the maximum $V_{GS,peak}$ is 2.96 V for Case4, as given in Table 3.18.

AC magnitude and impedance matching are simulated as presented in Figure 3.77(a) and (b). With the decreased parameter values in secondary ESD protection, bandwidth of the T-coil can be improved up to 26.5 GHz for Case4, which is comparable to that of the inductance-halving T-coil. With $R_{ser} = 0 \Omega$ and decreased P_{dio} , secondary protection only provides pure small capacitance at the receiver input, which can achieve higher bandwidth. Figure 3.77(b) gives S11 magnitude response. The frequency points where S11 becomes -10 dB (f_{RL10}) are around 13-15 GHz for all the cases (see Table 3.18), but $\sim 20\%$ smaller than that of inductance halving (18.5 GHz).

The overall performance between different split cases for secondary ESD protection is summarized in Table 3.18 and plotted in Figure 3.78. For comparison, the performance of inductance halving is also plotted in Figure 3.78. Bandwidth and voltage overshoot are improved by decreasing R_{ser} and P_{dio} , but not for f_{RL10} . This fact shows that the inductance-halving scheme is still competitive considering overall performance metrics, against the best optimized secondary protection.

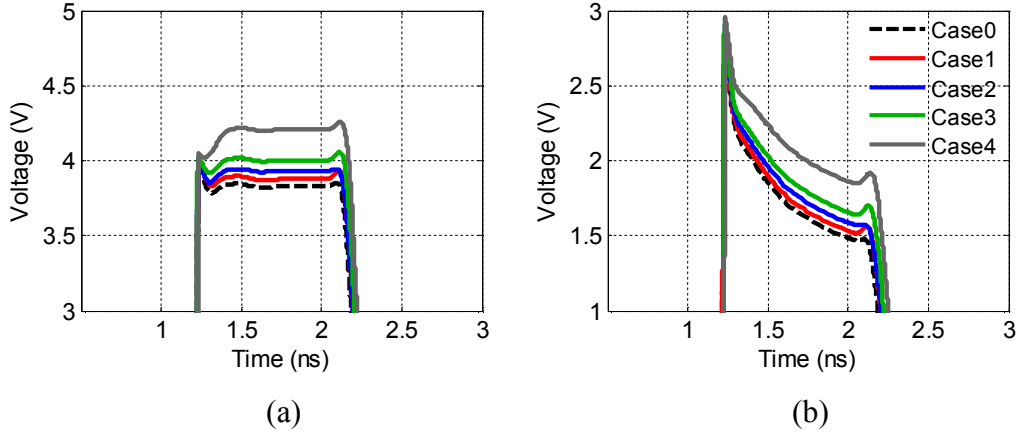


Figure 3.76. VF-TLP waveforms for the different secondary-protection cases: (a) voltage waveforms at the input transistor gate (X1 in Figure 3.67) and (b) gate-to-source voltage (V_{GS}) of the input NMOS transistor.

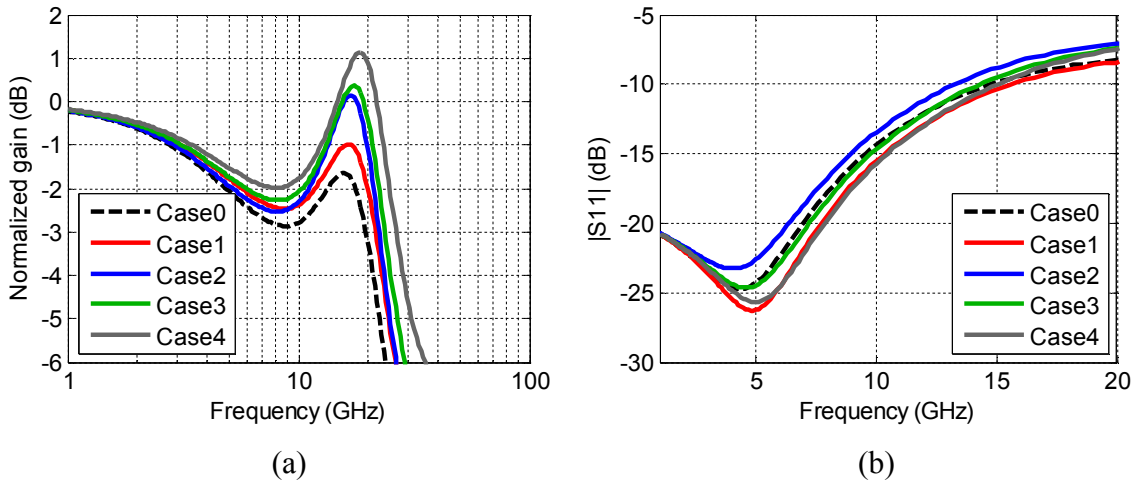


Figure 3.77. (a) AC magnitude response and (b) return-loss (S11), for all the secondary-protection cases.

Table 3.18. Performance summary of the different secondary ESD protection cases.

	$V_{GS,peak}$	f_{-3dB}	f_{RL10}
Case0 ($R_{ser} = 50 \Omega$, $P_{dio} = 40 \mu m$)	2.82 V	19.6 GHz	14.7 GHz
Case1 ($R_{ser} = 50 \Omega$, $P_{dio} = 25 \mu m$)	2.82 V	21.6 GHz	15.6 GHz
Case2 ($R_{ser} = 20 \Omega$, $P_{dio} = 40 \mu m$)	2.90 V	22.3 GHz	13.3 GHz
Case3 ($R_{ser} = 20 \Omega$, $P_{dio} = 25 \mu m$)	2.91 V	23.7 GHz	14.3 GHz
Case4 ($R_{ser} = 0 \Omega$, $P_{dio} = 25 \mu m$)	2.96 V	26.5 GHz	15.0 GHz

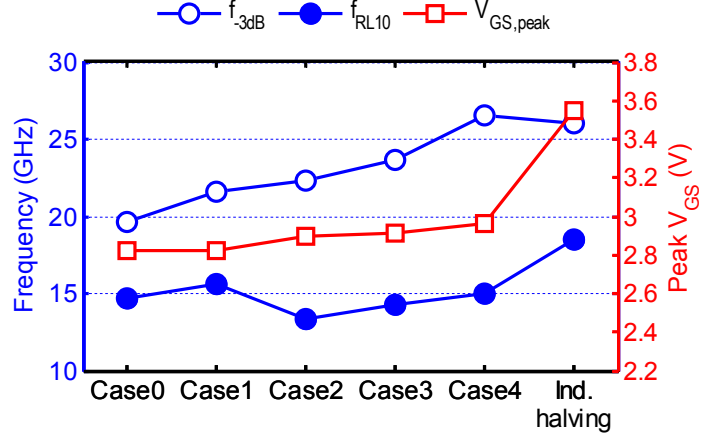


Figure 3.78. Performance summary: plots of Table 3.18.

3.7.3 Performance of inductive neutralization

The receiver utilizing inductive neutralization, HSRX3, is simulated and compared to the receiver without the inductive neutralization technique (HSRX2). Eye diagrams are simulated with 2^7-1 PRBS data at 20 Gb/s on the 10-inch channel, as characterized in Figure 3.73. HSRX3 was initially targeted for 25-Gb/s, but the T-coil and the load inductor are not fully optimized for the data-rate. Thus, simulation is performed at a lower data-rate and less lossy channel.

Figure 3.79 shows the SBR plots for HSRX3 and HSRX2, at the channel output (gray) and the CTLE output (blue). The channel output of HSRX3 in Figure 3.79(a) is distorted, compared to that of HSRX2 in Figure 3.79(b). As inductive neutralization also influences the current flow in the input T-coil, the SBR of the channel output can be changed. Furthermore, the CTLE output for HSRX3 (Figure 3.79(a)) shows overcompensation by the negative value at the first pre-cursor ISI, which potentially deteriorate the eye diagram.

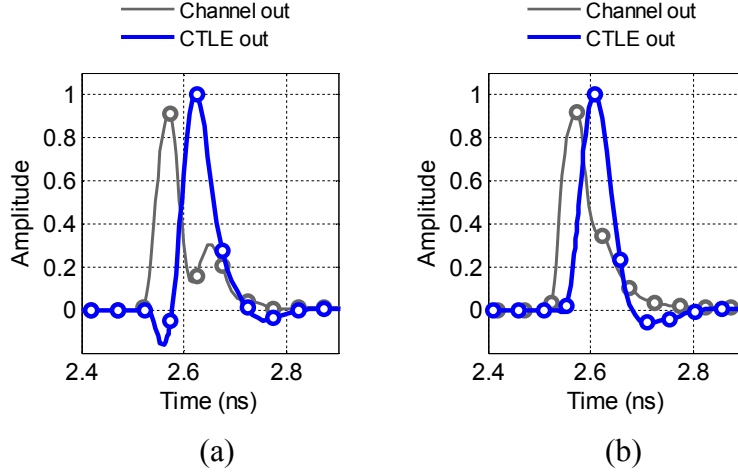


Figure 3.79. Single-bit responses for (a) HSRX3 and (b) HSRX2.

Figure 3.80 shows the eye diagram simulation results; the plots in the first column (Figure 3.80(a) and (c)) are for HSRX3 and the second-column plots (Figure 3.80(b) and (d)) show eye diagrams for HSRX2. The eye diagrams in the first row are measured at the output of the CTLE, and the second-row eye diagrams are obtained at the output driver. The eye diagram simulation results are summarized in Table 3.19. HSRX3 shows 207.9 mV eye height at the CTLE output, which is about 6% larger than that of HSRX2. However, the eye width of HSRX3 is about 6% worse than that of HSRX2, which is due to the multiple transitions on the eye (see Figure 3.80(a)). This is because the designed inductance and magnetic coupling factor are not fully optimized, thereby leading to over-peaking on the magnitude response of the CTLE, which can be anticipated from the SBR in Figure 3.79(a). In comparison, the eye diagram at the output driver shows that the eye widths of the two receivers are very similar. This is because the limited bandwidth of the subsequent stages can suppress unnecessary peaking of the CTLE and eliminate multiple transitions at time-domain.

Therefore, it is proved that the proposed inductive neutralization works. However, more optimization is needed for the T-coil and the load inductor. By locating the load inductor underneath the T-coil, the original magnetic coupling factor inside the T-coil (i.e., k_{12} , coupling between L_1 and L_2 in Figure 3.60) will be changed. Hence, the original T-coil should also be redesigned.

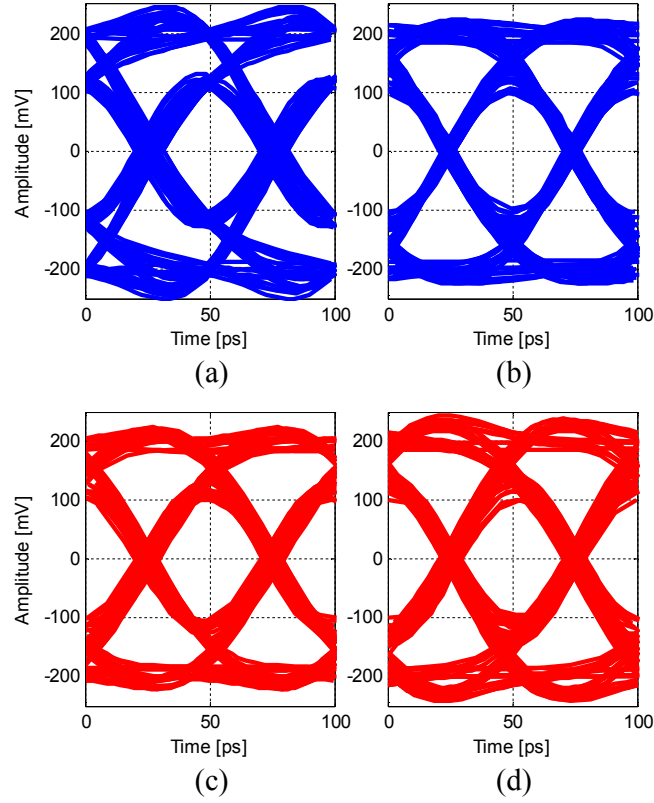


Figure 3.80. Eye diagram simulation results at 20 Gb/s on 10-inch channel. Eye diagrams measured at the CTLE for (a) HSRX3 and (b) HSRX2, and measured at the output driver for (c) HSRX3 and (d) HSRX2.

Table 3.19. Performance summary for the eye diagram simulation results.

Splits \ Items	@ CTLE		@ Output driver	
	Eye height	Eye width	Eye height	Eye width
HSRX3	207.9 mV	38.3 ps	204.7 mV	40.6 ps
HSRX2	196.8 mV	43.8 ps	194.4 mV	41.1 ps

3.8 Test Board Design

This test chip is assembled with a 40-pin QFN package with 5 mm × 5 mm outer dimension and 0.4 mm pin-pitch. Figure 3.81 shows the schematic for the printed circuit board (PCB). Unrelated pins at the top of the QFN package for the system-level ESD study are properly disabled by grounding or floating. High-speed line traces are marked in red; they require 50-Ω impedance control. Register control is performed externally by microcontroller board (MCU board) via three control signals (LE_uC, SDI_uC

and SCK_uc), which should be level-down-converted from 3.3 V to 1.2 V swing using the level shifter IC. Supply voltages for 2.5 V and 1.2 V are generated on the board using voltage regulator ICs. Bias currents are externally generated by potentiometers to set the necessary currents.

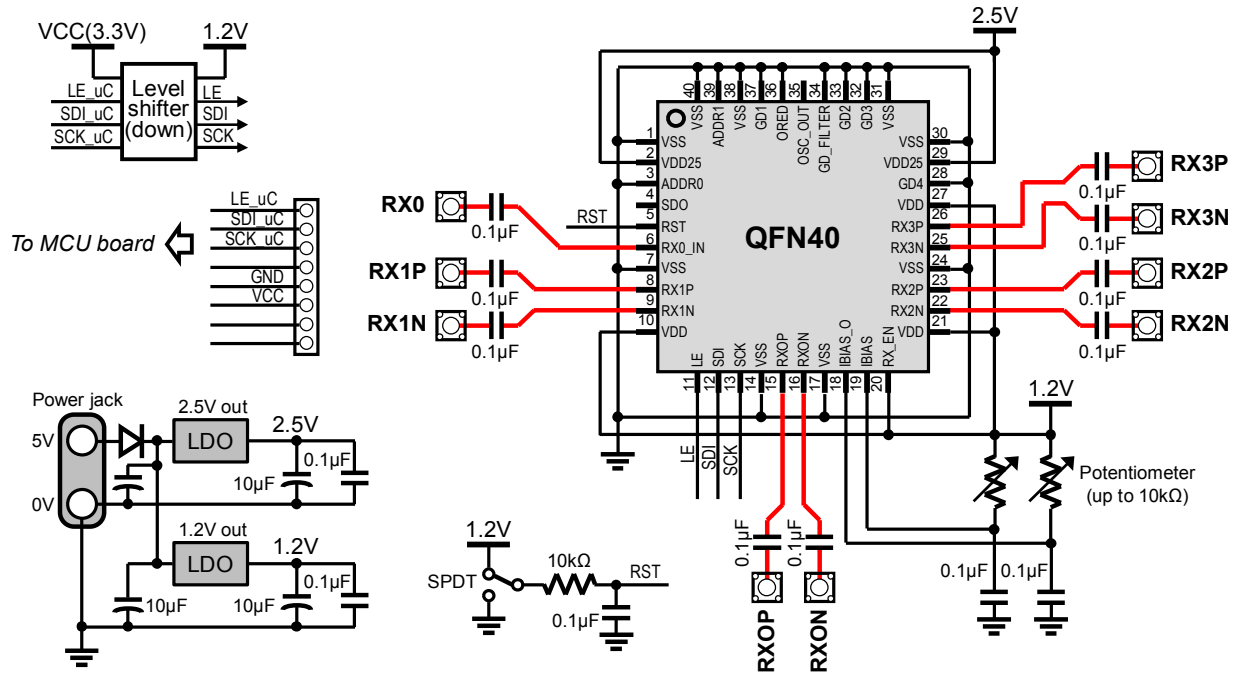


Figure 3.81. Test board schematic. Red lines are high-speed lines that require impedance control.

To support high-speed data transmission of 25 Gb/s, PCB dielectric material is selected as “Rogers” [69], not as conventional FR4. Rogers material has superior high-speed performance to FR4 because of lower dielectric loss (about 1/10 of FR4). Figure 3.82 shows the eye diagram simulation results; $2^{31}-1$ PRBS data are transmitted at 25 Gb/s on FR4 (Figure 3.82(a)), and Rogers (Figure 3.82(b)) on the same 1.5-inch PCB trace. Other components such as package parasitic, connectors and cables are also modeled in the simulation setup. Rogers shows about 23% larger eye-height opening than FR4 for $1E-12$ bit-error rate.

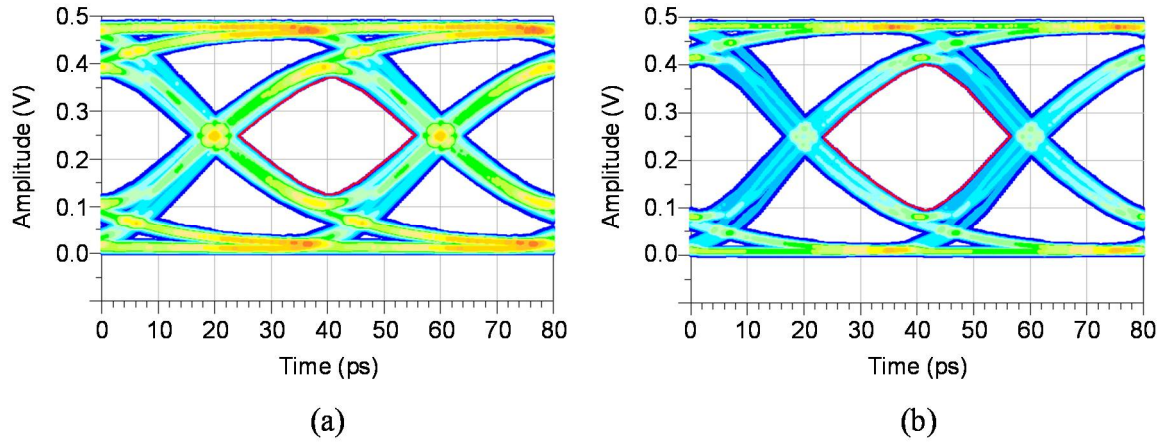


Figure 3.82. Eye diagram simulation results with different PCB materials: (a) FR4 and (b) Rogers.

Layout and the populated PCB are shown in Figure 3.83(a) and (b), respectively. It is laid out in a four-layer board: top and bottom copper layers for signal traces, and two inner layers for ground and power supply planes. To reduce signal loss on the traces, the board dimension is kept small: 2.4 inch \times 3.2 inch. The high-speed lines are designed as grounded coplanar waveguide (GCPW) with 28-mil width and 10-mil spacing in 1-oz copper (\sim 1.37-mil thickness); the dielectric thickness between the signal trace and ground plane will be controlled for 50- Ω characteristic impedance. The most important lines are the differential output traces (RXOP and RXON), which have trace length of 1.1 inch and are laid out symmetrically. To further improve signal integrity on the output lines, the insulating coats on the signal traces and adjacent ground copper for GCPW are stripped out using the solder mask layers. The length of each trace leading to the input differential pair is approximately 1.5 inches.

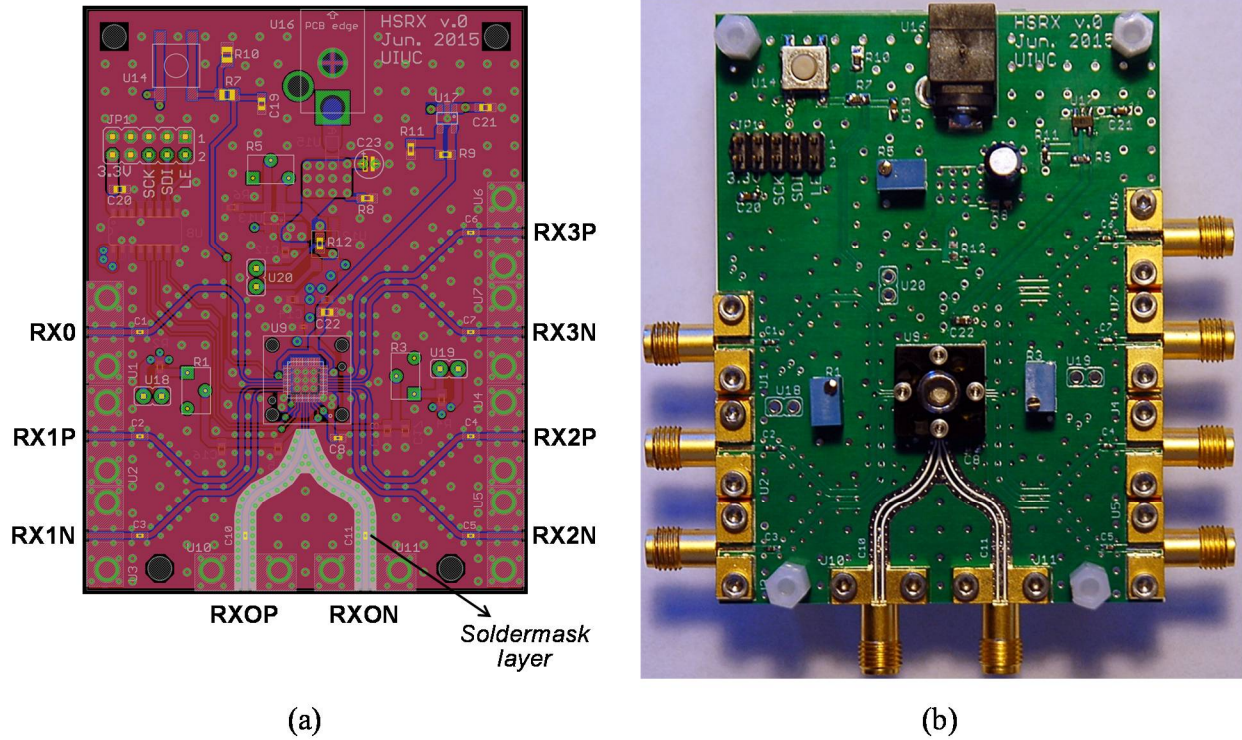


Figure 3.83. (a) PCB layout and (b) the assembled PCB with all connectors, parts and an IC socket. The board size is 2.4 inch \times 3.2 inch. The trace length of the differential outputs (RXOP and RXON) is 1.1 inch.

3.9 Measurement Results

Figure 3.84(a) shows the measured channel loss (S_{21}). The lossy channel is composed of a 4-ft flexible cable and a 6-inch semi-rigid cable, which are connected together by a female-to-female adapter, a 1.5-inch on-board trace, the QFN package leadframe, and a 1.6-mm bondwire. The S_{21} response gives approximately 15-dB loss at 12.5 GHz with several notches in the mid-band, due to impedance discontinuities at the connectors. Figure 3.84(b) shows the measured eye diagram at the output of the receiver HSRX1 with the equalizer function off ($EQ_OFF = 1$). The transmitter outputs 2^7-1 PRBS data at 25-Gb/s with 400-mV differential amplitude on the channel at Figure 3.84(a). Without equalization, the measured eye is totally closed.

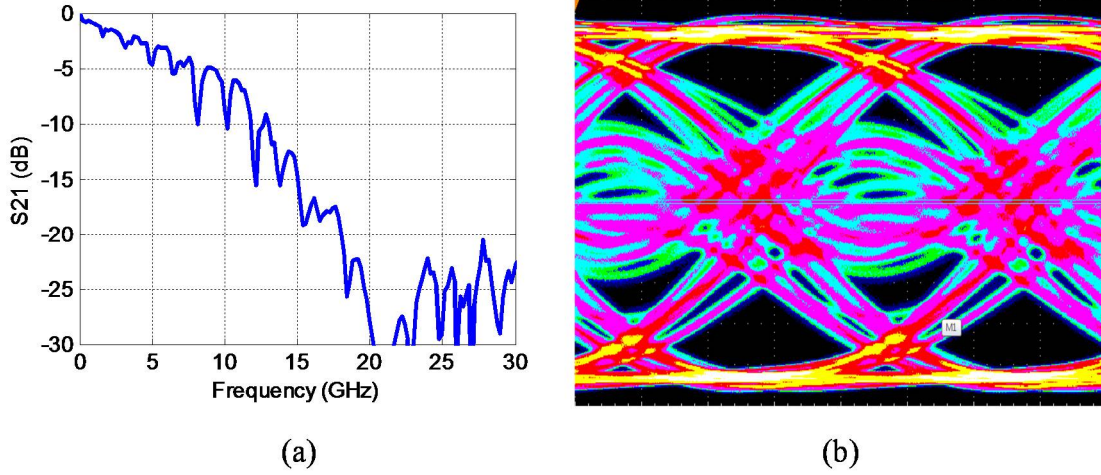


Figure 3.84. (a) Loss characteristic of the channel used for the eye diagram measurements and (b) the eye diagram of the HSRX1 output with the equalizer turned off (x-div: 10 ps/div, y-div: 100 mV/div). The TX data are 2^7-1 PRBS with 400 mV amplitude at 25-Gb/s.

The measured eye diagrams for HSRX1 and HSRX2 are presented in Figure 3.85(a) and (b), respectively, on the channel characterized in Figure 3.84(a). The eye of HSRX1 (Figure 3.85(a)) is opened by the aid of the equalizer, compared to the equalizer-off case of Figure 3.84(b), which verifies the equalizer functionality. The vertical eye opening of HSRX1 is 110 mV, while HSRX2 shows only 60-mV vertical eye opening; the superior eye diagram performance of HSRX1 was expected, based on the simulation results presented in Section 3.7.2.

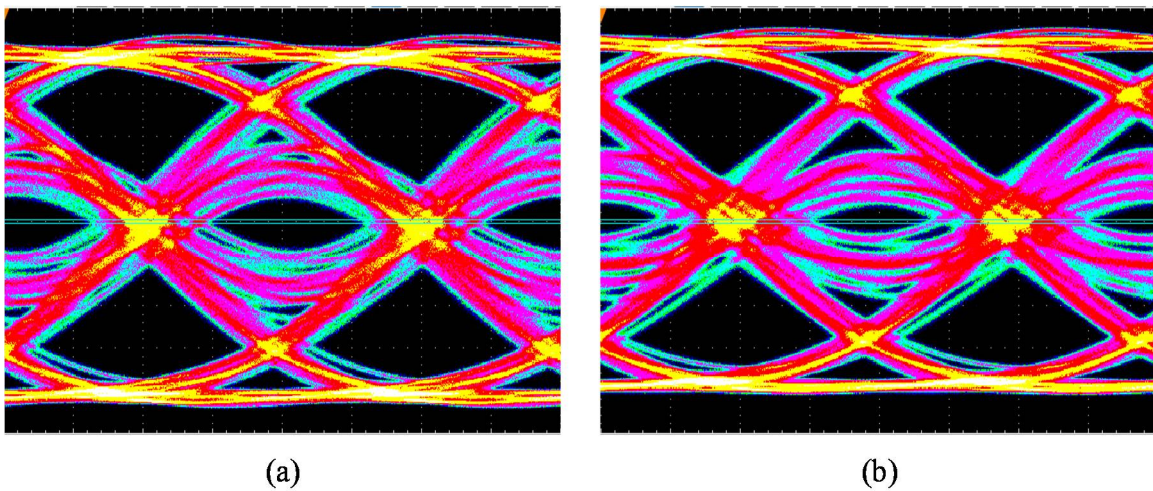


Figure 3.85. Eye diagram comparison: (a) HSRX1 and (b) HSRX2. The channel and data are the same as those in Figure 3.84 (x-div: 10 ps/div, y-div: 100 mV/div).

BER performance with different sampling points is plotted in Figure 3.86 for HSRX1 and HSRX2; these are often referred to as “bathtub” curves. For a BER of $1\text{E-}12$, HSRX1 shows 15.5-ps horizontal eye opening, which is 29% wider than that of HSRX2 (12 ps).

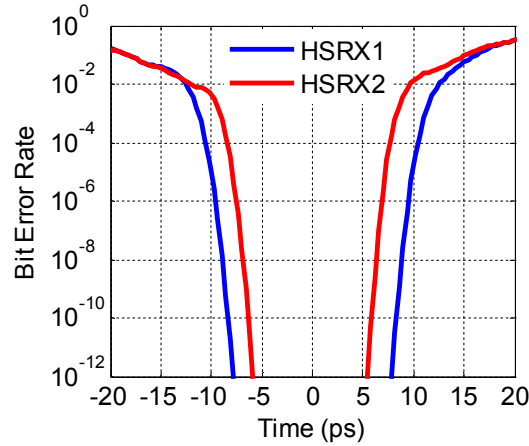


Figure 3.86. BER performance with different sampling time points for HSRX1 and HSRX2 from the eye diagrams in Figure 3.85.

VF-TLP has been performed on the bare dies with 200-ps rise time and 5-ns pulse width. Table 3.20 gives the measured failure currents for two bare dies and the averaged results. Surprisingly, HSRX1 shows slightly larger failure current than HSRX2, which is different from the simulation results in Section 3.7.1. Presumably, secondary ESD protection in HSRX2 does not protect well against a stress with a fast rise-time. This hypothesis can be verified by increasing the rise time of the VF-TLP pulse. If HSRX2 shows larger failure current than HSRX1, secondary protection works as intended; in other words, inductance halving will be more beneficial for the ESD events with a faster rise-time.

Table 3.20. Measured failure current under VF-TLP stresses: $t_r = 200\text{ps}$ and $t_w = 5\text{ ns}$.

	$I_{\text{fail,chip1}}$	$I_{\text{fail,chip2}}$	Average
Inductance halving (HSRX1)	8.23 A	8.22 A	8.23 A
Secondary protection (HSRX2)	7.53 A	7.65 A	7.59 A

3.10 Summary

Two variants of the T-coil bandwidth extension circuit are studied to select a better scheme for wireline I/Os. T-coil2, where transceiver circuits and termination resistors are located at the same port, will be a better choice because of its superior bandwidth performance. Magnetic coupling between two windings of the T-coil is identified as a potential CDM hazard; it will induce voltage overshoot at the receiver circuit side and potentially damage the gate oxide of the receiver transistors. A novel, inductance-halving T-coil circuit is introduced to reduce mutual inductance during ESD. To achieve the optimum T-coil for the ESD protection purpose, several layout techniques are proposed and proved by EM simulation. To compare the effectiveness of the proposed T-coil structures, 25-Gb/s wireline receivers with the CTLE are designed.

The inductance-halving scheme effectively reduces the voltage overshoot, as verified by simulation; it provides good voltage clamping at the active circuit input and is outperformed only by secondary ESD protection. However, the T-coil with inductance-halving provides 33% larger bandwidth and 42% wider eye height than does one with secondary protection at an even smaller layout area. From measurement, inductance-halving T-coil shows superior eye diagram performance. For the measured VF-TLP with 200-ps rise time, T-coil with inductance halving shows larger failure current than secondary protection, which potentially shows that inductance halving will be more advantageous for the faster rise-time ESD events.

The proposed inductance-halving T-coil improves bandwidth by $3.9\times$ relative to that at the input of a receiver without a T-coil. Despite the parasitics introduced by the inductance-halving circuitry, the T-coil2 topology with inductance halving provides a larger bandwidth enhancement than the theoretical limit of $2.8\times$ for the conventional T-coil1 topology (with no overshoot compensation). Even though the inductance-halving technique is not expected to provide quite as high CDM protection as does secondary ESD protection, it will be the only solution for very high-speed wireline or wireless applications such as Gigabit Ethernet or Ultra-wideband (UWB) RF transceivers, which are too performance-sensitive to use secondary protection.

As a new bandwidth extension technique, inductive neutralization has been proposed, which utilizes magnetic coupling between the input T-coil and the load inductor, to boost up current flow. It has the additional benefit of reducing layout area by overlapping both the load inductor and the main T-coil: about 24% area reduction in the 25-Gb/s CTLE in 65-nm CMOS. However, the composite inductor geometry was not fully optimized, thereby showing no better eye diagram performance than the conventional circuit.

Chapter 4. BER-Optimal ADC for ADC-Based Receivers

4.1 Motivation

To satisfy the demand for high-speed data communications such as high-speed internet or super-computing, the data-rate in serial I/O links now exceeds multi-Gb/s range. However, the channels that connect the transmitting and receiving ends have a non-uniform frequency response, due to the skin effect, impedance mismatch, intersymbol interference (ISI), and other causes. Figure 4.1 shows the non-uniform frequency response of a typical FR4 printed-circuit board (PCB) with different trace lengths [70].

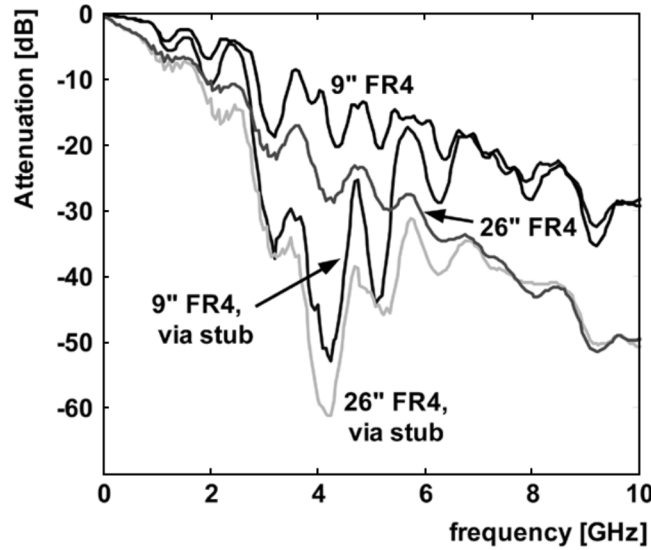


Figure 4.1. Frequency responses of different channel lengths of FR4 printed-circuit board [70].

Various equalization techniques are adopted to compensate for this channel non-uniformity. The fundamental operation of an equalizer is to boost high-frequency magnitude with a high-pass or band-pass filter. ADC-based equalizers have been widely investigated, as a possible equalization technique [71], [72], [73], especially for compensating high-loss and high-ISI channels. Figure 4.2 illustrates a typical block diagram of an ADC-based equalization system. The popularity of ADC-based equalizers is due to the adoption of versatile digital equalization algorithms by back-end digital signal processors. Moreover, digital signal processing provides high reliability and scalability with device scale-down.

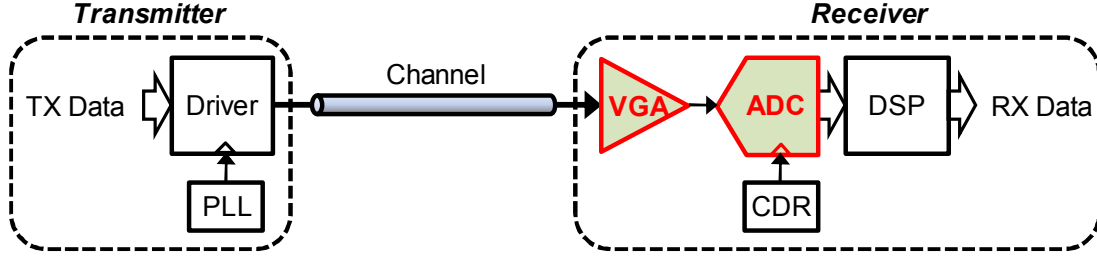


Figure 4.2. High-speed serial I/O link system with an ADC-based receiver.

Despite the ability of ADC-based receivers to compensate highly lossy channels, the application of ADC-based equalizers is still limited because of the high power consumption of multi-GS/s ADCs. In this dissertation, a new ADC design method is proposed, which adopts a new system performance metric, bit-error rate (BER) in case of wireline communication links, to mitigate ADC requirements for energy-efficient receivers. The test chip has been implemented and verified in measurement results.

4.2 BER-Optimal ADC

Typically, the required ADC resolution for high-speed links ranges from 4-6 bits considering the required signal-to-noise ratio (SNR) for back-end processing. However, ADC bit resolution can be mitigated by adopting BER as an ADC performance metric, as opposed to the SNR or SNDR used in conventional stand-alone ADCs [74], [75]. According to [74] and [75], ADC with non-uniform quantization levels optimized for BER does not need as much bit-resolution as ADC with uniform quantization levels. As shown in Figure 4.3, BER performance of 3-bit non-uniform-level ADC is better than that of 4-bit uniform-level ADC from 25-dB channel SNR. At the point of $1\text{E-}12$ BER, the channel SNR margin of 3-bit non-uniform-level ADC becomes as much as 4 dB, compared to 4-bit uniform-level ADC. Using 3-bit ADC instead of 4-bit can save power consumption of ADCs, which leads to overall system power-efficiency; this is the case for flash-type ADCs [76], whose power consumption is proportional to 2^N (N: the ADC resolution). This new type of ADC is “BER-optimal ADC,” which has been optimized for BER performance.

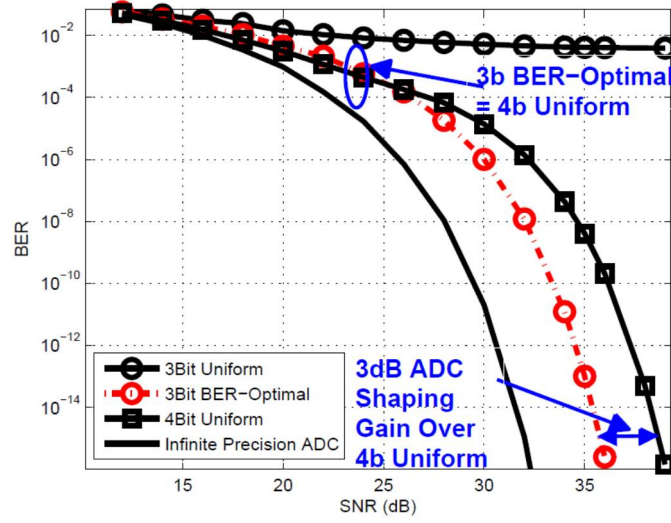


Figure 4.3. BER vs. SNR for 3-bit uniform ADC, 3-bit non-uniform ADC, and 4-bit uniform ADC [75].

The overall block diagram including BER-optimal ADC is shown in Figure 4.4. ADC output data are processed by the back-end digital, which is implemented in a separate FPGA board. An adaptive least-mean squared (LMS) equalizer, composed of F-block and Weight update, performs channel equalization. The quantization level of ADC is controlled by a Reference Level Update (RL-UD) block, which calculates reference levels based on equalizer coefficients ($\mathbf{w}[n]$) and error ($e[n]$). BER-optimal ADC is treated as one of the front-end components of the composite receiver, not a stand-alone block.

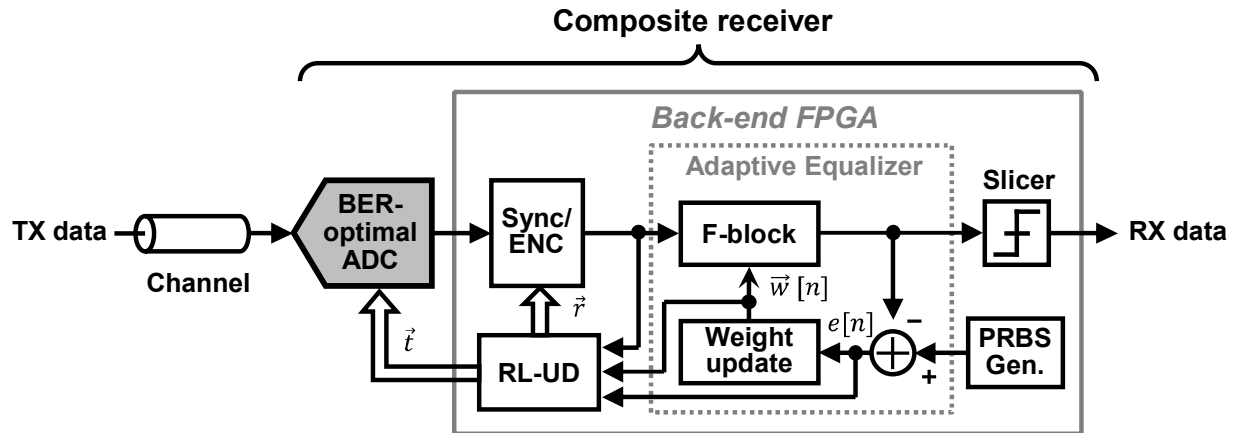


Figure 4.4. Block diagram of the overall BER-optimal ADC-based receiver system: ADC chip and back-end digital implemented in FPGA.

4.3 ADC Design

Figure 4.5 shows the overall ADC block diagram. Based on the system simulation results in Figure 4.3, 3-bit non-uniform level ADC has been implemented. In order to compare BER performance, conventional 4-bit uniform level ADC is also needed. Hence, this ADC was designed to be reconfigured as uniform or non-uniform, with 3-bit or 4-bit quantization levels, by using a digital-to-analog converter (DAC). The on-chip 2^7-1 PRBS generator is integrated, whose data will be multiplexed with ADC output for initial training of the back-end adaptive equalizer and output channel skew cancelation.

The ADC architecture used in this dissertation is a flash ADC [76], considering the high conversion rate of 4 GS/s. Flash ADC is known as the fastest architecture because it can convert analog input at once without any pipeline delay. Although some latch or flip-flop stages can be inserted for stable operation in recent high-speed flash ADCs, all of the digital codes are generated simultaneously. As shown in Figure 4.5, this consists of a comparator array followed by a thermometer-to-binary encoder, a reference generator by 8-bit DAC, and a clock distribution network. Usually quantization levels are set by a reference voltage generator implemented by a resistor string. In this dissertation, however, a DAC followed by a storage capacitor array has been designed for reference generation due to non-uniform quantization level control. The comparator array generates thermometer codes, so binary code encoding is needed to simplify back-end digital processing in binary format. The main drawback of flash ADC is its high power consumption and large area. These tradeoffs are incurred by maximizing parallelism to achieve the fastest possible conversion rate. Usually, the required number of comparators is $2^N - 1$ for an N-bit flash converter. Hence, the power and area consumption increase exponentially with the number of bits. As comparators consume most power and area in a flash converter, decreasing the number of bits can save power and area.

Comparators in flash ADCs provide an inherent sample-and-hold function. Thus a sample-and-hold amplifier (SHA) in front of the ADC is not necessary [76]. In the case of SHA elimination, power can be

saved but each comparator characteristic should be matched well and the clock distribution network for comparators should be carefully designed.

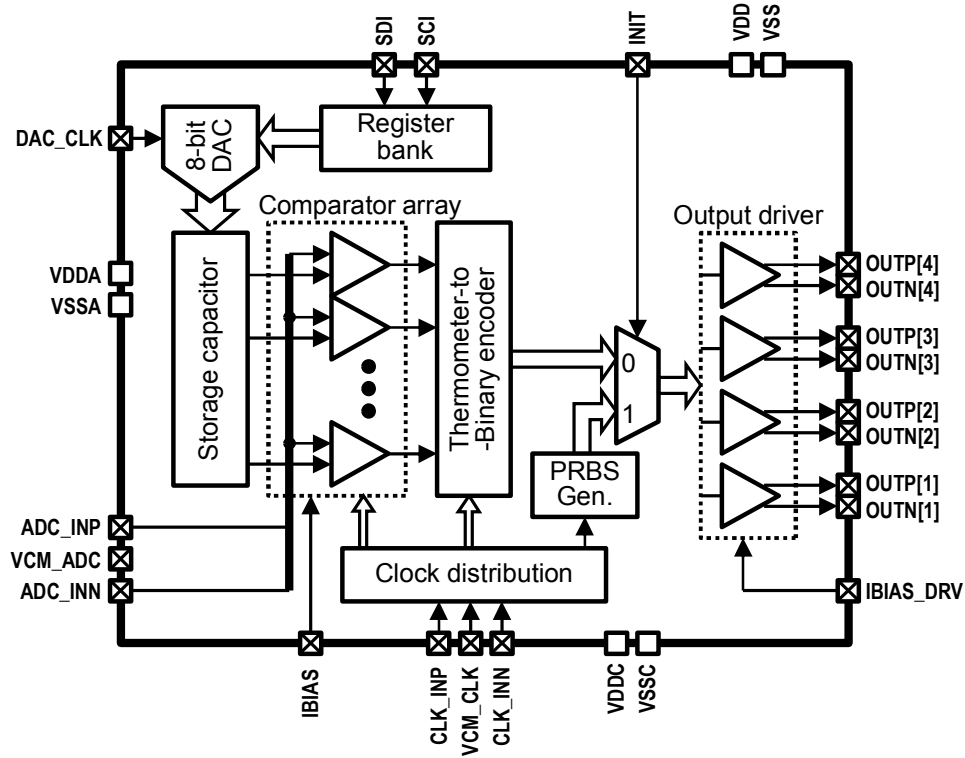


Figure 4.5. Block diagram of 3-bit/4-bit-reconfigurable 4-GS/s flash ADC.

4.3.1 Single-core multiple-output DAC

The DAC requires 8-bit resolution, considering system requirements and comparator offset calibration. DAC architecture can be (1) multiple-core multiple-output, or (2) single-core multiple-output. Single-core multiple-output DAC architecture reduces power consumption and area, and avoids mismatch from multiple DAC cores. Figure 4.6 illustrates a DAC schematic and a timing diagram. This two-step R-C-string DAC topology is modified from [77] for multiple outputs, whose output voltages will be stored in a switched capacitor array with 30 capacitor banks ($C_{REF,0} - C_{REF,29}$) for 15 pseudo-differential reference voltage pairs supporting 4-bit ADC resolution. The DAC operates in a sub-ranging manner; 8-bit resolution is divided by MSB 4-bit (M value) and LSB 4-bit (N value), which pre-select voltage points in a resistor string by S_{MSB1} and S_{MSB2} , and capacitor values of C_1 and C_2 by S_{LSB1} and S_{LSB2} . A two-phase

clock, ϕ_1 and ϕ_2 is used to consecutively access each storage capacitor and refresh reference voltage values ($V_{REF,0} - V_{REF,29}$). During the ϕ_1 phase, the charge stored at C_1 (Q_1) and C_2 (Q_2) are as follows:

$$Q_1 = C_1 V_1 = N C_u (V_{TOP} - V_{BOT}) \frac{M+1}{16} \quad (4.1)$$

$$Q_2 = C_2 V_2 = (16-N) C_u (V_{TOP} - V_{BOT}) \frac{M}{16}. \quad (4.2)$$

In the ϕ_2 phase, DAC output is stored as $V_{REF,i}$ at i -th storage capacitor ($C_{REF,i}$), which is expressed as follows by the law of charge conservation:

$$V_{REF,i} = \frac{C_u}{C_{REF,i}} \frac{16M+N}{16} (V_{TOP} - V_{BOT}). \quad (4.3)$$

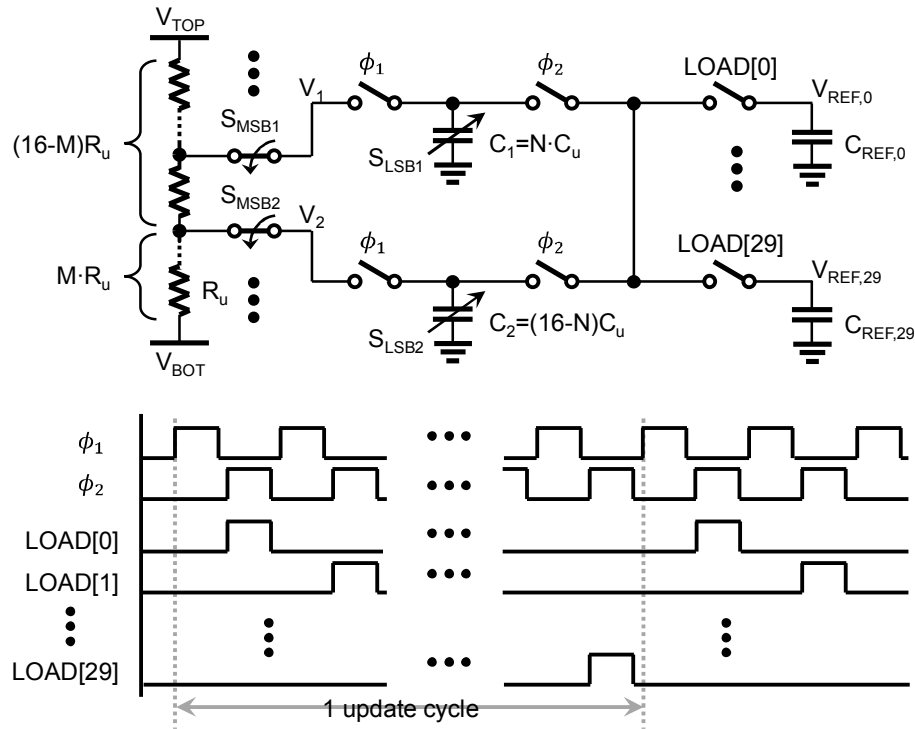


Figure 4.6. 8-bit single-core multiple-output DAC architecture and timing diagram.

4.3.2 Clock distribution network

As this chip does not include an SHA to reduce power consumption, the clock distribution network should be carefully designed to achieve good dynamic characteristics of the ADC. The clock specification

is determined by the comparator requirements for stable operation without metastability and no degradation of SNDR, as summarized in Table 4.1. Note that the RMS jitter does not need to be 2 ps in the case of a 3-bit ADC mode, but the jitter specification is set to the 4-bit requirement.

Table 4.1. Specifications of the clock distribution network.

Items	Min	Typ.	Max	Units
Frequency	-	-	4	GHz
Rise/fall time	-	-	30	ps
Duty cycle	45	50	55	%
Differential skew	-	-	10	ps
RMS jitter	-	-	2	ps

The overall clock distribution network is illustrated in Figure 4.7. The clock distribution network can be divided into two parts: clock delivery and clock tree. Because this chip is fabricated as a MOSIS multi-project wafer (MPW), the die size is fixed to $2\text{ mm} \times 2\text{ mm}$. Thus, there is a distance of about $500\text{ }\mu\text{m}$ from the clock input pads to the ADC core (comparator array and encoder). A “clock delivery” circuitry composed of an input buffer, repeater and differential-skew canceller has been designed to drive this long line length. To provide the comparator array and encoder with equal phase-delay clocks, a tree-style layout is needed.

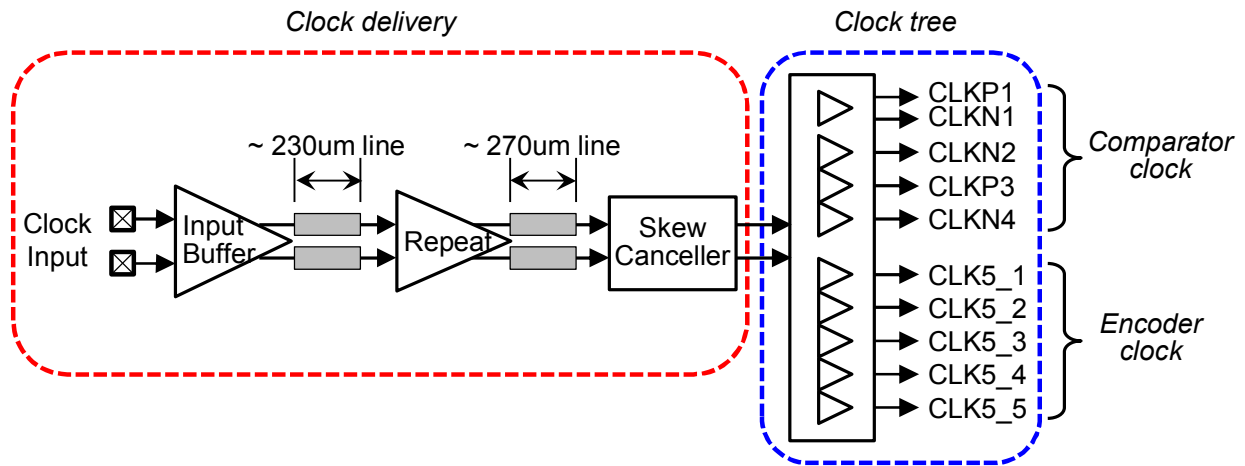


Figure 4.7. The clock distribution network.

Figure 4.8 shows the clock input buffer schematic. As a 4-GHz differential clock signal comes from the 50- Ω source over a 50- Ω transmission line, clock input lines are differentially-matched with differential 100 Ω for the maximum power transfer. The two 50- Ω termination resistors are implemented with p+ poly-silicon resistors on the chip. The common node of the two 50- Ω resistors (VCM_CLK) is connected to an external DC voltage regulator with 0.6 V because the clock lines will be AC-coupled through series capacitors on the board.

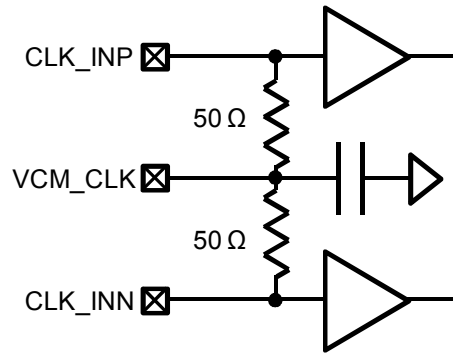


Figure 4.8. Clock input buffer. The common-mode input voltage (VCM_CLK) will be provided externally, to set the appropriate DC voltage of input buffer (external AC coupling capacitors on the clock input pair are not shown).

As differential clock signals travel through a long metal interconnection line, the differential (or complementary) skew can increase, which should be canceled out before distribution to the local sites. One of the straightforward implementations of the skew canceller is the simple cross-coupled inverter stage depicted in Figure 4.9. This scheme incorporates a feedback path to force skewed differential signals to be zero-crossed and works well for a small amount of skew at low frequency. However, if the skew becomes excessive (i.e., as much as 25% of unit-interval (UI)), this scheme fails to suppress skew under 10 ps. Furthermore, it degrades duty-cycle. Contention between feedback and feedforward paths can hinder full signal swing, which consequently leads to duty cycle distortion. The simulation result of this duty-cycle degradation is shown in Figure 4.10 at differential skew inputs of 20 ps (Figure 4.10 (a)) and 60 ps (Figure 4.10 (b)) for different process corner conditions. Clock names on the x-axis represent different clock points on the comparator stage and encoder, as shown in Figure 4.7. For 20 ps skew, the

duty-cycle is within the specification window (45~50%) for all corner conditions, though 60 ps skew shows several spec-out clock points.

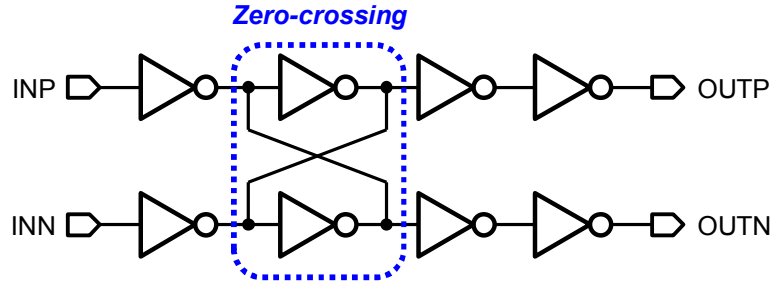


Figure 4.9. Conventional skew canceller incorporating zero-crossing inverter pair.

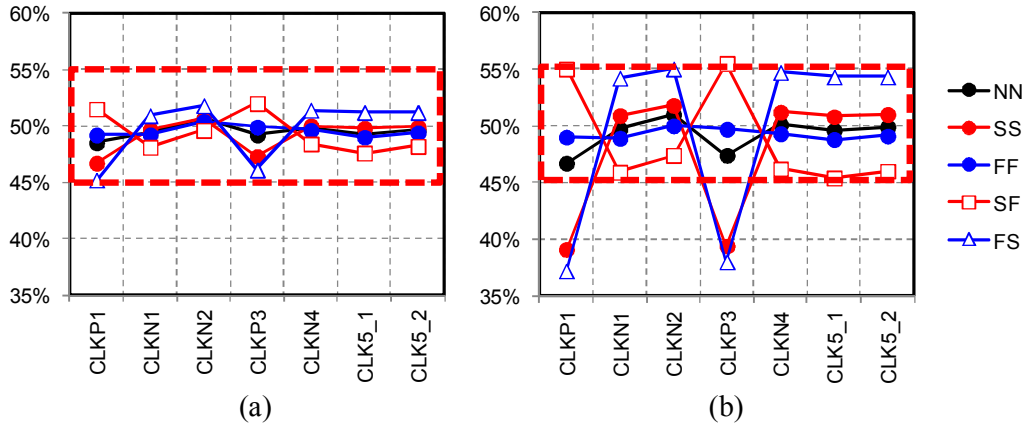


Figure 4.10. Duty cycle simulation results observed at different clock sites (x-axis) with different skew inputs under various process corner conditions: (a) 20 ps differential skew input, (b) 60 ps differential skew input.

To solve this problem, phase-blending with only feedforward path will be used together with zero-crossing [78]. Figure 4.11 demonstrates the proposed skew compensation circuit for this combined technique. First, phase-blending roughly compensates skew by using feedforward path. Then, the residual skew will be further canceled in the next zero-crossing stage by weak cross-coupled inverters. The simulation results for the same conditions as Figure 4.10 are shown in Figure 4.12. Compared with the previous results, the duty-cycle is well within its specification with enough margin even at 60 ps skew (nearly equivalent to 0.25 UI). The differential skew is suppressed under 5 ps for all process conditions. However, this improved performance comes at the cost of increased current consumption in the clock delivery part, from 7 mA to 8.2 mA.

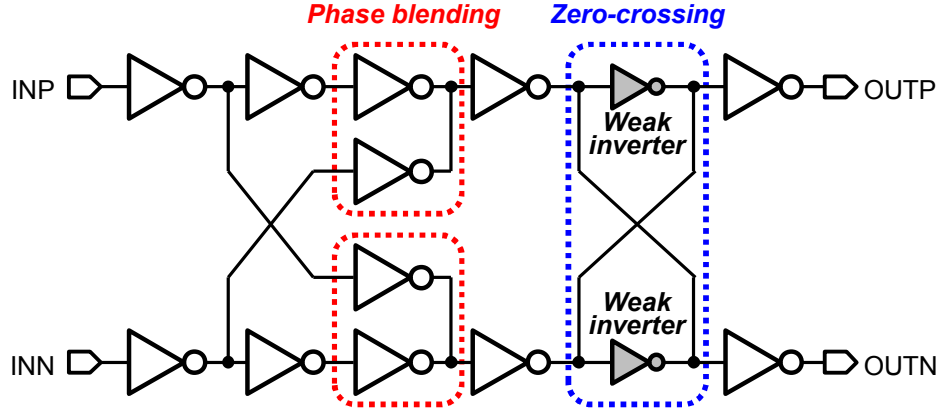


Figure 4.11. The proposed skew compensation circuit incorporating phase-blending and zero-crossing.

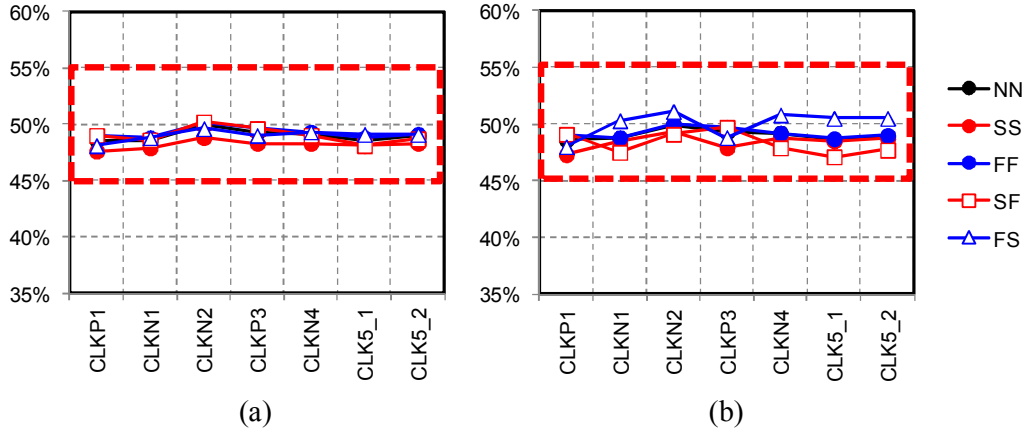


Figure 4.12. Duty cycle simulation results of proposed skew compensation circuit: (a) 20 ps differential skew input, (b) 60 ps differential skew input.

Once the differential skew is removed, the clock signal is ready to be distributed to all local clock points through the clock tree. Figure 4.13 is a perspective layout view of the vertical tree structure. Only the top three metal layers (metal6-8) are used because the lower metal layers should be used for internal block routing. To equalize parasitic capacitance, the metal lines are intentionally stretched to cover every sub-tree routing with the same metal patterns.

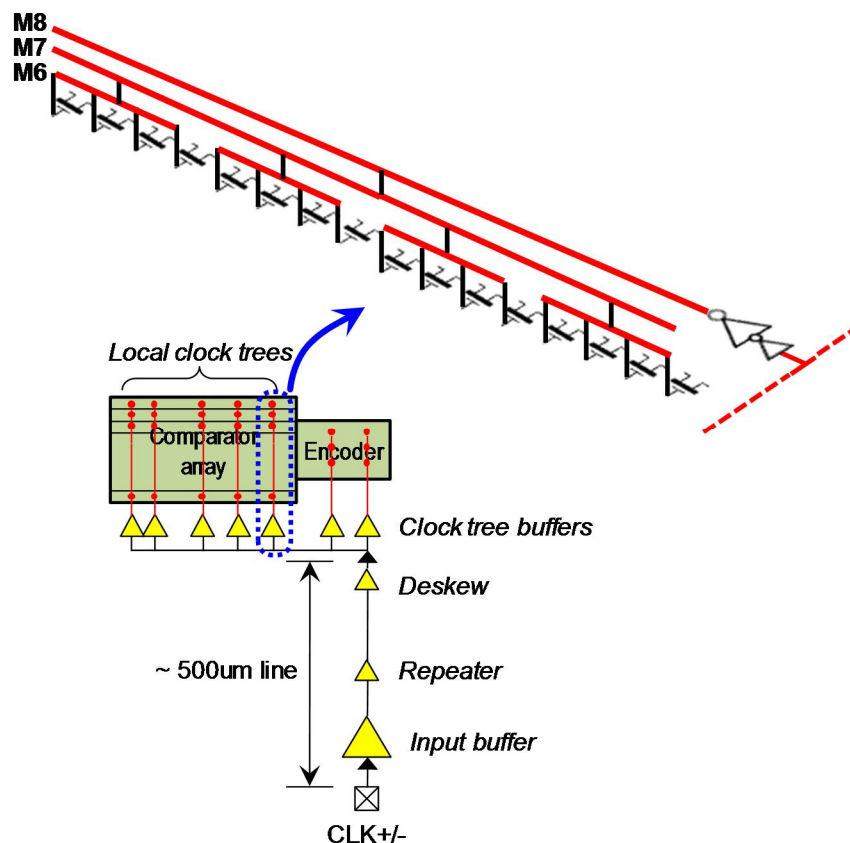


Figure 4.13. Clock tree layout for comparator input clock.

4.3.3 Output driver

The output driver provides 4-Gb/s data off the chip. Because the PCB trace and receiver of the back-end FPGA are all 50- Ω matched, the output of the driver should be matched to 50 Ω . The output driver specifications are summarized in Table 4.2.

Table 4.2. Specifications of the output driver.

Items	Min	Typ	Max	Units
Data-rate	-	-	4	Gb/s
Differential output swing	80	300	800	mV
Duty cycle	45	50	55	%
Differential skew	-	-	10	ps
Single-line impedance	43	50	57	Ω

Figure 4.14 illustrates the 50-Ω output driver schematic. It is a current-mode-logic (CML) type driver with a 50-Ω passive load for single-line impedance matching. At the source of two NMOS input transistors, a source degeneration resistor (R_1) is inserted to improve gain linearity and immunity to process variation. The voltage gain can be expressed as follows:

$$A_v = -\frac{g_{m1}R_2}{1 + g_{m1}R_1/2} \approx -\frac{2R_2}{R_1}. \quad (4.4)$$

Compared to the gain of the CML driver without source degeneration ($-g_{m1}R_2$), the voltage gain is decreased but is less sensitive to transconductance (g_m) of the transistor, which highly depends on process-variation. Especially for large g_m with $1/g_m \ll R_1$, the voltage gain can be approximated by a resistance ratio, as in (4.4).

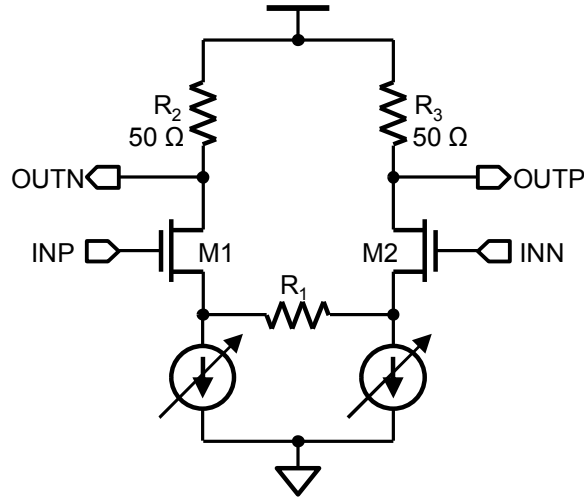


Figure 4.14. Schematic of a CML-based output driver.

4.4 Full-Chip Simulation

Combining all of the functional blocks in the ADC, a full-chip simulation has been performed as shown in Figure 4.15. At the beginning, on-chip PRBS generator data are outputted. Then, by switching 2:1 MUX selection, ADC data for a 100 MHz sinusoid come out. Binary 4-bit ADC outputs are post-processed by an ideal DAC block in Spectre to reconstruct them as a digitized analog waveform.

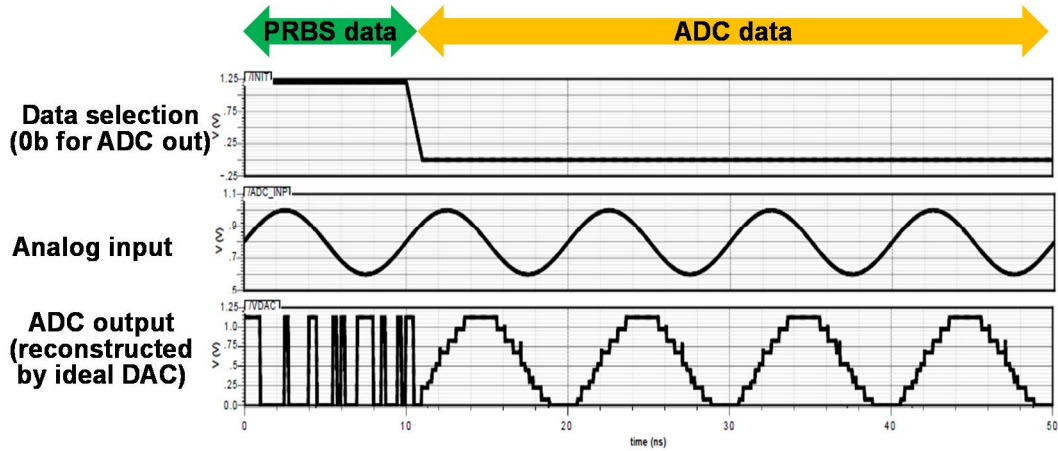


Figure 4.15. Waveforms from a full-chip transient simulation.

4.5 Full-Chip Floorplan and Layout

The test chip has been designed in a 90-nm low-power CMOS process with eight metal layers. The full-chip layout is demonstrated in Figure 4.16. The full die size is $2\text{ mm} \times 2\text{ mm}$ and the ADC core layout size is 0.38 mm^2 . The storage capacitor arrays of DAC are placed between analog input pads and comparator arrays. The vacant areas are filled with de-coupling capacitors designed by MOS capacitor for on-site power noise suppression. To suppress power and ground noise propagation, power rails are separated into three different domains: analog, digital, and clock. The layouts of sub-blocks including related pad cells for each power domain are also completely separated by triple-wells to prevent noise injection through the substrate. Ground pads are assigned especially adjacent to the high-speed signal I/Os to minimize return path. All signal and power pads have proper ESD protection diodes. The chip microphotograph is shown in Figure 4.17.

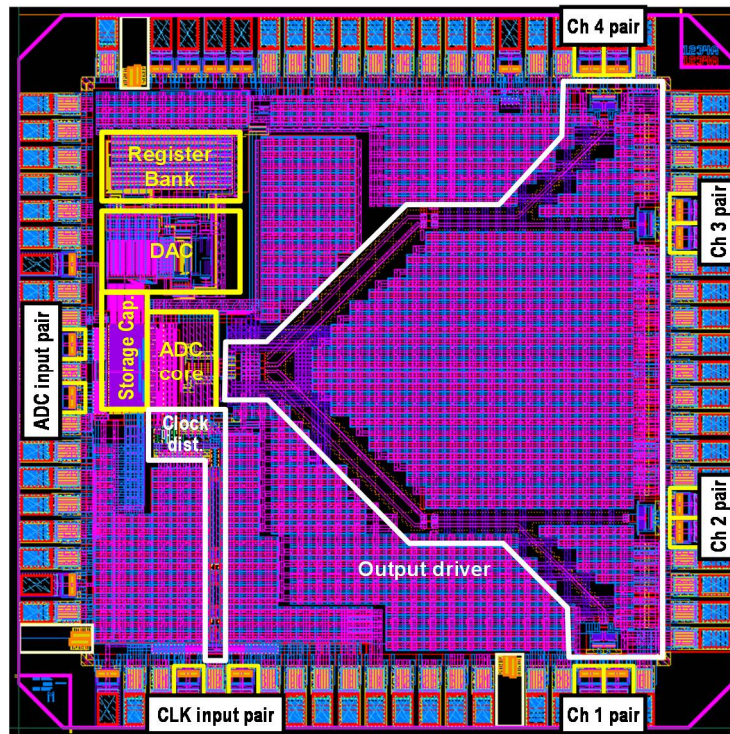


Figure 4.16. Full-chip layout.

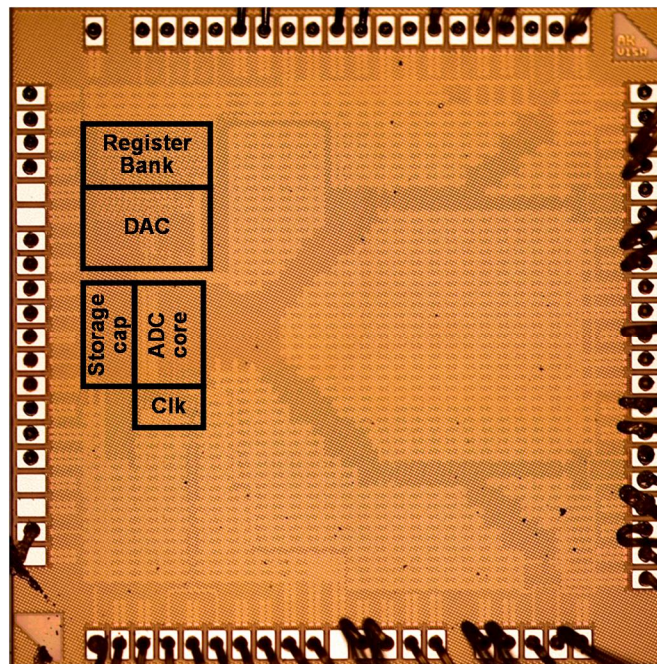


Figure 4.17. Chip microphotograph.

4.6 Test Board Design

This chip was assembled using a 40-pin open-cavity QFN package, which can support 4-GS/s ADC operation. This package provides a huge ground plane on the die-attach area and backside of the package for a heat sink and a solid ground connection. Thus, all ground pads are down-bonded to this ground plane, which is good for noise performance. By saving ground pins on the package, smaller packages with a smaller pin count can be used, which is beneficial to reduce bondwire length and parasitic inductance.

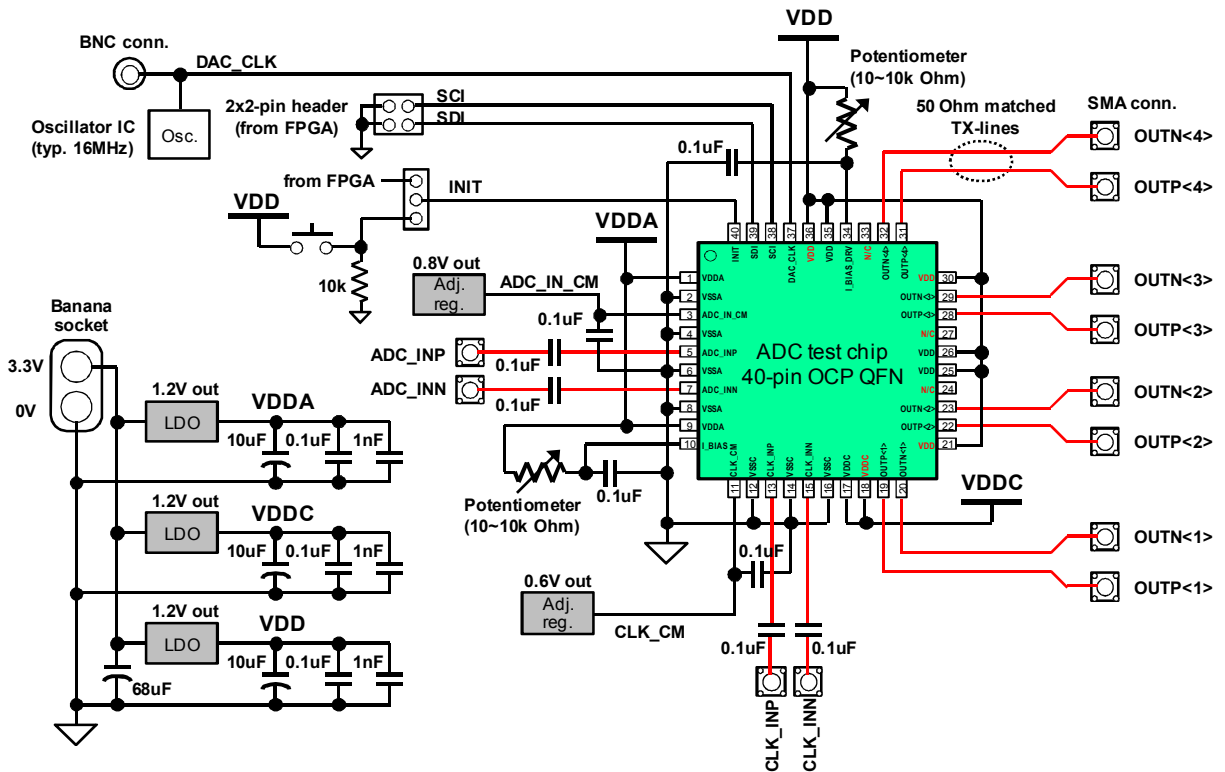


Figure 4.18. Test board circuit.

Figure 4.18 shows the test circuit for PCB. This board has only one 3.3 V external power supply, and each domain voltage is generated by on-board voltage regulators. The differential ADC and clock input pairs are AC-coupled using on-board capacitors. Four ADC output differential pairs are also AC-coupled using integrated AC-coupling capacitors at the back-end FPGA. Potentiometers are used to adjust the amount of bias current and to set the appropriate common-mode voltages for AC-coupled differential inputs.

This board is manufactured from 4-layer FR4 PCB as shown in the layout and photograph in Figure 4.19. The PCB dimension is 3.5 inch \times 5.0 inch. High-speed differential traces are designed with a coplanar waveguide (CPW) with a vertical ground plane for 50- Ω characteristic impedance. The transmission line width has been designed at 20 mils which is the same width as 0402-size surface-mount devices, to minimize potential impedance discontinuity by footprint patterns. The input differential traces (ADC analog input and clock) are exactly symmetric; they are matched in line length and shape to eliminate complementary skew. The ADC chip is mounted without soldering on a high-speed socket, which has the elastomer with tiny gold-plated brass balls.

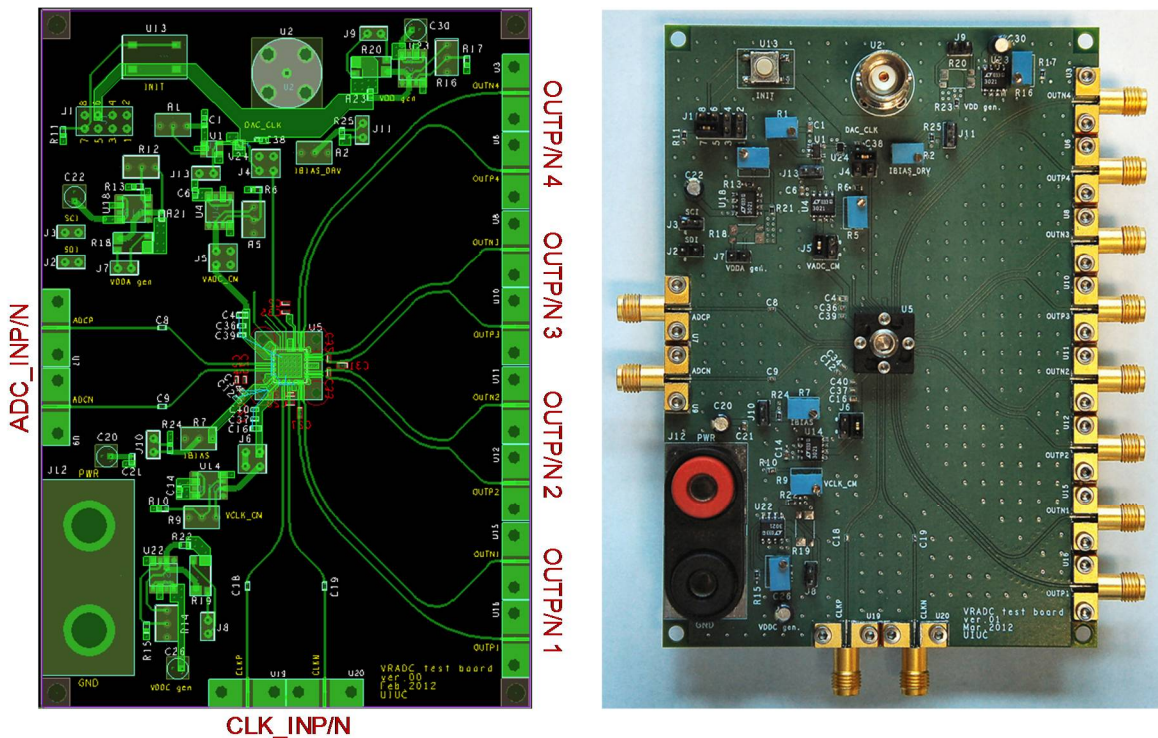


Figure 4.19. Test board layout (left) and an actual photo of the test board with all sockets, connectors and components assembled (right).

4.7 Measurement Results

The chip test was performed in two phases: a stand-alone ADC test and a link test with an FPGA.

4.7.1 Stand-alone ADC test

The ADC performance as a stand-alone block was measured in a 4-bit uniform level mode. Figure 4.20 shows linearity characteristics: differential nonlinearity (DNL), and integral nonlinearity (INL). ADC calibration is performed manually by tuning the DAC code from each nominal value to compensate input DC offsets of comparators. Before comparator offset calibration, DNL was as high as ± 1.3 LSB as shown in Figure 4.20(a). After calibration, the DNL was improved to ± 0.04 LSB. In the case of INL, 4 LSB before calibration can be improved to 0.14 LSB. These values are sufficient for 4-bit ADC operation.

The full-scale range (FSR) of the ADC was set to $800 \text{ mV}_{\text{ppd}}$ by default, and can be configured to any value. From the default FSR, the effective number of bits (ENOB) is measured as shown in Figure 4.21 in 4-bit uniform ADC mode. ENOB of 3.4 bits was measured at the Nyquist frequency ($\sim 2 \text{ GHz}$).

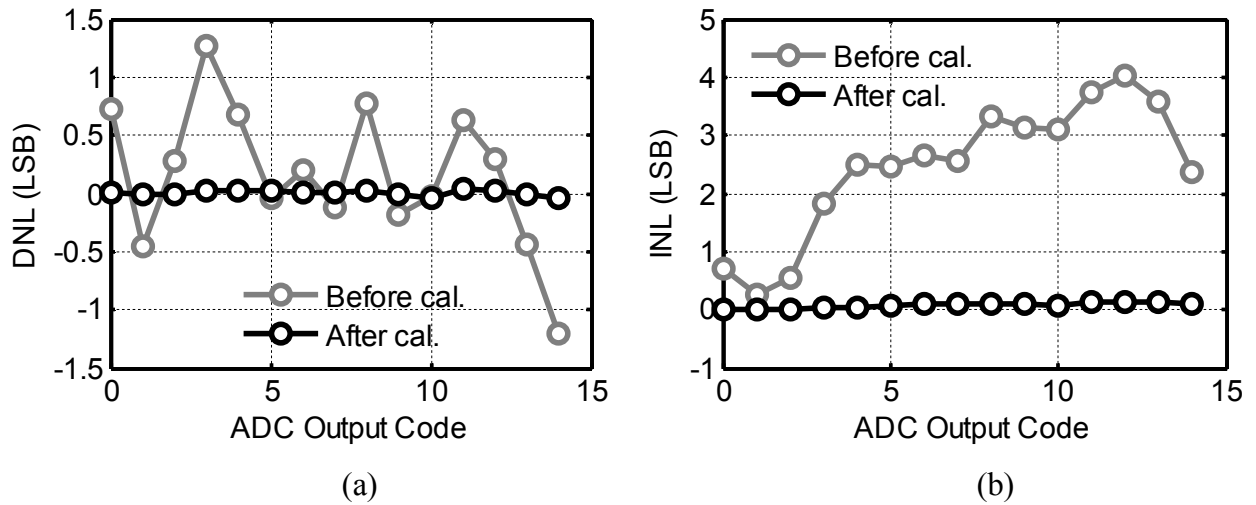


Figure 4.20. ADC linearity performance: (a) DNL and (b) INL, before and after comparator calibration.

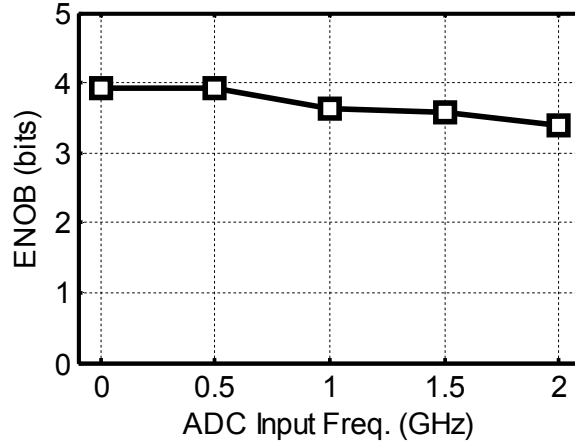


Figure 4.21. The effective number of bits (ENOB) at $\text{FSR} = 800 \text{ mV}_{\text{ppd}}$ in a 4-bit uniform ADC mode.

4.7.2 Link test

The test setup for the full link test is illustrated in Figure 4.22. The bit-error-rate tester (BERT) generates 4-Gb/s $2^{23} - 1$ PRBS data and 4-GHz clock. PRBS data go through an FR4 channel board and enters the ADC inputs. The phase of differential clock signals is adjusted to sample ADC input data at a proper time point. ADC outputs are fed into an FPGA board which contains all back-end processing as shown in Figure 4.4. BER is measured by the BER counter implemented in the FPGA.

Because this ADC chip does not contain a front-end variable-gain amplifier (VGA), ADC input voltage can be adjusted by BERT, and equivalently by FSR of ADC using DAC programming. In the link test, the FSR of the ADC was set to $100 \text{ mV}_{\text{ppd}}$. The ENOB characteristics of the ADC in different resolution modes was measured in Figure 4.23 at $\text{FSR} = 100 \text{ mV}_{\text{ppd}}$; 4-bit uniform ADC shows the best performance while 3-bit non-uniform ADC gives the worst performance even when compared to 3-bit uniform ADC. This was already well expected because 3-bit non-uniform ADC cannot reproduce an input signal with a good linearity due to non-uniform quantization levels. Because of the reduced FSR ($100 \text{ mV}_{\text{ppd}}$) and smaller 1-LSB voltage, ENOB in 4-bit uniform ADC is degraded by more than 1-bit, compared to that of Figure 4.21 with $\text{FSR} = 800 \text{ mV}_{\text{ppd}}$. Because there is no SHA in front of the comparator array, mismatches in input signal routing and bandwidth of the comparator array can result in

different aperture uncertainty [79] for each comparator channel, which can deteriorate ENOB performance.

Figure 4.24 shows BER performance with respect to the relative sampling phase from the center of the data window in unit-interval (UI) units. This measurement is done with a $220\text{ mV}_{\text{ppd}}$ transmit signal amplitude on a 20-inch FR4 channel board. As expected, 3-bit non-uniform ADC shows the best BER performance, $1\text{E-}12$, compared to 4-bit non-uniform ADC with 0.2 UI data window opening. This proves the concept that BER-optimal ADC can possibly reduce the resolution of ADC by non-uniform quantization level control toward a better system performance metric (BER, in this case).

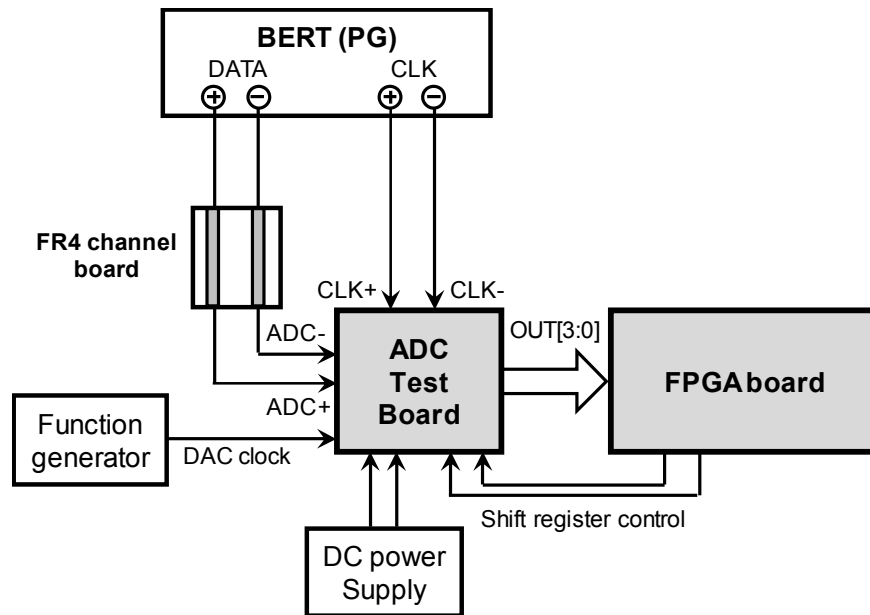


Figure 4.22. Link test setup using back-end FPGA board.

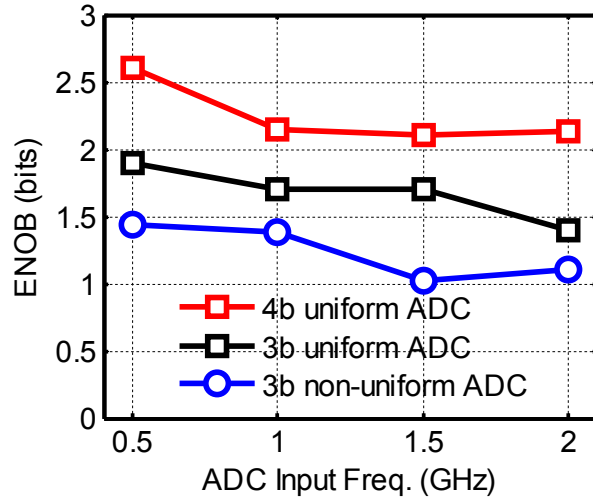


Figure 4.23. ENOB comparison in different ADC resolution modes at FSR = 100 mV_{ppd}.

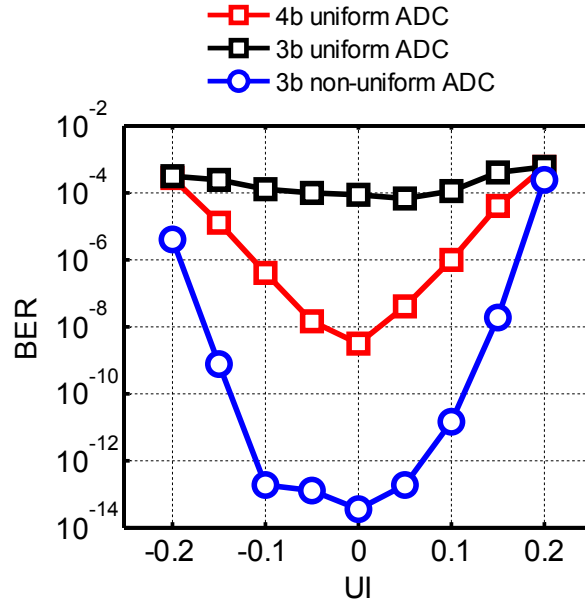


Figure 4.24. BER performance vs. relative sampling phase. Measurement conditions include a transmitter amplitude of 220 mV_{ppd} at 4-Gb/s $2^{23}-1$ PRBS on 20-inch FR4 board and ADC FSR = 100 mV_{ppd}.

The ADC performance is summarized in Table 4.3. As the back-end digital block was implemented on an FPGA board, its power consumption was estimated from RTL synthesis results in the same 90-nm CMOS process. The ADC in 3-bit non-uniform mode can save ADC power consumption almost by half, compared to regular 4-bit uniform ADC. In order to support non-uniform reference level control, more digital control blocks are required, but the additional power consumption by the back-end digital is much

smaller than the amount of reduced power in ADC. Therefore, the energy efficiency of the overall system can be improved.

Table 4.4 compares the performance of the proposed link to the previous works of ADC-based receivers. The receiver energy efficiency is calculated to 21.5 pJ/bit from the overall power consumption of 86 mW, which is comparable to or smaller than the other receivers.

Table 4.3. Performance summary.

ADC operating mode		3b non-uniform (BER-optimal)	4b uniform
Technology		90nm LP CMOS (1P8M)	
Core die area		0.38 mm ²	
Supply voltage		1.2V for analog, 1.28V for digital & clock	
Data rate		4 Gb/s	
Power consumption	ADC	56 mW	109 mW
	B/E digital	*30 mW	*28 mW

Table 4.4. Performance comparison of ADC-based receivers.

	This work	VLSI'11 [80]	ISSCC'09 [81]	ISSCC'11 [82]
Process	90nm	65nm	65nm	65nm
Data Rate [Gb/s]	4	10	10.3	5
Number of bits	3	3	6	5
BER @channel	< 1E-12 @20"	< 1E-7 @-17dB	< 1E-15 @-26dB	< 1E-12 @34"
RX power [mW]	86	130	500	192
RX efficiency [pJ/bit]	21.5	13.0	48.5	38.4

4.8 Summary

The concept of a BER-optimal ADC targeting system performance (BER) was proposed and demonstrated with measurement results. The 4-GS/s flash ADC with programmable 3-bit/4-bit

quantization levels was implemented in a 90-nm CMOS process to compare performance both in stand-alone ADC metrics and system performance metrics. The proposed ADC-based receiver can achieve as low as $1\text{E-}12$ BER with a 0.2-UI eye opening in 21.5 pJ/bit power-efficiency. The ordinary performance metrics from stand-alone ADCs cannot be applied to this ADC. However, in terms of system performance metrics, the proposed ADC can outperform normal uniform-level ADCs with higher resolutions. This feature can lead to better system energy-efficiency, which enables practical ADC-based receiver systems.

Chapter 5. Conclusion and Future Work

5.1 Conclusion

Very high-speed I/O design (> 10 Gb/s) often compromises ESD reliability, because of the trade-off relationship between the I/O speed and the allowable ESD-device parasitics. The research in this dissertation starts from the question on how to break this strict trade-off relationship to achieve both requirements.

From the first research in Chapter 2 on the self-protecting I/O using “adaptive active bias conditioning (AABC),” SSTL I/Os fabricated in 130-nm CMOS verified that about 30% CDM reliability was achieved with only 10% additional silicon area overhead. A polarity-aware ESD protection scheme using the ESD direction detector enables ESD protection in any polarity. Even with this approach, however, CDM could still be a problem in a very high-speed serial links such as over 10 Gb/s in nano-scale process technologies.

This issue was addressed again in Chapter 3 in a different approach; ESD device size is increased sufficiently to protect I/O circuits, and the increased capacitance is nullified by a bandwidth extension technique, a T-coil. Theoretically, the version of the T-coil in which both the transceivers and termination resistors are at the same port has wider bandwidth. A different T-coil scheme, which is preferable for wireline I/Os, was selected based on the performance metrics of bandwidth and impedance matching. The potential CDM failure mechanism of the T-coil scheme was identified as transient voltage overshoot through magnetic coupling. A novel peak voltage mitigation scheme, “inductance halving” has been proposed. The design and layout strategy for the inductance-halving T-coil was explored, and the most optimal T-coil layout was demonstrated. The test chip has been designed to verify the efficiency of the proposed T-coil scheme in the 25-Gb/s wireline receivers. From simulations in 65-nm CMOS PDK, the efficiency of the inductance halving scheme was verified in overshoot suppression by 2 V for 5 A VF-TLP with 100-ps rise time, compared to the uncompensated T-coil. However, the inductance-halving T-

coil cannot exceed the secondary ESD protection in peak voltage suppression. Regarding circuit performance, bandwidth is extended by 33% at the receiver input, and eye height is improved by 42% at the CTLE output, compared to the receiver with the T-coil and secondary ESD protection.

The third research in Chapter 4 focuses on energy-efficiency in the ADC-based receivers. The ADC specifications have been redefined to achieve a better system performance. A new “BER-optimal ADC” concept was proposed, to reduce the required number of bits by controlling the quantization level non-uniformly, according to the channel environment. The 3/4-bit-reconfigurable 4-GS/s flash ADC was implemented in 90-nm CMOS to verify the concept of the BER-optimal ADC. From the measurement results, it is verified that the 3-bit BER-optimal ADC which has only 1.5-bit ENOB can outperform 4-bit uniform-quantization-level ADC, with 0.2-UI eye opening for $1\text{E-}12$ BER at the 20-inch FR4 channel.

5.2 Future Work

The transient voltage overshoot reduction by the inductance-halving scheme, which has been proposed in Chapter 3, can be further improved with the scheme in Figure 5.1. The basic idea is to mimic the operation of the secondary ESD protection: to flow some branch current through the parasitic diodes. Different from the original scheme in Figure 3.18(a), the N-well bias of M1 is connected to VDD, as shown in Figure 5.1(a). During the PS-mode ESD, the voltage at node B can be higher than VDD, and the parasitic diode (D2) between p+ source junction of M1 and N-well tap (refer to the cross-section of M1 in Figure 5.1(b)) is forward-biased. Extra current will flow through D2, which will further reduce the voltage at P2 (receiver input), as the secondary ESD protection does. To utilize this “implicit” secondary protection with the stray diode (D2), the failure current of D2 should be sufficiently larger than the branch ESD current.

Another scheme with a better bandwidth performance for overshoot-canceling T-coils may be achieved by combining both direct switching (inductance halving) and indirect switching (shadow inductor). Shadow inductor loops can decrease mutual inductance partially, which can mitigate switch size requirements for inductance halving.

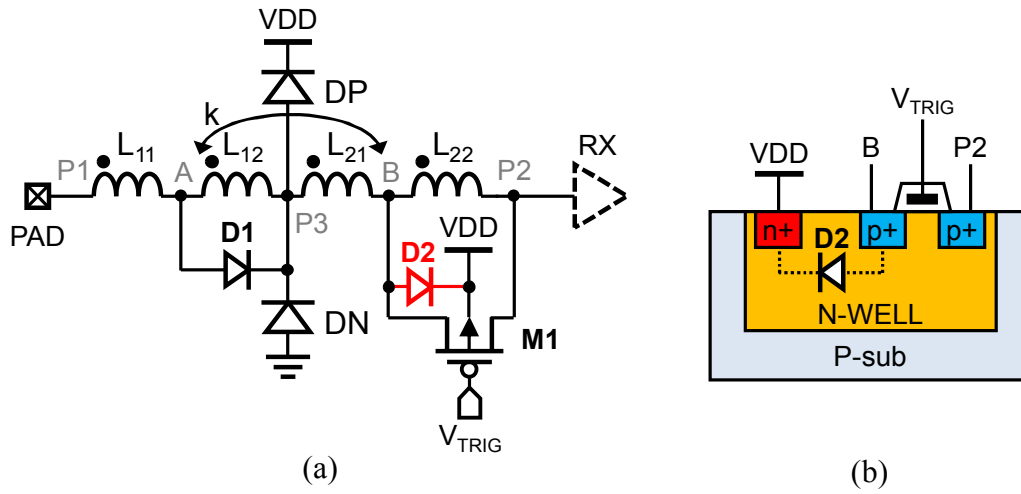


Figure 5.1. Improved inductance-halving scheme for a better overshoot voltage reduction: (a) schematic and (b) cross-section of the M1 transistor with a parasitic diode, D2.

Appendix A. Power-On I-V Curve Method

To accurately diagnose failure points after ESD, an I-V measurement performed on a power-on system is proposed. Using the proposed method, shunt resistance, R_T and R_B of ESD-induced failure model in Figure 2.11, can be calculated. The measurement setup is illustrated in Figure A.1. DC pin voltage sweep for PAD relative to VSS is done in the setup of Figure A.1(a). The semiconductor parameter analyzer (Agilent 4156C in this measurement) is used to generate DC voltage and to measure current through a coaxial SMU (Source/ Monitor Unit) channel. The chassis ground of the semiconductor parameter analyzer is connected to VSS pin of the device-under-test (DUT). To avoid voltage contention between chassis ground and DUT reference ground, the DUT board is powered by a mobile battery. In a similar way, the PAD-to-VDD I-V curve can also be measured as shown in Figure A.1(b); chassis ground is shorted to VDD pin of the DUT. When PAD voltage is swept from a negative voltage to a positive voltage (e.g., -1 V to 1 V in this design), I-V curves in Figure A.2 can be obtained.

In the PAD-to-VSS setup before ESD failure, an I-V curve only for the bottom diode is obtained, because the top diode is always in reverse-bias, if the voltage sweep range is less than the reverse breakdown voltage of diodes. The dashed line in Figure A.2(a) shows the unstressed I-V curve characterized from the bottom diode. However, when there exist R_T and R_B from ESD-damaged devices, straight lines from R_T (blue line) and R_B (red line) are added on the original diode curve; the composite I-V curve is distorted as shown in the thick black line in Figure A.2(a). The straight lines from DC shunt resistance, R_T and R_B , are expressed as follows:

$$I_{R_B} = \frac{1}{R_B} V \quad (\text{A.1})$$

$$I_{R_T} = \frac{1}{R_T} (V - VDD). \quad (\text{A.2})$$

Hence, the linear region of the composite curve can be expressed as follows:

$$I = \left(\frac{1}{R_T} + \frac{1}{R_B} \right) V - \frac{VDD}{R_T}, \quad \text{for } V \geq -V_{bot,on}, \quad (\text{A.3})$$

where $V_{bot,on}$ represents bottom diode turn-on voltage. From the measured I-V curve after the ESD stress, x-intercept (V_0) and slope ($1/R_0$) in the linear region of the curve can be recorded. From (A.3) with the measured parameters R_0 and V_0 , the shunt resistance, R_T and R_B can be calculated as follows:

$$R_T = \frac{VDD}{V_0} R_0 \quad (\text{A.4})$$

$$R_B = \frac{VDD}{VDD - V_0} R_0. \quad (\text{A.5})$$

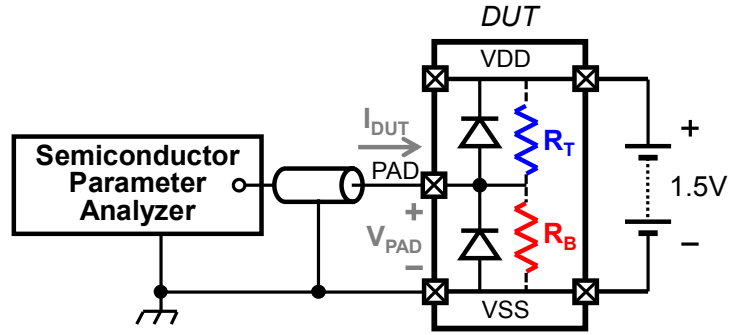
From Figure A.2(b) for the PAD-to-VDD I-V curve, the top diode characteristic is distorted by the failed devices with R_T and R_B , which is calculated as follows in the same procedure as the PAD-to-VSS case:

$$R_T = \frac{VDD}{VDD + V_0} R_0 \quad (\text{A.6})$$

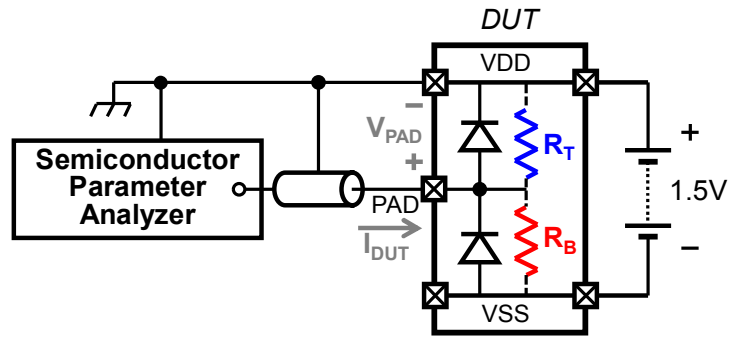
$$R_B = -\frac{VDD}{V_0} R_0. \quad (\text{A.7})$$

Using (A.4)-(A.7), shunt resistors from the actual power-on I-V curves after ESD failure were calculated as shown in Table 2.4. and Table 2.5.. Theoretically, both PAD-to-VSS and PAD-to-VDD I-V curves should give identical results. Thus, only one test setup is sufficient to obtain both shunt resistance.

The advantages of the proposed power-on I-V curve method are: (1) to ensure no unnecessary current flow during leakage test by configuring a DUT chip, and (2) to calculate both top and bottom shunt resistance with only one test configuration, compared to the conventional power-off leakage measurement which needs two separate test setups.



(a)



(b)

Figure A.1. Power-on I-V curve measurement setup. Chassis ground connection of the semiconductor parameter analyzer differentiates the reference point of DC sweep: (a) DC sweep from PAD to VSS, and (b) DC sweep from PAD to VDD. In order to avoid voltage contention, DUT board should be battery-powered.

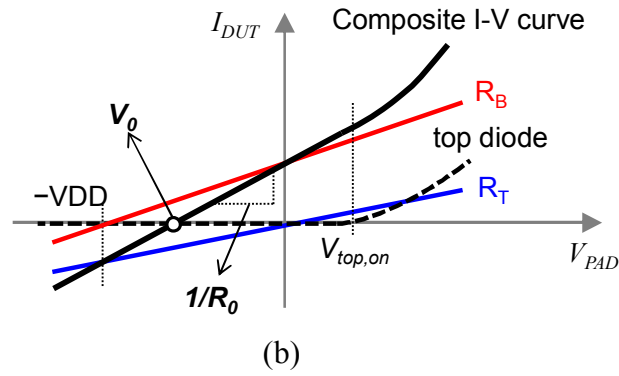
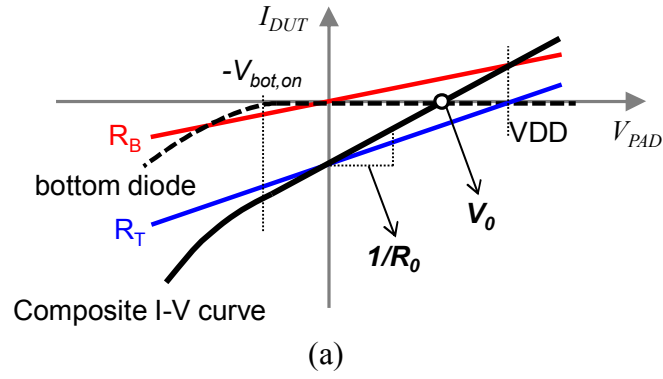


Figure A.2. Power-on I-V curves from (a) PAD-to-VSS, and (b) PAD-to-VDD. Composite I-V curves are plotted as thick black lines by summing up curves from diode, R_T , and R_B . Two measurement parameters, x-intercept (V_0) and slope ($1/R_0$) are indicated.

Appendix B. T-Coil Parameter Calculation from Multi-Port S-Parameters

In this appendix, the equations to calculate T-coil parameters from multi-port spiral inductors will be derived. Basically, T-coil implemented by a center-tapped spiral inductor has three ports. When more inner points are tapped out (e.g. inductance-halving T-coil), the EM simulation generates huge S-parameter matrices with the same number of ports as that of the inductor.

B.1 Definitions of Inductance and Quality Factor

To begin with, various definitions of inductance and quality factor (Q-factor) should be revisited to use the proper definition for this study. Figure B.1(a) shows equivalent circuit model for a silicon spiral inductor [83], which is composed of effective self-inductance (L_{eff}), series resistance (R_s), shunt capacitance (C_s) between two terminals, and vertical parasitics including vertical capacitance from the inter-layer dielectric (C_{ILD}) and silicon substrate models with resistance (R_{sub}) and capacitance (C_{sub}). To expedite the analysis, the model components are lumped into admittance denoted as Y_0 for series admittance between P1 and P2, and Y_1 and Y_2 for vertical admittance, as shown in Figure B.1(b). Each lumped admittance, Y_0 - Y_2 , can be calculated from 2-port Y-parameters, as follows:

$$Y_0 = -Y_{21} \text{ (or } -Y_{12} \text{)} \quad (\text{B.1})$$

$$Y_1 = Y_{11} + Y_{12} \quad (\text{B.2})$$

$$Y_2 = Y_{22} + Y_{21}. \quad (\text{B.3})$$

According to the circuit configuration, inductance and Q-factor could be defined differently. For example, differential LC-VCO use a spiral inductor as differential mode; each terminal P1 and P2 are connected to the signals with opposite polarity, and the ground connections of Y_1 and Y_2 no longer exist but they are virtually grounded [84], as shown in Figure B.2(a). Figure B.2(b) presents single-ended excitation, where P1 is connected to the stimulus input and P2 is grounded; this is the case of inductively-peaked amplifiers in which spiral inductors are used as loads. Therefore, in differential excitation in Figure

B.2(a), inductance and Q-factor are calculated as follows using 2-port Y-parameters, if the additional admittance Y1 and Y2 are negligible:

$$L = \frac{\text{Im}(-1/Y_{21})}{\omega} \quad (\text{B.4})$$

$$Q = \frac{\text{Im}(-1/Y_{21})}{\text{Re}(-1/Y_{21})}. \quad (\text{B.5})$$

With the same way, inductance and Q-factor for singled-ended excitation in Figure B.2(a) are calculated as below [85],

$$L = \frac{\text{Im}(1/Y_{11})}{\omega} \quad (\text{B.6})$$

$$Q = \frac{\text{Im}(1/Y_{11})}{\text{Re}(1/Y_{11})} = -\frac{\text{Im}(Y_{11})}{\text{Re}(Y_{11})}. \quad (\text{B.7})$$

The physical meanings of inductance definitions, (B.4) and (B.6) are slightly different. Inductance defined in (B.4) is usually called as “effective inductance,” which may be not a physical parameter, but a decomposed reactive component from the π -model in Figure B.1(a). Thus, this is useful to optimize stand-alone inductor or transformer structures in the device design. However, (B.6) provides inductance which incorporates all resistive and reactive components in a spiral inductor; this will be used in the context of the functional circuit design with inductors.

In T-coil design, inductance in each primary (L_1) and secondary (L_2) winding should be defined first, and from the total inductance (L_{tot}) between two outer terminals (P1 and P2), coupling factor (k) can be calculated. For this purpose, self-inductance should be calculated as effective inductance defined in (B.4). The series resistance can also be calculated in similar way to (B.4), by taking the real part of $-1/Y_{21}$. Q-factor of the total inductor (L_{tot}) will be calculated from (B.7); Q-factor of a T-coil is not a critical parameter, but it is to monitor the behavior of T-coil.

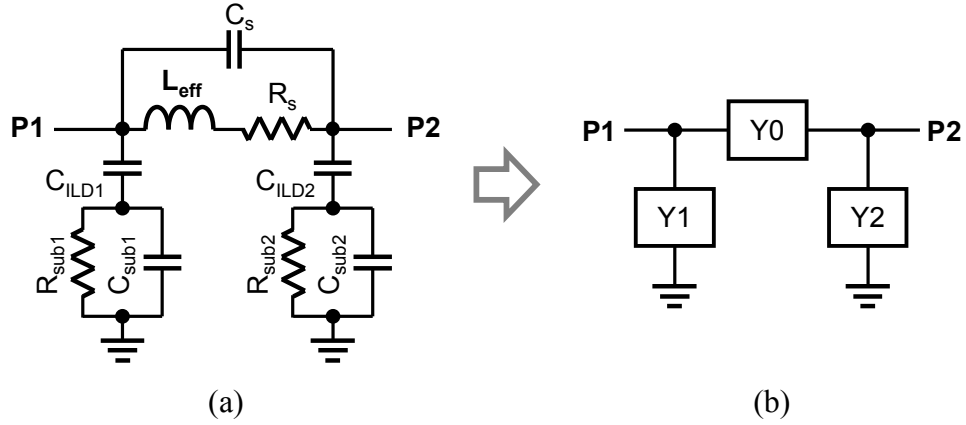


Figure B.1. On-chip spiral inductor model: (a) equivalent π -model and (b) simplified model with lumped admittance elements.

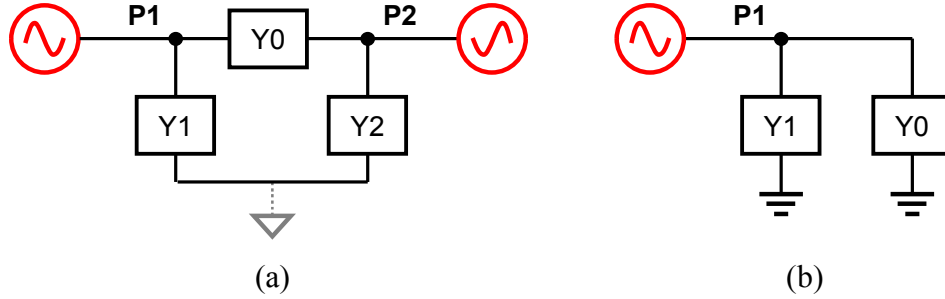


Figure B.2. Stimulus for Q-factor and inductance calculation: (a) differential excitation; P1 and P2 are differentially excited and ground connection of Y_1 and Y_2 are virtually grounded, and (b) single-ended excitation; P1 is connected to stimulus input and P2 is grounded (Y_2 no longer exists).

B.2 Parameter Extraction from Multi-Port Inductors

The previously derived equations cannot be directly applied to the 3-port T-coil devices. In this dissertation, similar methods used in [86] will be used but further modified to calculate k . Figure B.3(a) shows a black-box representation of 3-port T-coil with 3×3 Y-parameters, and Figure B.3(b) depicts simplified equivalent circuit with important T-coil parameters to be calculated.

First, to calculate total inductance (L_{tot}) between P1 and P2, input stimulus is placed at P1 with current I_x and voltage V_x , and P2 is grounded with 0V but I_2 current flows; see Figure B.4. Port P3 is floated for this calculation setup; voltage V_3 is seen at P3 but no current flows. The matrix representation for this situation is as follows:

$$\begin{bmatrix} I_x \\ I_2 \\ 0 \end{bmatrix} = \begin{bmatrix} Y_{11} & Y_{12} & Y_{13} \\ Y_{21} & Y_{22} & Y_{23} \\ Y_{31} & Y_{32} & Y_{33} \end{bmatrix} \begin{bmatrix} V_x \\ 0 \\ V_3 \end{bmatrix}. \quad (\text{B.8})$$

To calculate L_{tot} , transimpedance $Z_{\text{eff}12}$ from P1 to P2 should be calculated from (B.8), and then L_{tot} is calculated by taking imaginary part of $Z_{\text{eff}12}$:

$$Z_{\text{eff}12} = -\frac{V_x}{I_2} = \frac{Y_{33}}{Y_{23}Y_{31} - Y_{21}Y_{33}} \quad (\text{B.9})$$

$$L_{\text{tot}} = \frac{\text{Im}(Z_{\text{eff}12})}{\omega}. \quad (\text{B.10})$$

Total Q-factor (Q_{tot}) is calculated from the input impedance, $Z_{\text{in}12}$ as follows:

$$Z_{\text{in}12} = \frac{V_x}{I_x} = \frac{Y_{33}}{Y_{11}Y_{33} - Y_{13}Y_{31}} \quad (\text{B.11})$$

$$Q_{\text{tot}} = \frac{\text{Im}(Z_{\text{in}12})}{\text{Re}(Z_{\text{in}12})}. \quad (\text{B.12})$$

Now, parameters in the primary and secondary winding will be calculated from the setup in Figure B.5.

Inductance (L_1) and resistance (R_1) in the primary side can be calculated from transimpedance $Z_{\text{eff}13}$ from P1 to P3 as follows (assuming that the tap-resistance on P3 is negligible):

$$Z_{\text{eff}13} = -\frac{V_x}{I_3} = \frac{Y_{22}}{Y_{32}Y_{21} - Y_{31}Y_{22}} \quad (\text{B.13})$$

$$L_1 = \frac{\text{Im}(Z_{\text{eff}13})}{\omega} \quad (\text{B.14})$$

$$R_1 \approx R_{\text{eff}13} = \text{Re}(Z_{\text{eff}13}). \quad (\text{B.15})$$

With the same method, secondary winding parameters can be calculated in the following equations:

$$Z_{\text{eff}23} = -\frac{V_x}{I_3} = \frac{Y_{11}}{Y_{31}Y_{12} - Y_{32}Y_{11}} \quad (\text{B.16})$$

$$L_2 = \frac{\text{Im}(Z_{eff\,23})}{\omega} \quad (\text{B.17})$$

$$R_2 \approx R_{eff\,23} = \text{Re}(Z_{eff\,23}). \quad (\text{B.18})$$

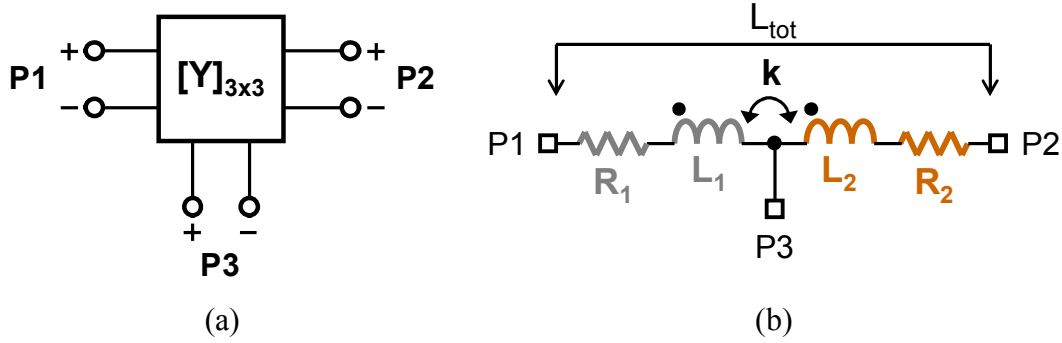


Figure B.3. T-coil model: (a) a network model with 3-port Y-parameters from center-tapped inductor for T-coil and (b) simplified equivalent circuit with important parameters.

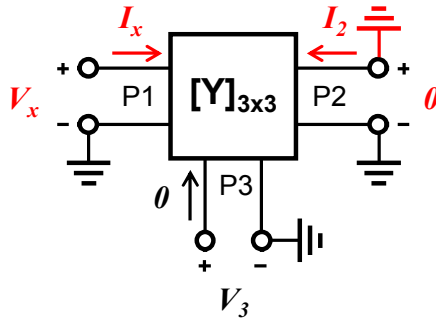


Figure B.4. Setup for L_{tot} calculation. Input stimulus is at P1 and P2 is grounded.

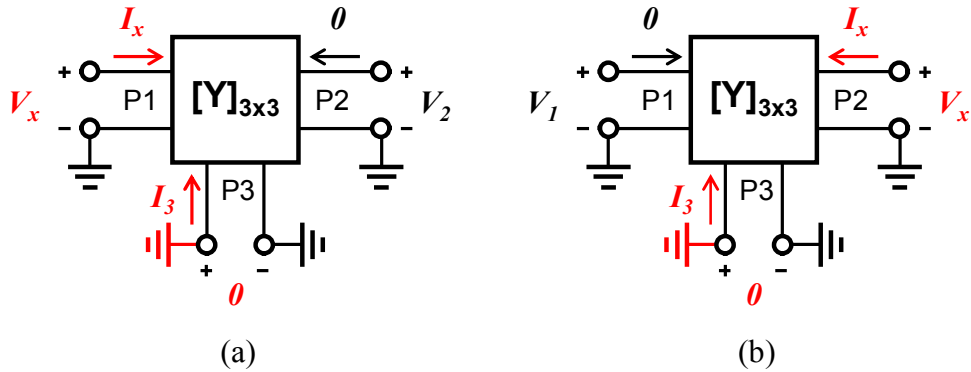


Figure B.5. Calculation setups for (a) primary winding parameters (from P1 to P3) and (b) secondary winding parameters (from P2 to P3).

To calculate coupling factor, the original T-coil in Figure B.6(a) can be decoupled as shown in Figure B.6(b), considering the dot position in the transformer. The definition of mutual inductance, M is the same as in (3.9). From Figure B.6(b), the total inductance L_{tot} can be expressed by

$$L_{tot} = L_1 + L_2 + 2M = L_1 + L_2 + 2k\sqrt{L_1L_2}. \quad (B.19)$$

Thus, coupling factor, k can be calculated as follows, using inductances expressed by (B.10), (B.14) and (B.17):

$$k = \frac{L_{tot} - L_1 - L_2}{2\sqrt{L_1L_2}}. \quad (B.20)$$

For multi-port T-coils with more than three ports, the same method can be applied to calculate each individual parameter.

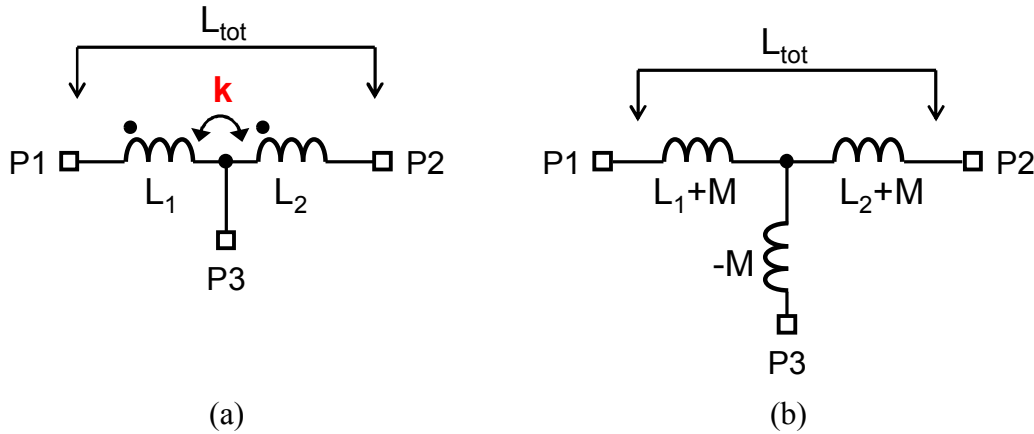


Figure B.6. T-coil circuit conversion: (a) coupled transformer through k and (b) decoupled transformer.

B.3 Calculation of Tap-Resistance

As explained in Section 3.4.2, the resistance between the center-tap and P3 in T-coils is directly related to the quasi-static voltage during ESD. This section shows how to calculate tap-resistance. Figure B.7 illustrates the equivalent circuit of the T-coil; tap-resistance, R_{tap} , is added between the center-tap (CT) and P3. In (B.15) and (B.18), tap-resistance was ignored because it is relatively smaller than the series resistance on the primary and secondary inductors (R_1 and R_2 in Figure B.7). However, R_{tap} is included in

this section for more accurate calculation. The effective resistance (R_{eff12}) between P1 and P2 is calculated as follows using Z_{eff12} in (B.9):

$$R_{eff12} = \text{Re}(Z_{eff12}) = R_1 + R_2. \quad (\text{B.21})$$

Rewriting (B.15) and (B.18) for R_{eff13} and R_{eff23} ,

$$R_{eff13} = \text{Re}(Z_{eff13}) = R_1 + R_{tap} \quad (\text{B.22})$$

$$R_{eff23} = \text{Re}(Z_{eff23}) = R_2 + R_{tap}. \quad (\text{B.23})$$

Equating (B.21)-(B.23), the actual series resistance at each inductor leg and tap-resistance can be calculated as follows:

$$R_1 = \frac{1}{2}(R_{eff12} + R_{eff13} - R_{eff23}) \quad (\text{B.24})$$

$$R_2 = \frac{1}{2}(R_{eff12} + R_{eff23} - R_{eff13}) \quad (\text{B.25})$$

$$R_{tap} = \frac{1}{2}(R_{eff13} + R_{eff23} - R_{eff12}). \quad (\text{B.26})$$

Using (B.24) and (B.25), more accurate series resistance on L_1 and L_2 can be calculated, and (B.26) denotes tap-resistance on P3.

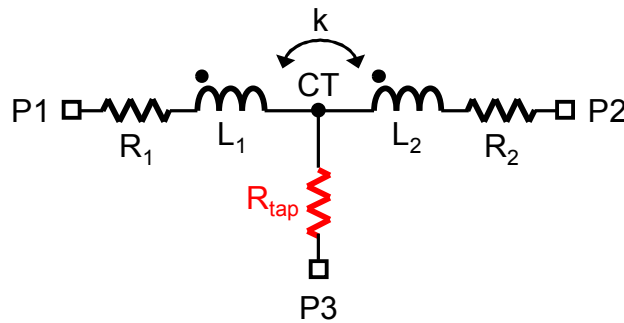


Figure B.7. Equivalent circuit of the T-coil with tap-resistance, R_{tap} .

References

- [1] S. Narendra, L. Fujino and K. Smith, "Through the looking glass? The 2015 edition: Trends in solid-state circuits from ISSCC," *IEEE Solid-State Circuits Magazine*, vol. 7, no. 1, pp. 14-24, 2015.
- [2] C. Duvvury, "ESD qualification changes for 45nm and beyond," in *Int. Electron Devices Meeting*, 2008.
- [3] JEDEC, "Recommended ESD-CDM target levels," JEP157, 2009.
- [4] JEDEC, "Recommended ESD target levels for HBM/MM qualification," JEP155A, 2012.
- [5] R. Merrill and E. Issaq, "ESD design methodology," in *Proc. EOS/ESD Symp.*, 1993.
- [6] M. S. Keel, N. Jack and E. Rosenbaum, "ESD-resilient active biasing scheme for high-speed SSTL I/Os," in *Proc. EOS/ESD Symp.*, 2013.
- [7] N. Jack and E. Rosenbaum, "ESD protection for high-speed receiver circuits," in *Intl. Reliability Physics Symp.*, 2010.
- [8] G. D. Lee, J. H. Chun, S. Cao, S. G. Beebe, K. W. Kwon and R. W. Dutton, "Investigation on the behavior of stacked devices within output drivers under ESD conditions," *IEEE Trans. Electron Devices*, vol. 59, no. 12, pp. 3313-3320, 2012.
- [9] M. G. Khazhinsky, J. W. Miller, M. Stockinger and J. C. Weldon, "Engineering single NMOS and PMOS output buffers for maximum failure voltage in advanced CMOS technologies," in *Prof. EOS/ESD Symp.*, 2004.
- [10] Y. Li, J. D. Sarro, S. Chang, J. Li, R. Gauthier and R. Halbach, "Maximizing ESD robustness of current-mode logic (CML) driver with internal gate bias network," in *Proc. EOS/ESD Symp.*, 2013.
- [11] J. Li, R. Mishra, M. Shrivastava, Y. Yang, R. Gauthier and C. Russ, "Technology scaling effects on the ESD performance of silicide-blocked PMOSFET devices in nanometer bulk CMOS technologies," in *Proc. EOS/ESD Symp.*, 2011.
- [12] JEDEC, "Stub series terminated logic for 1.8V (SSTL_18)," JESD8-15A, 2003.
- [13] M. Baze, "Two novel fully complementary self-biased CMOS differential amplifier," *IEEE J. Solid-State Circuits*, vol. 26, no. 2, pp. 165-168, 1991.
- [14] ESDA/JEDEC, "For electrostatic discharge sensitivity testing, human body model (HBM) – component level," ANSI/ESDA/JEDEC JS-001-2010, 2010.
- [15] J. Lee and E. Rosenbaum, "Voltage clamping requirements for ESD protection of inputs in 90nm CMOS technology," in *Proc. EOS/ESD Symp.*, 2008.
- [16] JEDEC, "Field-induced charged-device model test method for electrostatic-discharge-withstand thresholds of microelectronic components," JESD22-C101F, 2009.
- [17] J. M. Rabaey, A. Chandrakasan and B. Nikolic, *Digital Integrated Circuits: A Design Perspective*, Second Edition, Prentice Hall, 2003.
- [18] "Spectre Circuit Simulator," Cadence, [Online]. Available: http://www.cadence.com/products/cic/spectre_circuit/pages/default.aspx.
- [19] T. J. Maloney and N. Khurana, "Transmission line pulsing techniques for circuit modeling of ESD phenomena," in *Proc. EOS/ESD Symp.*, 1985.

- [20] H. Gieser and M. Haunschild, "Very-fast transmission line pulsing of integrated structures and the charged device model," in *Proc. EOS/ESD Symp.*, 1996.
- [21] S. Cao, T. W. Chen, S. G. Beebe and R. W. Dutton, "ESD design challenges and strategies in deeply-scaled integrated circuits," in *Proc. Custom Integrated Circuits Conf.*, 2009.
- [22] N. Jack and E. Rosenbaum, "Comparison of FICDM and wafer-level CDM test methods," *IEEE Trans. Device Mater. Rel.*, vol. 13, no. 2, pp. 379-387, 2013.
- [23] J. Li, K. Chatty, R. Gauthier, R. Mishra and C. Russ, "Technology scaling of advanced bulk CMOS on-chip ESD protection down to the 32nm node," in *Proc. EOS/ESD Symp.*, 2009.
- [24] S. Shekhar, J. S. Walling and D. J. Allstot, "Bandwidth extension techniques for CMOS amplifiers," *IEEE J. Solid-State Circuits*, vol. 41, no. 11, pp. 2424-2439, 2006.
- [25] T. H. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*, New York: Cambridge Univ. Press, 1998.
- [26] S. Gondi and B. Razavi, "Equalization and clock and data recovery techniques for 10-Gb/s CMOS serial-Link receivers," *IEEE J. Solid-State Circuits*, vol. 42, no. 9, pp. 1999-2011, 2007.
- [27] S. Galal and B. Razavi, "10-Gb/s limiting amplifier and laser/modulator driver in 0.18- μ m CMOS technology," *IEEE J. Solid-State Circuits*, vol. 38, no. 12, pp. 2138-2146, 2003.
- [28] J. W. Jung and B. Razavi, "A 25 Gb/s 5.8 mW CMOS equalizer," *IEEE J. Solid-State Circuits*, vol. 50, no. 2, pp. 515-526, 2015.
- [29] J. Kim, J. K. Kim, B. J. Lee and D. K. Jeong, "Design optimization of on-chip inductive peaking structures for 0.13- μ m CMOS 40-Gb/s transmitter circuits," *IEEE. Tans. Circuits and Systems*, vol. 56, no. 12, pp. 2544-2555, 2009.
- [30] M. Kossel, C. Menolfi, J. Weiss, P. Buchmann, G. von Bueren, L. Rodoni, T. Morf, T. Toifl and M. Schmatz, "A T-coil-enhanced 8.5 Gb/s high-swing SST transmitter in 65 nm bulk CMOS with < 16 dB return loss over 10 GHz bandwidth," *IEEE. J. Solid-State Circuits*, vol. 43, no. 12, pp. 2905-2920, 2008.
- [31] S. Galal and B. Razavi, "Broadband ESD protection circuits in CMOS technology," *IEEE. J. Solid-State Circuits*, vol. 38, no. 12, pp. 2334-2340, 2003.
- [32] J. D. Irwin, *Basic Engineering Circuit Analysis*, Fifth Edition, Prentice-Hall, Inc., 1996.
- [33] D. Linten, S. Thijs, J. Borremans, M. Dehan, D. Tremouilles, M. Scholz, M. I. Natarajan, P. Wambacq and G. Groeseneken, "T-Diode - A novel plug-and-play wideband RF circuit ESD protection methodology," in *Proc. EOS/ESD Symp.*, 2007.
- [34] T.-S. Horng, J.-M. Wu, L.-Q. Yang and S.-T. Fang, "A novel modified-T equivalent circuit for modeling LTCC embedded inductors with a large bandwidth," *IEEE Trans. Microwave Theory and Techniques*, vol. 51, no. 12, pp. 2327-2333, 2003.
- [35] F. A. Muller, "High-frequency compensation of RC amplifiers," *Proc. IRE*, vol. 42, no. 8, pp. 1271-1276, 1954.
- [36] P. Staric and E. Margan, *Wideband Amplifiers*, Springer, 2006.
- [37] J. Paramesh and D. J. Allstot, "Analysis of the bridged T-coil circuit using the extra-element theorem," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 53, no. 12, p. 1408-1412, 2006.
- [38] S. C. Dutta Roy, "Comments on 'analysis of the bridged T-coil circuit using the extra-element

- theorem'," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 54, no. 8, pp. 673-674, 2007.
- [39] J. Hoagg and D. Bernstein, "Nonminimum-phase zeros - Much to do about nothing - Classical control revisited - Part II," *IEEE Control Systems Magazine*, vol. 27, no. 3, pp. 45-57, 2007.
 - [40] D. Linten, S. Thijs, M. Okushima, M. Scholz, J. Borremans, M. Dehan and G. Groeseneken, "A 4.5 kV HBM, 300 V CDM, 1.2 kV HMM ESD protected DC-to-16.1 GHz wideband LNA in 90 nm CMOS," in *Proc. EOS/ESD Symp.*, 2009.
 - [41] S. M. Yim and K. K. O, "Switched resonators and their applications in a dual-band monolithic CMOS LC-tuned VCO," *IEEE Trans. Microwave Theory and Techniques*, vol. 54, no. 1, pp. 74-81, 2006.
 - [42] M. Demirkan, S. P. Bruss and R. R. Spencer, "Design of wide tuning-range CMOS VCOs using switched coupled-inductors," *IEEE J. Solid-State Circuits*, vol. 43, no. 5, pp. 1156-1163, 2008.
 - [43] "Advanced Design System (ADS)," Keysight Technologies, [Online]. Available: <http://www.keysight.com/en/pc-1297113/advanced-design-system-ads?cc=US&lc=eng>.
 - [44] C. P. Yue and S. S. Wong, "On-chip spiral inductors with patterned ground shields for Si-based RF ICs," *IEEE J. Solid-State Circuits*, vol. 33, no. 5, pp. 743-752, 1998.
 - [45] A. Niknejad and R. Meyer, "Analysis of eddy-current losses over conductive substrates with applications to monolithic inductors and transformers," *IEEE Trans. Microwave Theory and Techniques*, vol. 49, no. 1, pp. 166-176, 2001.
 - [46] W. L. Wang, S. A. Saller, W. R. Hunter, P. K. Chatterjee and P. Yang, "Direct moat isolation for VLSI," *IEEE J. Solid-State Circuits*, vol. 17, no. 2, pp. 171-177, 1982.
 - [47] S. Hyvonen and E. Rosenbaum, "Diode-based tuned ESD protection for 5.25-GHz CMOS LNAs," in *Proc. EOS/ESD Symp.*, 2005.
 - [48] "Momentum 3D planar EM simulator," Keysight Technologies, [Online]. Available: <http://www.keysight.com/en/pc-1887116/momentum-3d-planar-em-simulator?cc=US&lc=eng>.
 - [49] S. J. Pan, L. W. Li and W. Y. Yin, "Performance trends of on-chip spiral inductors for RFICs," *Progress In Electromagnetic Research*, vol. PIER 45, pp. 123-151, 2004.
 - [50] B. Razavi, *RF Microelectronics*, Prentice Hall, 2011.
 - [51] J. R. Andrews, "RZ vs. NRZ," *Appl. Note AN-12*, Picosecond Pulse Labs, 2001.
 - [52] M. Brown, M. Dudek, A. Healey, L. B. Artsi, R. Mellitz, C. Moore, A. Ran and P. Zivny, "The state of IEEE 802.3bj 100 Gb/s backplane ethernet," in *DesignCon*, 2014.
 - [53] IEEE P802.3bj/D3.0, "IEEE Draft Standard for Ethernet Amendment 2: Physical Layer Specifications and Management Parameters for 100 Gb/s Operation over Backplanes and Copper Cables," 2014.
 - [54] P. Hanumolu, G. Wei and U. Moon, "Equalizers for high-speed serial links," *Int. J. High Speed Elec. Sys.*, vol. 15, no. 2, pp. 429-458, 2005.
 - [55] J. Lee, "A 20-Gb/s adaptive equalizer in 0.13- μ m CMOS technology," *IEEE J. Solid-State Circuits*, vol. 41, no. 9, pp. 2058-2066, 2006.
 - [56] R. Apfel and P. Gray, "A fast-settling monolithic feedforward op-amp using doublet compression techniques," in *IEEE Int. Solid-State Circuits Conf. Digest of Tech. Papers (ISSCC)*, 1974.
 - [57] R. York, "Frequency response and Bode plots," 2009. [Online]. Available: <http://my.ece.ucsb.edu/>

York/Bobsclass/2B/Frequency%20Response.pdf.

- [58] B. Razavi, "Devices and circuits for phase-locked systems," in *Phase-Locking in High-Performance Systems, From Devices to Architectures*, Wiley, 2003.
- [59] B. Razavi, *Design of Integrated Circuits for Optical Communications*, Hoboken, NJ: John Wiley & Sons, Inc., 2012.
- [60] JEDEC, "POD12 - 1.2 V pseudo open drain interface," JESD8-24, 2011.
- [61] S.-J. Bae, K.-I. Park, J.-D. Ihm, H.-Y. Song, W.-J. Lee, H.-J. Kim, K.-H. Kim, Y.-S. Park, M.-S. Park, H.-K. Lee, S.-Y. Bang, G.-S. Moon, S.-W. Hwang, Y.-C. Cho, S.-J. Hwang, D.-H. Kim, J.-H. Lim and J, "An 80 nm 4 Gb/s/pin 32 bit 512 Mb GDDR4 graphics DRAM with low power and low noise data bus inversion," *IEEE J. Solid-State Circuits*, vol. 43, no. 1, pp. 121-131, 2008.
- [62] JEDEC, "DDR4 SDRAM," JESD79-4A, 2013.
- [63] Fairchild Semiconductor, "Terminations for advanced CMOS logic," *Appl. Note AN-610*, 2006.
- [64] J. A. Mataya, G. W. Haines and S. B. Marshall, "IF amplifier using Cc compensated transistors," *IEEE J. Solid-State Circuits*, vol. 3, no. 12, pp. 401-407, 1968.
- [65] J. M. Miller, "Dependence of the input impedance of a three-electrode vacuum tube upon the load in the plate circuit," *Scientific Papers of the Bureau of Standards*, vol. 15, no. 351, pp. 367-385, 1920.
- [66] S. S. Mohan, C. P. Yue, M. del Mar Hershenson, S. S. Wong and T. H. Lee, "Modeling and characterization of on-chip transformers," in *Intl. Electron Devices Meeting Technical Digest*, 1998.
- [67] R. Mertens, "Understanding, modeling, and mitigating system-level ESD in integrated circuits," Ph.D. dissertation, University of Illinois at Urbana-Champaign, Urbana, IL, 2015.
- [68] T. Smedes and N. Guitard, "Harmful voltage overshoots due to turn-on behaviour of ESD protections during fast transients," in *Proc. EOS/ESD Symp.*, 2007.
- [69] Rogers corporation, [Online]. Available: <http://www.rogerscorp.com/index.aspx>.
- [70] V. Stojanovic, A. Ho, B. W. Garlepp, F. Chen, J. Wei, G. Tsang, E. Alon, R. T. Kollipara, C. W. Werner, J. L. Zerbe and M. A. Horowitz, "Autonomous dual-mode (PAM2/4) serial link transceiver with adaptive equalization and data recovery," *IEEE J. Solid-State Circuits*, vol. 40, no. 4, pp. 1012-1026, 2005.
- [71] H.-M. Bae, J. Ashbrook, J. Park, N. Shanbhag, A. Singer and S. Chopra, "An MLSE receiver for electronic dispersion compensation of OC-192 fiber links," *IEEE J. Solid-State Circuits*, vol. 41, no. 11, pp. 2541-2554, 2006.
- [72] M. Harwood, N. Warke, R. Simpson, T. Leslie, A. Amerasekera, S. Batty, D. Colman, E. Carr, V. Gopinathan, S. Hubbins, P. Hunt, A. Joy, P. Khandelwal, B. Killips, T. Krause, S. Lytollis, A. Pickering, M. Saxton, D. Sebastio, G. Swanson, A. Szczepanek, T. Ward, J. Williams, R. Williams and T. Willwerth, "A 12.5Gb/s SerDes in 65nm CMOS using a baud-rate ADC with digital receiver equalization and clock recovery," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, 2007.
- [73] E. H. Chen and C. K. K. Yang, "ADC-based serial I/O receivers," *IEEE Trans. Circuits and Systems I*, vol. 57, no. 9, pp. 2248-2258, 2010.
- [74] M. Lu, N. Shanbhag and A. Singer, "BER-optimal analog-to-digital converters for communication links," in *Int. Symposium on Circuit and Systems*, 2010.
- [75] R. Narasimha, N. Shanbhag and A. Singer, "BER-based adaptive ADC-equalizer based receiver for

- communication links," in *IEEE Workshop on Signal Processing System*, 2010.
- [76] W. Kester, "ADC architectures I: the flash converter," Application Note MT-020, Analog Devices, Inc., 2009.
 - [77] C.-M. Hsu, M. Straayer and M. 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, no. 12, pp. 2776-2786, 2008.
 - [78] K. Nakamura, M. Fukaishi, Y. Hirota, Y. Nakazawa and M. Yotsuyanagi, "A CMOS 50% duty cycle repeater using complimentary phase blending," in *Symp. VLSI Circuits Dig. Tech. Papers*, 2000.
 - [79] B. Brannon and A. Barlow, "Aperture uncertainty and ADC system performance," Application Note AN-501, Analog Devices, Inc., 1998.
 - [80] E.-H. Chen, R. Yousry, T. Ali and C.-K. Yang, "10Gb/s serial I/O receiver based on variable reference ADC," in *Symp. VLSI Circuits Dig. Tech. Papers*, 2011.
 - [81] J. Cao, B. Zhang, U. Singh, D. Cui, A. Vasani, A. Garg, W. Zhang, N. Kocaman, D. Pi, B. Raghavan, H. Pan, I. Fujimori and A. Momtaz, "500mW digitally calibrated AFE in 65nm CMOS for 10Gb/s Serial links over backplane and multimode fiber," in *IEEE Int. Solid-State Circuits Conf. Digest of Tech. Papers (ISSCC)*, 2009.
 - [82] B. Abiri, A. Sheikholeslami, H. Tamura and M. Kibune, "A 5Gb/s adaptive DFE for 2x blind ADC-based CDR in 65nm CMOS," in *IEEE Int. Solid-State Circuits Conf. Digest of Tech. Papers (ISSCC)*, 2011.
 - [83] C. P. Yue and S. S. Wong, "Physical modeling of spiral inductors on silicon," *IEEE Tans. Electron Devices*, vol. 47, no. 3, pp. 560-568, 2000.
 - [84] R. L. Bunch, D. I. Sanderson and S. Raman, "Quality factor and inductance in differential IC implementations," *IEEE Microwave Mag.*, vol. 3, no. 6, pp. 82-93, 2002.
 - [85] K. O, "Estimation methods for quality factors of inductors fabricated in silicon integrated circuit process technologies," *IEEE J. Solid-State Circuits*, vol. 33, no. 8, pp. 1249-1252, 1998.
 - [86] K. Okada and K. Masu, "Modeling of Spiral Inductors," in *Advanced Microwave Circuits and Systems*, V. Zhurbenko, Ed., InTech, 2010, pp. 291-311.