Title
Low-Power Techniques for CMOS Wireline Receivers

Permalink
https://escholarship.org/uc/item/02f6b4hz

Author
Manian, Abishek

Publication Date
2016

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Low-Power Techniques for CMOS Wireline Receivers

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Electrical Engineering

by

Abishek Manian

2016

ABSTRACT OF THE DISSERTATION

Low-Power Techniques for CMOS Wireline Receivers

by

Abishek Manian

Doctor of Philosophy in Electrical Engineering

University of California, Los Angeles, 2016

Professor Behzad Razavi, Chair

With the ever-increasing need for high throughput from chip-to-chip I/Os, wireline transceivers are being pushed to operate at higher speeds. With the increase in data rates, the power consumption of broadband receivers has become critical in multi-lane applications like the Gigabit Ethernet. It is therefore desirable to minimize the power drawn by all of the building blocks.

This work introduces a 40-Gb/s CMOS wireline receiver that advances the art by achieving a tenfold reduction in power and an efficiency of 0.35 mW/Gb/s. An innovative aspect of the proposed NRZ receiver is our "minimalist" approach, which recognizes that every additional stage in the data or clock path consumes more power *and* limits the bandwidth. The minimalist mentality avoids multiple stages in the front-end continuous-time linear equalizer (CTLE), quadrature oscillators in the clock and data recovery (CDR) circuit, clock or data buffers, or phase interpolation. Moreover, building blocks are shared among different functions so as to reduce the number of current paths between V_{DD} and ground. Using charge-steering techniques extensively, the receiver contains only a few static bias

currents adding up to about 6 mA. The minimalist approach also leads to a small footprint, about 110 μm × 175 μm, for the entire receiver, making it possible to design a multi-lane system in a small area and with short interconnects.

This receiver incorporates a one-stage CTLE with 5.5-dB boost, a one-tap discrete-time linear equalizer (DTLE) with 5.4-dB boost, a half-rate CDR circuit, a half-rate/quarter-rate decision-feedback equalizer, a 1:4 deserializer, and two new latch topologies. Since in recent designs, the CTLE draws significant power, this work introduces the DTLE as an efficient means of creating a high-frequency boost with only 0.3 mW. Fabricated in 45-nm CMOS technology, the receiver achieves a BER $< 10^{-12}$ with a recovered clock jitter of 0.515 ps_{rms}, a jitter tolerance of 0.45 UI_{pp} at 5 MHz, with a channel loss of 18.6 dB at Nyquist, while consuming 14 mW from a 1-V supply.

The dissertation of Abishek Manian is approved.

Milos D. Ercegovac

William J. Kaiser

Danijela Cabric

Behzad Razavi, Committee Chair

University of California, Los Angeles

2016

To the memory of my grandfather

List of Figures

xvi

LIST OF TABLES

Vita

2011	B.E. (Electronics and Telecommunication), University of Mumbai, India.
2012–2015	Graduate Student Researcher, Electrical Engineering Department, University of California, Los Angeles, USA.
2012	Intern, Qualcomm, San Diego, USA.
2012–2013	Teaching Assistant, Electrical Engineering Department, University of California, Los Angeles, USA.
2013	M.S. (Electrical Engineering), University of California, Los Angeles, USA.
2013	Analog/Mixed-Signal Design Intern, Xilinx Inc., San Jose, USA.
2013–2014	Teaching Associate, Electrical Engineering Department, University of California, Los Angeles, USA.
2014, 2015	Design Engineering Intern, Texas Instruments, Atlanta, USA.
2015	Teaching Fellow, Electrical Engineering Department, University of California, Los Angeles, USA.

Publications

A. Manian and B. Razavi, "A 32-Gb/s 9.3-mW CMOS Equalizer with 0.73-V Supply," *Proc. IEEE Custom Integrated Circuits Conference (CICC)*, pp. 1-4, Sept 2014.

A. Manian and B. Razavi, "A 40-Gb/s 9.2-mW CMOS Equalizer," *Symposium on VLSI Circuits Dig. of Tech. Papers*, pp. 226-227, June 2015.

A. Manian and B. Razavi, "A 40Gb/s 14mW CMOS Wireline Receiver," *IEEE ISSCC Dig. Tech. Papers*, pp. 412-413, Feb 2016.

CHAPTER 1

Introduction

1.1 Motivation

The push for higher data rates in copper media continues unabated. The use of fewer lanes to carry faster data is attractive, especially if the power dissipation per lane can be maintained relatively constant. In the limit, most of the transceiver power is dissipated in the unscalable termination resistors at the transmitter output and the receiver input. It is therefore desirable to minimize the power drawn by all of the building blocks.
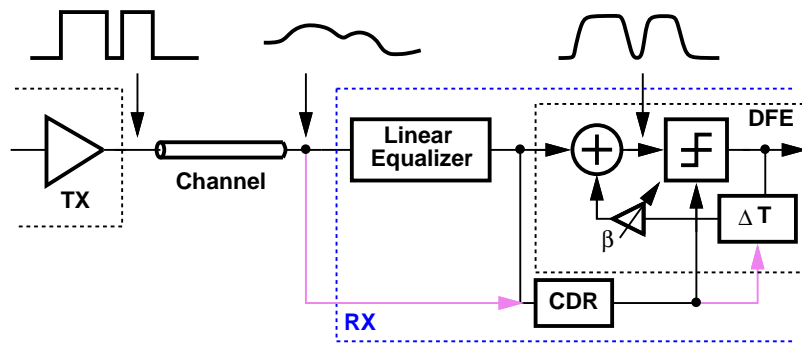
Figure 1.1: Generic wireline transceiver link.

Shown in Fig. 1.1 is a generic wireline link with a lossy channel with a typical channel profile as shown in Fig. 1.2. At higher data rates, the channel frequency response results in loss and intersymbol interference (ISI). A transmitted square pulse appears smeared across multiple bit periods at the input of the receiver, also
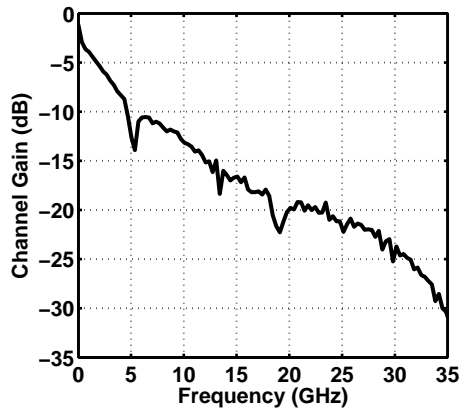
Figure 1.2: Frequency response of lossy channel.

with a reduced amplitude, as shown in Fig. 1.3. Hence, on the receiver side we typically need both a linear equalizer and a decision-feedback equalizer (DFE), to sharpen the pulse such that its effect remains only within its bit period, as shown in Fig. 1.3, such that the previous and the next sampling instants remain unaffected by the present bit.
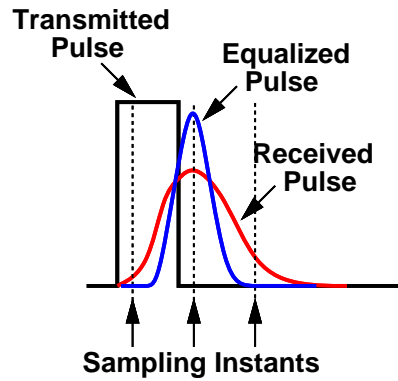
Figure 1.3: Pulse response at transmitter output, receiver input and equalizer output.

The role of equalizers are better appreciated by looking at the eye diagrams at various points in the system, as depicted in Fig. 1.4. The transmitter output is typically a waveform with sharp rise and fall times, with a wide open eye. When this data travels through the channel, it experiences both loss and ISI, which
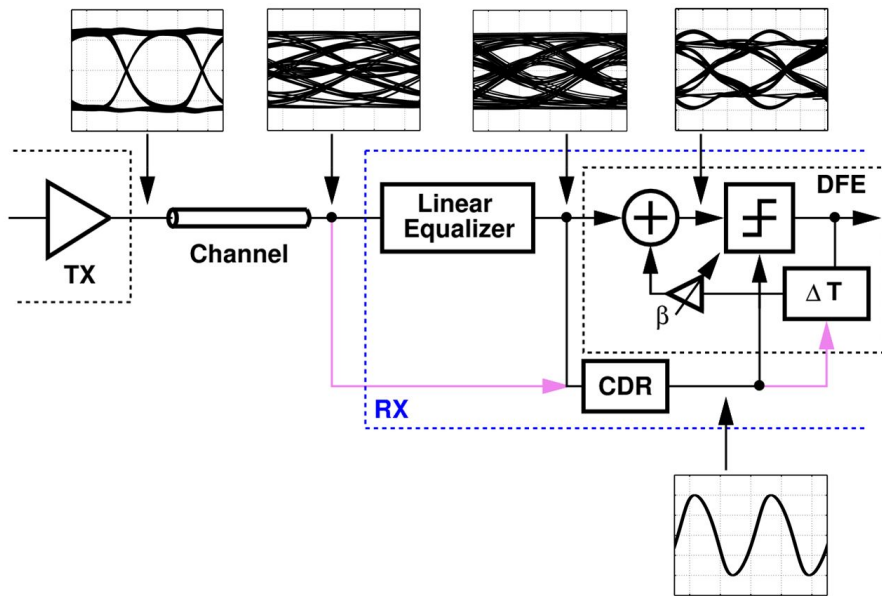
Figure 1.4: Eye diagrams at various points in a wireline link.

results in a closed eye at the receiver input. The receiver front-end typically consists of a linear equalizer commonly implemented in continuous time, which effectively reduces the length of the channel, thus improving the eye diagram at its output. Since a continuous-time linear equalizer (CTLE) cannot correct for deep notches in the channel response, we also need a DFE (clocked or unclocked [1, 2]), which can increase the eye opening enough to achieve a bit error rate of typically 10^{-9} or lower, without any error correction schemes. In addition, the receiver also needs a clock and data recovery (CDR) circuit to recover the clock information from the input data and adjust the clock phase such that it samples the output of the summer around the middle of the eye to achieve the best bit error rate possible.

The performance of receivers in high-speed links is typically quantified by speed and power numbers. However, two other factors need to be considered as well: (1) Channel loss: The higher the loss, the more difficult the design and the higher the power consumption, and (2) Robustness in terms of the bit error

rate. It is possible to define a figure of merit (FOM) which takes into account the channel loss [3], but this FOM has not been widely adopted.

Table 1.1 shows a few examples of state-of-the-art receivers. At 40 Gb/s, recent receivers consume from 150 mW [4] to 1 W [5].

Table 1.1: State-of-the-art receivers at 40 Gb/s

Reference	Hsieh VLSI 2011	Chen JSSC Mar. 2012	Raghavan JSSC Dec. 2013
Data Rate (Gb/s)	40	40	40
Supply (V)	1.2 for DFE/CDR, 1.5 for CTLE	1.6	1
Channel Loss at Nyquist (dB)	23.5	19	>21
Bit Error Rate	$<10^{-12}$	$<10^{-12}$	$<10^{-12}$
Power (mW)	150	520	1050†
Power Efficiency (pJ/bit)	3.75	13	26.25
Recovered Clock Jitter (ps)	6.8 pp	0.319 rms	-
Jitter Tolerance	-	≈ 0.65 UI$_{pp}$ at 10 MHz	0.95 UI$_{pp}$ at 10 MHz‡
Area (mm^2)	0.278	1.1475*	3.9*
Technology	65–nm CMOS	65–nm CMOS	40–nm CMOS

* Includes pads   † Includes SFI–5.2 TX; 350 mW for line–side RX
‡ Measured for BER $= 10^{-9}$

1.2   Organization

This dissertation describes a number of architecture and circuit techniques that reduce the power consumption of wireline receivers around 40 Gb/s by a factor of ten. Chapter 2 describes some of the commonly used equalization techniques in high-speed wireline receivers.

Chapter 3 introduces the basics of clock and data recovery for wireline communication, and describes a few CDR architectures used in practice.

Chapter 4 presents a full-rate equalizer at 32 Gb/s that operates with a supply voltage as low as 0.73 V. This design uses a CTLE/DFE cascade incorporating inductor nesting to reduce chip area and latch feedforward to improve the loop speed.

Chapter 5 describes a half-rate discrete-time equalizer at 40 Gb/s based on charge-steering techniques. This circuit incorporates a power-efficient discrete-time linear equalizer and two new charge-steering latch topologies.

Chapter 6 introduces a 40-Gb/s CMOS wireline receiver that advances the art by achieving a tenfold reduction in power and an efficiency of 0.35 mW/Gb/s. This performance is achieved through the use of a minimalist approach, hardware sharing, and charge-steering techniques.

Chapter 7 summarizes the contributions of this work and offers suggestions for future work.

CHAPTER 2

Equalization in Receivers

As described in Chapter 1, wireline receivers typically employ equalization techniques to compensate for the loss and ISI introduced by the channel. From a circuit-design point-of-view, it is easier to design continuous-time linear equalizers (CTLE), but these cannot account for deep notches in the channel's frequency response. In such cases, a DFE can be used for equalization. However, a DFE cannot correct for pre-cursor ISI. Hence, a linear equalizer is often combined with a DFE in high-speed receivers.

2.1 Bit Error Rate

Figure 2.1 shows a generic equalizer architecture with a linear equalizer and a one-tap DFE. Assuming that noise exhibits a Gaussian distribution with zero mean, the probability of error, also known as the bit error rate (BER), can be written as [7]

$$\text{BER} = Q\left(\frac{V_{pp,eq}}{2\sqrt{\overline{V_{n,eq}^2}}}\right), \tag{2.1}$$

where $V_{pp,eq}$ is the peak-to-peak signal swing at the summing node X, $\overline{V_{n,eq}^2}$ is the noise variance, and

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left(\frac{-u^2}{2}\right) du. \tag{2.2}$$

6
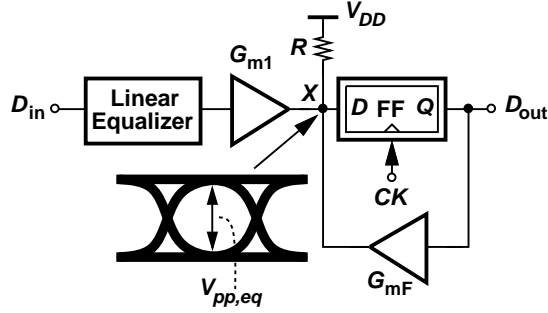
Figure 2.1: Generic receiver equalizer.

$Q(7) \approx 10^{-12}$. In other words, to achieve a BER of 10^{-12},

$$\frac{V_{pp,eq}}{2} \approx 7\sqrt{\overline{V_{n,eq}^2}}. \tag{2.3}$$

Equation (2.1) takes into account the probability of making an error due to noise alone. Including other factors such as offsets and sensitivity of the sampling flipflop of Fig. 2.1, the BER can be expressed as [3]

$$\text{BER} = \frac{1}{2} \, Q\left(\frac{V_{pp,eq}/2 - V_{os} - V_{sens}}{\sqrt{\overline{V_{n,eq}^2}}}\right), \tag{2.4}$$

where V_{os} is the total offset referred to the summing node X in Fig. 2.1, and V_{sens} is the sensitivity of the flipflop.

Thus, for BER $= 10^{-12}$, the requirement on the vertical eye opening $V_{pp,eq}$ at the summing node is given as

$$V_{pp,eq} \geq 14\sqrt{\overline{V_{n,eq}^2}} + 2(V_{os} + V_{sens}). \tag{2.5}$$

2.2 Linear Equalizers

A DFE can only correct for post-cursor ISI. However, for high-loss channels, the ISI introduced by pre-cursors can be significant, especially at higher data rates.

In order to correct for the pre-cursor ISI, it is necessary to have a linear equalizer either at the transmitter or the receiver or both.

There are several ways of implementing a linear equalizer: the passive linear equalizer, the FIR filter, and the capacitive degeneration amplifier.

2.2.1 Passive Linear Equalizer

A passive linear equalizer may be implemented as a high-pass filter, shown in Fig. 2.2(a) [8]. However, this circuit cannot provide any gain and hence, introduces loss at lower frequencies.

(a)   (b)
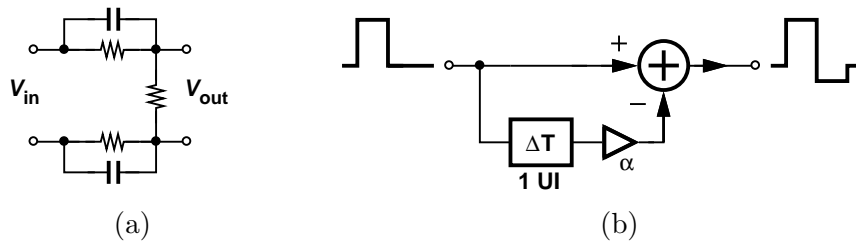
Figure 2.2: (a) Passive linear equalizer, (b) FIR implementation of a linear equalizer.

2.2.2 Discrete-Time FIR Filters

Linear equalization in receivers can also be in discrete time using FIR filters as shown in Fig. 2.2(b). As opposed to transmitters, the FIR filter in a receiver must have a linear delay element, which can be implemented using passive delay lines [9] or active stages [1], or a combination of both [10]. FIR filters used as linear equalizers in [11, 12] require multiple clock phases which can be power-hungry. Chapter 5 introduces a discrete-time linear equalizer that creates a boost of about 5.4 dB at 40 Gb/s while consuming only 0.3 mW.

2.2.3 Capacitive Degeneration Amplifier

The most common implementation of a linear equalizer is a continuous-time capacitive degeneration amplifier, commonly known as a continuous-time linear equalizer, or a CTLE, shown in Fig. 2.3(a) [13].

(a)   (b)
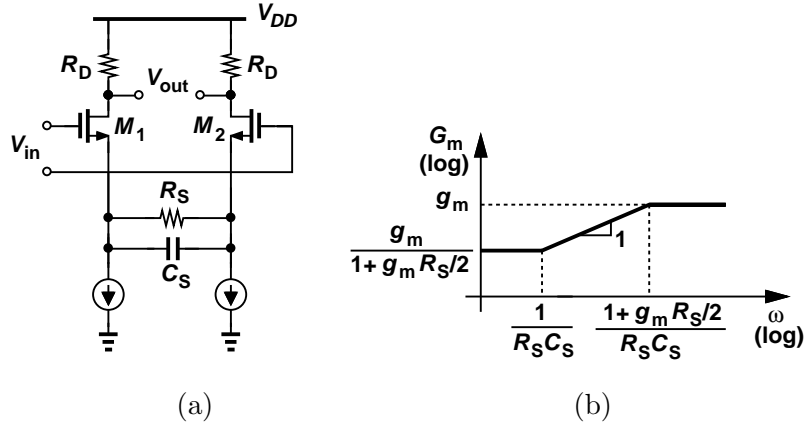
Figure 2.3: Capacitive degeneration amplifier: (a) Circuit, (b) frequency response.

This circuit creates a boost at higher frequencies by introducing a zero and a pole in the transfer function using capacitive degeneration, as seen from the frequency response of the transconductance in Fig. 2.3(b). The amount of boost is given by $1 + g_m R_S/2$, where g_m is the transconductance of M_1 and M_2. Thus, the high-frequency boost can be controlled by adjusting the value of R_S.

The bandwidth of this circuit is typically limited by the output pole given by $1/(R_D C_L)$, where C_L denotes the single-ended capacitance at the output. Section 2.4 discusses the choice of bandwidth for these stages based on noise and ISI considerations. At high data rates, it might be difficult to meet these bandwidth requirements. In order to extend the bandwidth beyond $1/(R_D C_L)$, broadband peaking techniques may be used. Some of these techniques are described in the next section.

9

2.3 Broadband Techniques

As the data rate increases, bandwidth requirements become more stringent. In order to extend the bandwidth of high-speed stages, the following techniques may be used.

2.3.1 Principle of Peaking

Broadband current-mode logic (CML) stage in Fig. 2.4(a) can be modeled by an equivalent circuit depicted in Fig. 2.4(b), where C represents the total capacitance at the output node.

(a)   (b)
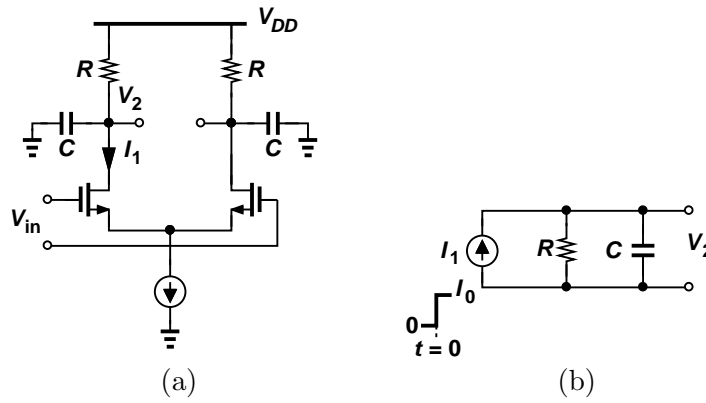
Figure 2.4: (a) CML stage, and (b) its equivalence to a parallel RC circuit.

For the circuit in Fig. 2.4(b), the time constant is given as RC. For a step at $t = 0$ in current I_1 from 0 to I_0, the voltage V_2 in this RC circuit is given as

$$V_2(t) = I_0 R \left[1 - \exp\left(\frac{-t}{RC} \right) \right], \tag{2.6}$$

as plotted in Fig. 2.5.

The output voltage V_2 reaches 99.33% of its final value, $I_0 R$, after a time $5RC$.
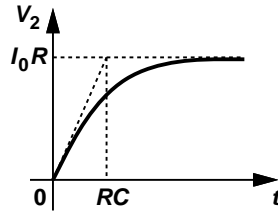
Figure 2.5: Transient response of a parallel RC circuit.

Now, let an ideal switch be connected in series with the resistor R, as shown in Fig. 2.6(a). Assume that this switch is somehow turned off at $t = 0$, and turned on at $t = RC$. At $t = 0^+$, since the switch is off, all of the current I_0 flows through the capacitor C, thus charging it at a rate I_0/C, as depicted in Fig. 2.6(b). Thus, for $0 < t < RC$, the output voltage V_2 can be written as

$$V_2(t) = \frac{I_0 t}{C}. \tag{2.7}$$
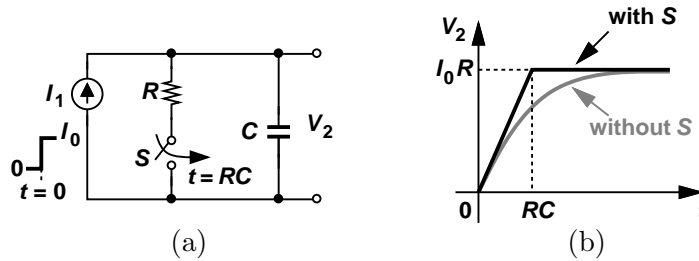
Figure 2.6: (a) Adding a switch to the parallel RC circuit, (b) transient response.

At $t = RC$, from the equation above, $V_2(RC) = I_0 R$. At this point, the switch turns on, thus causing all of the current to flow through the resistor R, thus maintaining a constant voltage $V_2 = I_0 R$, for $t > RC$. In other words, inserting the switch in series with the resistor reduces the time taken for V_2 to reach its final value from $\approx 5RC$ to RC, thus indicating a significant improvement in the bandwidth of the system. The only question that needs to answered now is how to control the turn-on and turn-off of this switch.

11

2.3.2 Inductive Shunt Peaking

The switch in the circuit of Fig. 2.6, can be implemented by an inductor L, as shown in Fig. 2.7.
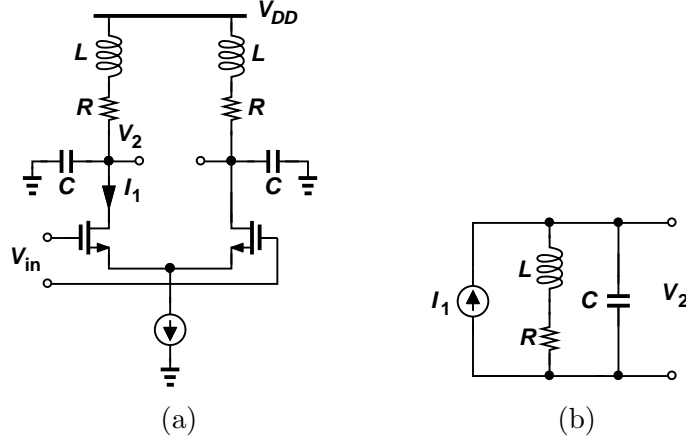
Figure 2.7: (a) Inductive shunt peaking, (b) equivalent circuit.

In order to intuitively understand the role of this inductor, let us focus on the time-domain behavior of this circuit. In Fig. 2.7(b), when I_1 is stepped from 0 to I_0 at time $t = 0$, the current through the inductor initially remains zero. This means all of the current I_0 flows through the capacitor initially, thus linearly charging it, similar to the operation of the circuit in Fig. 2.6. As the capacitor continues to charge, the inductor gradually starts carrying more and more current. The black curve in Fig. 2.8 shows the transient behavior of the current, I_L, through the inductor. This curve can be approximated by a piecewise waveform shown in gray in Fig. 2.8, thus indicating the behavior of this inductor as a switch.

The value of this inductor can be chosen according to the equation [14]

$$L = mR^2C, \tag{2.8}$$

where m is a design parameter typically chosen to be in the range 0.25 to 0.41
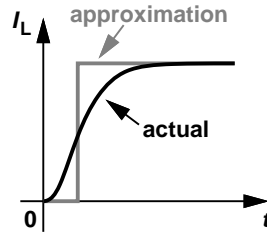
Figure 2.8: Transient response of current in the inductor.

for optimal peaking. $m = 0.41$ results in a maximally flat amplitude response, while extending the 3-dB bandwidth by nearly 73% [15].
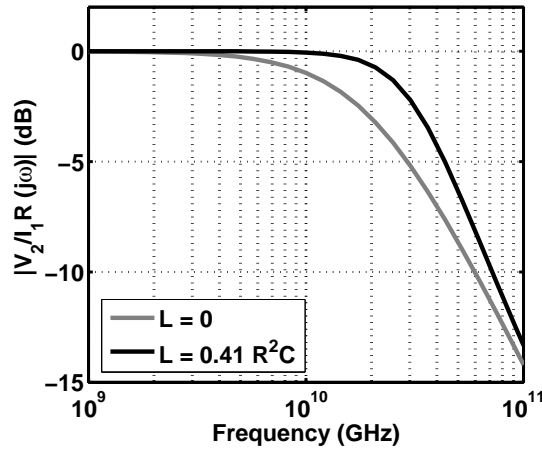
Figure 2.9: Frequency response without and with shunt peaking.

Figure 2.9 plots the magnitude response of $\dfrac{V_2}{I_1 R}(j\omega)$, for $R = 200\ \Omega$, and $C = 40$ fF, without (in gray) and with shunt peaking (in black, with $m = 0.41$), indicating the 73% bandwidth extension with an ideal inductor. Typical monolithic inductor characteristics (including inductor parasitics) limit this bandwidth improvement to approximately 50% [7].

2.3.3 Inductive Series Peaking

When the parasitic capacitance at the output of CML stage is much smaller or much larger than the load capacitance it is driving, series peaking can be

used for bandwidth extension. Figure 2.10(a) depicts a CML stage employing series peaking, where C_1 denotes the parasitic capacitance of the transistors in the CML stage, and C_2 denotes the input capacitance of the next stage. In the equivalent circuit shown in Fig. 2.10(b), series peaking produces useful results only if $C_1 < C_2$. Note that the circuit in Fig. 2.10(b) is reciprocal, and hence, it is possible to make sure that this condition is always satisfied. In other words, $\frac{V_2}{I_1}(s)$ in Fig. 2.11(a), is exactly equal to $\frac{V_1}{I_2}(s)$ in Fig. 2.11(b), by reciprocity. Thus, based on whether the parasitic capacitance of the stage (C_{par}) is smaller or larger than the load capacitance (C_L), we use either of the two configurations in Fig. 2.12, such that the smaller capacitance is always in parallel to R, i.e. $C_1 < C_2$.
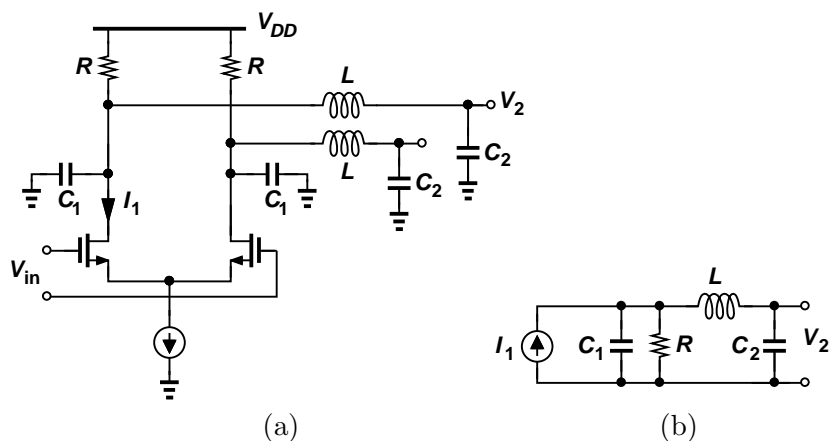
Figure 2.10: (a) Inductive series peaking, (b) equivalent circuit.
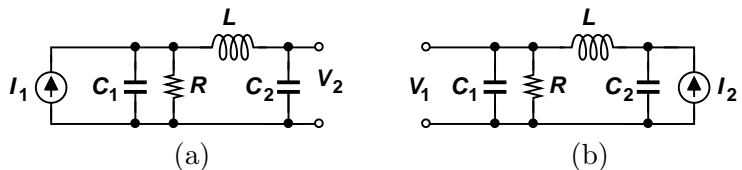
Figure 2.11: Reciprocity in a series-peaking circuit.

To intuitively understand the operation of the circuit, let's look at Fig. 2.11(b), where we can view the inductor as a switch, as we did for shunt peaking. Initially,
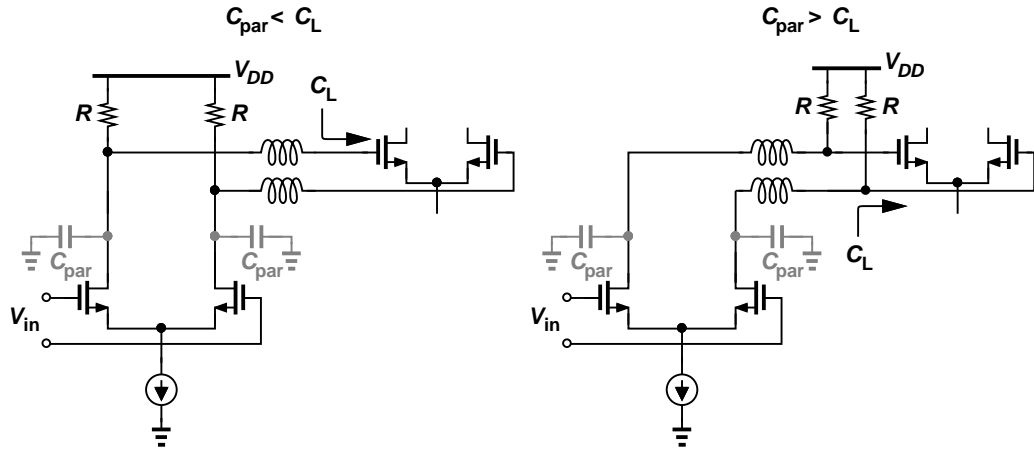
Figure 2.12: Series-peaking configurations.

all of the current flows through C_2, charging it linearly, after which C_1 is charged exponentially. Thus, if L were replaced by a ideal switch as in Section 2.3.1, the time it would take for V_1 to reach its final value can be approximately given as $RC_2 + 5RC_1$, which is much smaller than $5R(C_1 + C_2)$ when $L = 0$, thus indicating an improvement in bandwidth.

The value of this inductor can be chosen according to the equation [14]

$$L = mR^2(C_1 + C_2), \tag{2.9}$$

where m is typically chosen to be in the range 0.48 to 0.67, for optimal peaking. $m = 0.67$ gives a maximally flat amplitude response for $C_2/C_1 = 3$, while approximately doubling the 3-dB bandwidth. The choice of m is also dependent on the ratio C_2/C_1 [14].

Figure 2.13 plots the magnitude response of $\dfrac{V_2}{I_1 R}(j\omega)$, for $R = 200\ \Omega$, $C_1 = 10$ fF and $C_2 = 30$ fF, without (in gray) and with series peaking (in black, with $m = 0.67$), indicating the 100% bandwidth extension with an ideal inductor. Note that series peaking increases the roll-off rate to -60 dB/dec, because of the introduction of two additional poles.

15
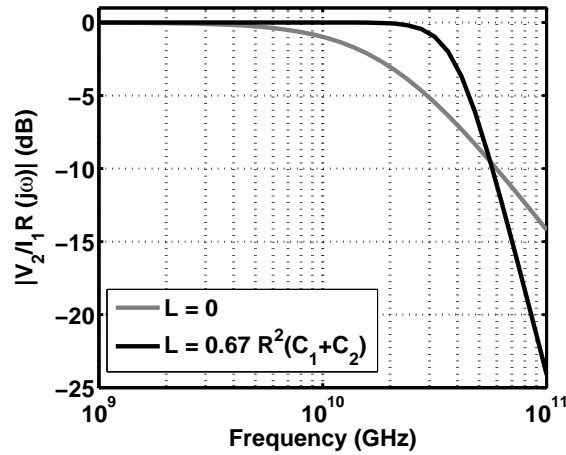
Figure 2.13: Frequency response without and with series peaking.

2.3.4   T-Coil Peaking

The bridged T-coil, often simply called the T-coil circuit uses a combination of series and shunt peaking, in addition to the mutual coupling between the inductors, to provide an extension in the 3-dB bandwidth by a factor of $2\sqrt{2} \approx$ 2.83, which is remarkably higher than series- or shunt-peaking techniques [16]. Figure 2.14 shows a common-source stage with T-coil peaking.
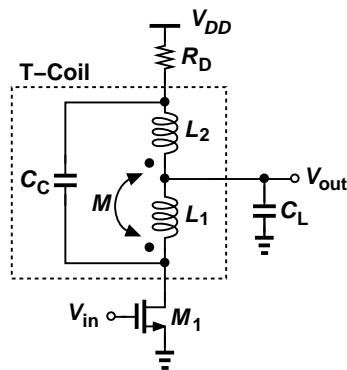
Figure 2.14: A common-source stage with T-coil peaking.

16

In Fig. 2.14, if $L_1 = L_2 = L$, and if the following two conditions hold, namely,

$$\frac{C_C}{C_L} = \frac{1}{4}\frac{1-k}{1+k}, \tag{2.10}$$

where k is the mutual coupling coefficient given by $M/\sqrt{L_1 L_2} = M/\sqrt{L}$, and

$$2C_C + C_L \frac{k}{1+k} = \frac{L(1+k)}{R_D^2}, \tag{2.11}$$

the transfer function assumes a second-order form [17]

$$\frac{V_{out}}{V_{in}}(s) = -g_m R_D \frac{\omega_0^2}{s^2 + 2\zeta\omega_0 s + \omega_0^2}, \tag{2.12}$$

where g_m is the transconductance of transistor M_1 and,

$$\omega_0^2 = \frac{2}{LC_L(1-k)} \tag{2.13}$$

$$\frac{2\zeta}{\omega_0} = C_L R_D - \frac{L(1+k)}{R_D}. \tag{2.14}$$

By setting the damping factor ζ for optimization constraints such as a maximally flat amplitude response, the other circuit parameters can be designed using the following equations [17]:

$$k = \frac{4\zeta^2 - 1}{4\zeta^2 + 1} \tag{2.15}$$

$$C_C = \frac{C_L}{16\zeta^2} \tag{2.16}$$

$$L = \frac{C_L R_D^2}{4}\left(1 + \frac{1}{4\zeta^2}\right). \tag{2.17}$$

2.3.5   Active Inductive Peaking

Inductors tend to occupy a large area on chip, and hence, if the chip area is critical, peaking can be realized by means of active circuits behaving as inductors.

Shown in Fig. 2.15(a) is a source follower circuit with a resistor R_S connected in series with the gate of M_1. For this circuit, the output impedance Z_{out} can be
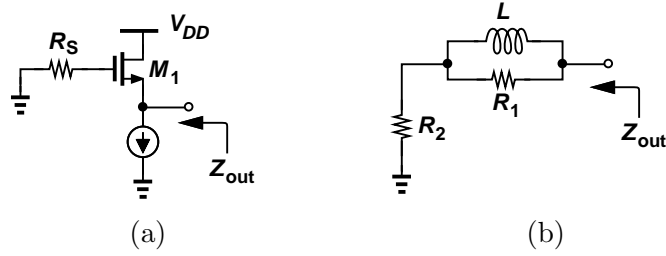
Figure 2.15: (a) Active inductor using a source follower, (b) simplified network.

written as [7]

$$Z_{out} = \frac{1 + sR_S C_{GS}}{g_m + sC_{GS}}, \tag{2.18}$$

where g_m is the transconductance of M_1, and C_{GS} is its gate-to-source capacitance. Note that $|Z_{out}(s = 0)| = 1/g_m$, and $|Z_{out}(s = \infty)| = R_S$. If $R_S > 1/g_m$, the impedance rises with frequency, thus exhibiting inductive behavior. The output impedance can be modeled as an equivalent circuit shown in Fig. 2.15(b), where

$$L = \frac{C_{GS}}{g_m}\left(R_S - \frac{1}{g_m}\right) \tag{2.19}$$

$$R_1 = R_S - \frac{1}{g_m} \tag{2.20}$$

$$R_2 = \frac{1}{g_m}. \tag{2.21}$$

The quality factor, Q, of the inductor can be improved by maximizing R_1 and minimizing R_2. Q of the parallel combination of L and R_1 is given by $\frac{R_1}{\omega L} = \frac{g_m}{\omega C_{GS}}$, which makes the overall Q independent of R_S.

Another implementation of an active inductor is depicted in Fig. 2.16, using a PMOS device. Note that $|Z_{out}(s = 0)| = 1/g_m$, and $|Z_{out}(s = \infty)| \approx R_1$. Thus, this circuit can provide an inductive output impedance if $1/g_m < R_1$.

The two implementations of active inductors described here however, severely limit the voltage headroom.

18
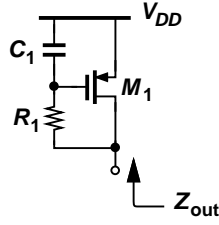
Figure 2.16: Realization of an active inductor using a PMOS device.

2.3.6 Negative Capacitance
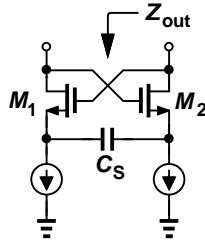
Figure 2.17: Negative capacitance circuit.

The bandwidth of a CML stage can be improved by connecting a negative capacitance circuit, shown in Fig. 2.17, to its output. This negative impedance converter transforms capacitor C_S to a negative capacitance in Z_{out}. Neglecting the gate-drain capacitances of M_1 and M_2, the output impedance can be expressed as [18]

$$Z_{out} = -\frac{1}{sC_S}\frac{g_m + s(C_{GS} + 2C_S)}{g_m - sC_{GS}}. \tag{2.22}$$

Hence, for frequencies well below f_T of the transistors, Z_{out} can be viewed as a series combination of a negative capacitance, $-C_S$, and a negative resistance, $-(C_{GS}/C_S + 2)/g_m$.

This negative capacitance circuit however, consumes headroom, because its bias current needs to flow through the drain resistors of the CML stage.

19

2.4   Bandwidth Considerations

It is known that the bandwidth of CML stages must be minimized for better noise performance. However, bandwidth limitation also results in ISI in terms of both the vertical (eye closure) and horizontal eye opening (jitter). Hence, the choice of bandwidth is based on a tradeoff between the noise performance and ISI. A typical CML stage has a dominant pole at its output. Hence, for simplicity, each stage can be modeled as a first-order RC circuit, as shown in Fig. 2.18.
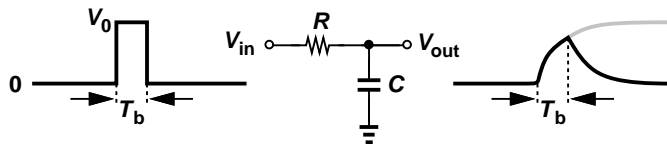
Figure 2.18: A first-order RC circuit with random input data.

Assuming that the input to this RC filter is a random bit sequence, we observe a maximum eye closure when the input consists of a number of ZEROs (or ONEs) followed by a ONE (or a ZERO), followed by a number of ZEROs (or ONEs), as depicted in Fig. 2.18. Let $R_b = 1/T_b$ denote the data rate and $f_{-3\text{dB}} = 1/(2\pi RC)$ indicate the bandwidth of the RC circuit. If the input amplitude is V_0, the output voltage can be written as

$$V_{out}(t) = V_0 \left[1 - \exp\left(\frac{-t}{RC} \right) \right]. \tag{2.23}$$

At the end of one bit period, the output voltage settles to a value

$$
\begin{aligned}
V_{out}(T_b) &= V_0 \left[1 - \exp\left(\frac{-T_b}{RC} \right) \right] \\
&= V_0 \left[1 - \exp\left(\frac{-2\pi f_{-3\text{dB}}}{R_b} \right) \right].
\end{aligned}
\tag{2.24}
$$

The ideal value that the output should have settled to is V_0. Hence, the error

is given by

$$V_0 - V_{out}(T_b) = V_0 \exp\left(\frac{-2\pi f_{-3\text{dB}}}{R_b}\right) \tag{2.25}$$

The total eye closure is twice this error, and hence can be written as

$$\text{Eye Closure} = 2V_0 \exp\left(\frac{-2\pi f_{-3\text{dB}}}{R_b}\right) \tag{2.26}$$

It is also important to examine the jitter caused by bandwidth limitation [7]. Figure 2.19 shows the slowest and the fastest rising edges at the output of the RC filter for a random input sequence. The time difference between these edges at $V_0/2$ indicates the amount of jitter.
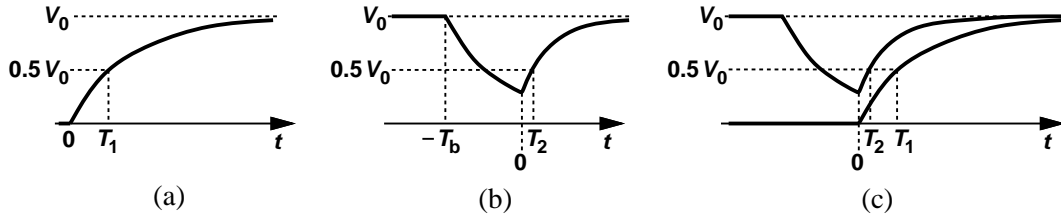
Figure 2.19: (a) Slowest rising edge, (b) fastest rising edge, and (c) calculation of jitter, for the RC filter with a random input data.

The slowest rising edge occurs when a long run of ZEROs is followed by a ONE. Shown in Fig. 2.19(a), the output voltage in this case can be written as

$$V_{out}(t) = V_0\left[1 - \exp\left(\frac{-t}{RC}\right)\right], \tag{2.27}$$

and thus, the time it takes for the output to reach $V_0/2$ is given by

$$T_1 = RC \ln 2. \tag{2.28}$$

The fastest rising edge occurs when a long run of ONEs is followed by a ZERO, which is followed by a long run of ONEs. Shown in Fig. 2.19(b), the

output voltage in this can be written as

$$
\begin{aligned}
V_{out}(t) &= V_0 \exp\left[\frac{-(t+T_b)}{RC}\right] + V_0\left[1 - \exp\left(\frac{-t}{RC}\right)\right] \\
&= V_0 - V_0 \exp\left(\frac{-t}{RC}\right)\left[1 - \exp\left(\frac{-T_b}{RC}\right)\right],
\end{aligned} \tag{2.29}
$$

and hence the time taken for the output to reach $V_0/2$ is given by

$$
T_2 = RC \ln\left[2 - 2\exp\left(\frac{-T_b}{RC}\right)\right]. \tag{2.30}
$$

The normalized jitter can be expressed as

$$
\begin{aligned}
\frac{T_1 - T_2}{T_b} &= \frac{-RC}{T_b} \ln\left[1 - \exp\left(\frac{-T_b}{RC}\right)\right] \\
&= \frac{-R_b}{2\pi f_{-3\mathrm{dB}}} \ln\left[1 - \exp\left(\frac{-2\pi f_{-3\mathrm{dB}}}{R_b}\right)\right]
\end{aligned} \tag{2.31}
$$

Table 2.1: Tradeoff between ISI and noise for different bandwidths

$\dfrac{f_{-3\mathrm{dB}}}{R_b}$	Eye Closure	Jitter	Normalized Integrated Noise
0.5	8.64%	1.41%	$\sqrt{0.5} = -3$ dB
0.7	2.46%	0.28%	$\sqrt{0.7} = -1.55$ dB
1	0.37%	0.03%	$1 = 0$ dB

Table 2.1 summarizes the eye closure, jitter and normalized integrated noise, for different bandwidths. As expected, the eye closure reduces, the jitter reduces, and the integrated noise increases, with increase in bandwidth. As an optimal choice, the bandwidth, $f_{-3\mathrm{dB}}$, is typically designed to be $0.7R_b$ [7].

2.5 Decision-Feedback Equalizers (DFE)

As mentioned before, continuous-time linear equalizers cannot correct for deep notches in frequency response. Hence, a DFE is almost always necessary to

counter reflections in the channel response. The DFE, first introduced in [19], can be thought of as an infinite impulse response (IIR) filter with a non-linear element in the loop. This should not be confused with IIR-DFEs [20, 21].[1]
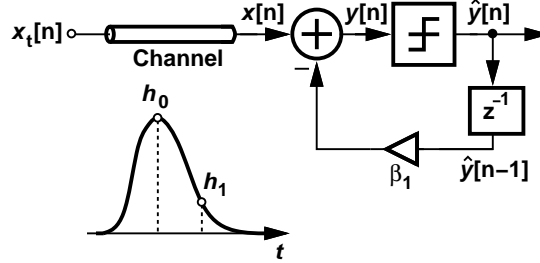
Figure 2.20: A one-tap DFE.

Shown in Fig. 2.20 is the implementation of a one-tap DFE. The slicer detects the previous bit and subtracts its effect from the incoming input signal to correctly equalize the data. Let us assume that, in discrete time, the transfer function of the channel can be written as

$$H(z) = h_0 + h_1 z^{-1} + h_2 z^{-2} + ... \tag{2.32}$$

where h_0 is called the main cursor and h_1 and h_2 are the first and the second post-cursors respectively. For now, let us assume that h_2 and higher post-cursors are zero. Thus, a discrete-time transmitted data $x_t[n]$, will appear at the output of the channel as

$$x[n] = h_0 x_t[n] + h_1 x_t[n-1] \tag{2.33}$$

The output of the DFE summer $y[n]$, can be written as

$$
\begin{aligned}
y[n] &= x[n] - \beta_1 \hat{y}[n-1] \\
&= h_0 x_t[n] + h_1 x_t[n-1] - \beta_1 \hat{y}[n-1].
\end{aligned}
\tag{2.34}
$$

---

[1]IIR-DFE is a class of DFEs that uses an IIR filter in DFE's feedback path.

23

If the DFE is operating correctly in the sense that it predicts the previous bit correctly, i.e. $\hat{y}[n-1] = x_t[n-1]$, and if we choose $\beta_1 = h_1$, then the effect of the post-cursor in the channel is completely canceled. This explanation can be extended to a continuous-time data $x(t)$ as shown in Fig. 2.21. Looking at the waveforms at the input and output of the summer, it is clear that the DFE cancels the effect of the first post-cursor. Consequently, it is possible to cancel multiple post-cursors in the channel response, by introducing multiple taps in the DFE as shown in Fig. 2.22.
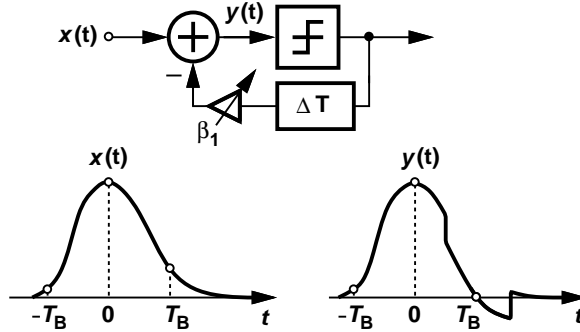
Figure 2.21: Operation of a one-tap DFE with a continuous-time input.
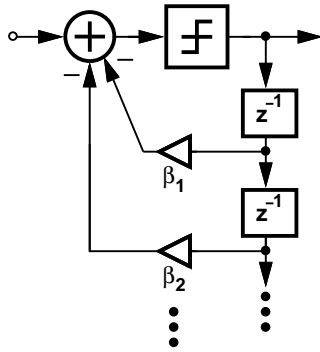
Figure 2.22: A multi-tap DFE.

DFEs have an advantage over linear equalizers in that they do not amplify noise. Let $w[n]$ be the noise waveform appearing at the input of the DFE, in Fig. 2.20. Thus, the input to the summer is modified to $h_0 x_t[n] + h_1 x_t[n-1] + w[n]$.

The output $y[n]$ can be written as

$$
\begin{aligned}
y[n] &= h_0 x_t[n] + h_1 x_t[n-1] + w[n] - \beta_1 \hat{y}[n-1] \\
&= h_0 x_t[n] + \beta_1 (x[n-1] - \hat{y}[n-1]) + w[n], \quad\quad (2.35)
\end{aligned}
$$

indicating no noise amplification.

One of the key challenges in using DFEs at high speeds is that the feedback signal must settle within one bit period. These timing constraints are discussed in Chapters 4 and 5.

CHAPTER 3

Clock and Data Recovery

A clock recovery circuit produces a clock signal from the incoming binary data stream. This chapter introduces the basics of CDR circuits, and describes various techniques used in practice.

3.1 Functions of a CDR

In a wireline system, the receiver might not have direct access to the clock of the transmitter. Hence, as mentioned in Chapter 1, these receivers need to recover clock from the received data and align its phase such that the clock samples the noisy data at its peaks, using a flipflop, as depicted in Fig. 3.1. The recovered data, also called the "retimed" data, is as clean as the recovered clock itself, thus removing the jitter (timing errors in zero crossings) of the incoming data [22].
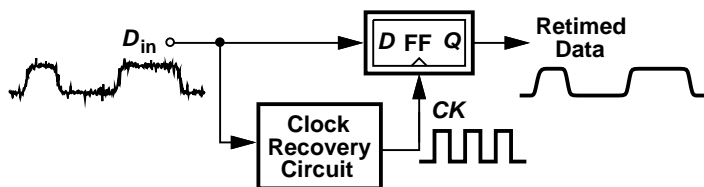
Figure 3.1: Retiming the received data using a CDR.

The clock recovery circuit of Fig. 3.1 is constructed using a phase-locked loop (PLL). As opposed to conventional PLLs which incorporate a periodic reference, the clock recovery circuit uses random data as input. Figure 3.2 shows the block

26

diagram of a clock recovery circuit consisting of a phase detector (PD) capable of measuring the phase difference of the clock and the edges of the incoming random data, a loop filter and a voltage-controlled oscillator (VCO).
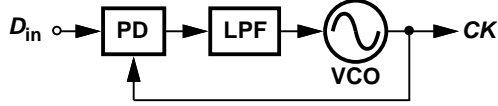
Figure 3.2: Block diagram of a clock recovery circuit.

Note that the CDR must generate the clock at the same frequency as the data rate, so that all the data bits are sampled correctly, and none are missed. This requirement is discussed in the next section.

3.2   Properties of Non-Return-to-Zero (NRZ) Data

The NRZ data shown in Fig. 3.3 does not have any spectral components at integer multiples of the data rate $R_b = 1/T_b$. This can be explained as follows. The autocorrelation function of the random binary sequence in Fig. 3.3 is given by [23]

$$R_x(\tau) = \begin{cases} 1 - \dfrac{|\tau|}{T_b}, & |\tau| \leq T_b \\ 0, & |\tau| > T_b. \end{cases}$$

Hence, its power spectral density is

$$S_x(f) = T_b \left[\frac{sin(\pi f T_b)}{\pi f T_b} \right]^2, \tag{3.1}$$

indicating no components at $f = n/T_b$, for integer values of n, as plotted in Fig. 3.4.

From another viewpoint, an NRZ sequence of data rate R_b when multiplied by $\cos(2\pi n R_b t + \phi)$, has a zero DC component, indicating that the waveform does

27
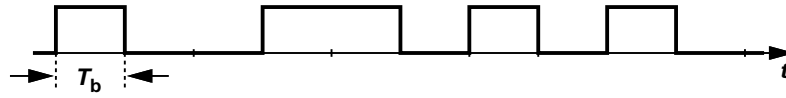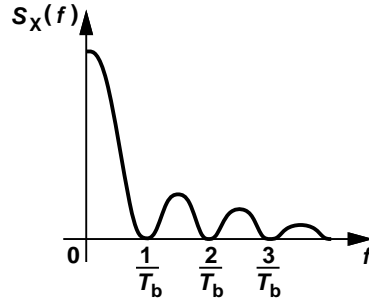
Figure 3.3: NRZ data.

Figure 3.4: Power spectral density of NRZ data.

not contain any frequency components at nR_b [22]. The absence of a spectral component at R_b, makes it difficult to recover clock from the NRZ data directly.

3.3 Edge and Phase Detection

The edges in the NRZ data can be detected (by differentiation operation), and rectified, as shown in Fig. 3.5, thus producing a waveform that has a non-zero frequency component at R_b. This can be verified by multiplying the resulting waveform with $\cos(2\pi R_b t + \phi)$, as shown in Fig. 3.6, which results in a non-zero DC component, indicating that the rectified edge waveform has a component at R_b.
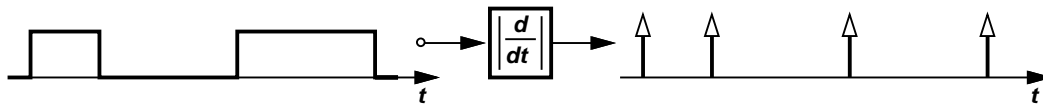
Figure 3.5: Differentiation and rectification of data.

After recovering the clock from the rectified edge detector, we must adjust
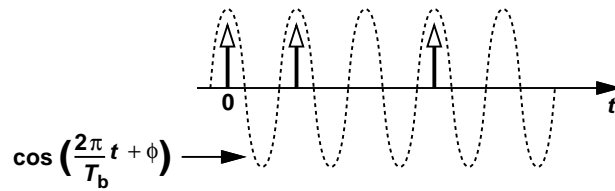
Figure 3.6: Multiplying the rectified edges by a sinuosid.

the phase of this clock with respect to data such that the clock samples the data bits at their peak values. This can be achieved by multiplying the rectified edge waveform by $\cos(2\pi R_b t + \phi)$, as shown in Fig. 3.7. The average value of the mixer output, V_{out}, indicates the relative phase difference between data and the cosine waveform. Figure 3.7 also plots the average value of V_{out} as a function of the excess phase, ϕ, in the cosine waveform, indicating zero average when the excess phase is $\pm\pi/2$. Note that at these points, the alternate zero crossings of the cosine clock waveform are aligned to the edges of the data, indicating that the clock samples the data at its peak values.
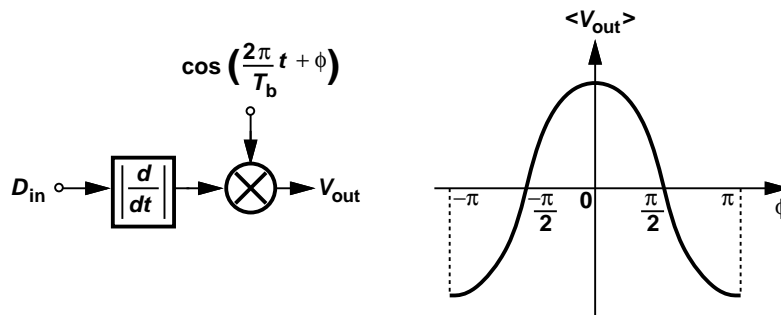
Figure 3.7: Phase detector using rectified edge detection.

The rectified edge detection can be performed by the circuit in Fig. 3.8, called the digital edge detector [22]. This circuit produces a positive pulse on each data edge. However, the delay element ΔT cannot be too large or too small [7]. If ΔT is too small, the finite bandwidth of the circuit prohibits the output of the XOR from reaching full swings. If ΔT is too large, the time overlap between the data

and its delayed version is too small, prohibiting edge detection.
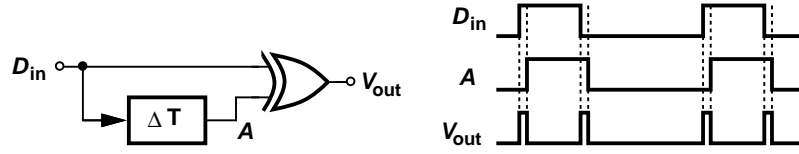
Figure 3.8: Digital edge detector.
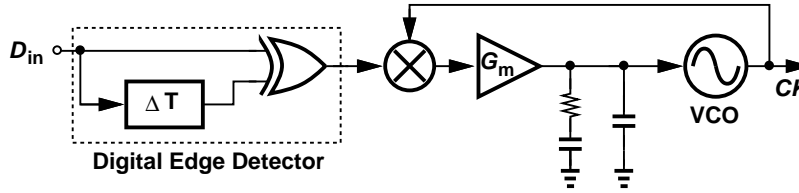
Figure 3.9: A clock recovery circuit using a digital edge detector.

Using this digital edge detector in the phase detector of Fig. 3.7, we can develop the clock recovery circuit, with a loop filter and a VCO, as shown in Fig. 3.9 [24].

3.3.1 Linear (Hogge) Phase Detector

The delay element in the digital edge detector of Fig. 3.8 may be implemented using a synchronous delay element. Figure 3.10 uses a flipflop to realize this synchronous delay [7]. However, this circuit produces an "error" signal indicating phase information between the incoming data, D_{in}, and clock only when there is a transition in the input data, and hence is input-pattern dependent. This can be observed from the waveforms in Fig. 3.10.

In order to overcome this ambiguity, another flipflop is added in cascade in Fig. 3.11, creating fixed-width "reference" pulses, which indicate the presence or absence of input transitions, thus eliminating data dependence from the phase detector output. This circuit is known as the "Hogge phase detector" [25]. Note

30
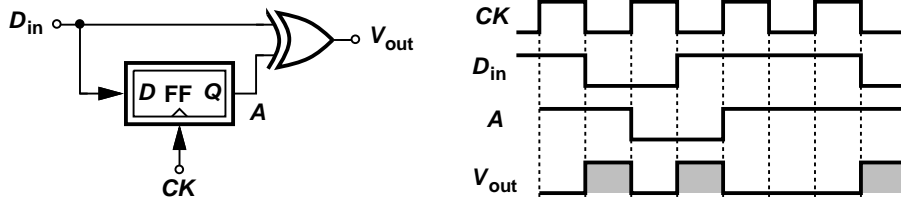
Figure 3.10: Digital edge detector using a synchronous delay element.
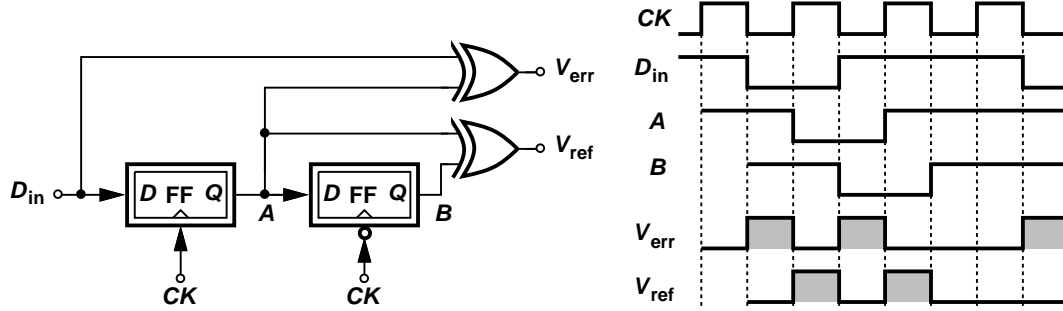
Figure 3.11: Hogge phase detector.

that the retimed data is available at both A and B, without the need for the additional flipflop of Fig. 3.1.
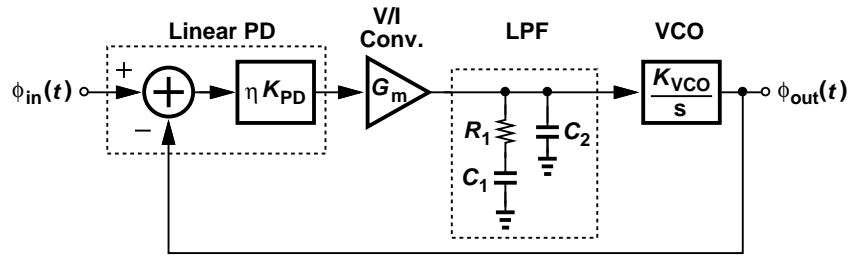
Figure 3.12: Model of a CDR employing a linear phase detector.

A CDR employing the Hogge phase detector can be modeled as shown in Fig. 3.12 similar to a type-II PLL [26]. The density of transitions in the input data is modeled by an activity factor, η, in the phase detector, where $0 \leq \eta \leq 1$. The loop bandwidth of the CDR is given by

$$\omega_{-3\text{dB}} = \eta R_1 G_m K_{PD} K_{VCO} \left(\frac{b-1}{b} \right),$$ (3.2)

31
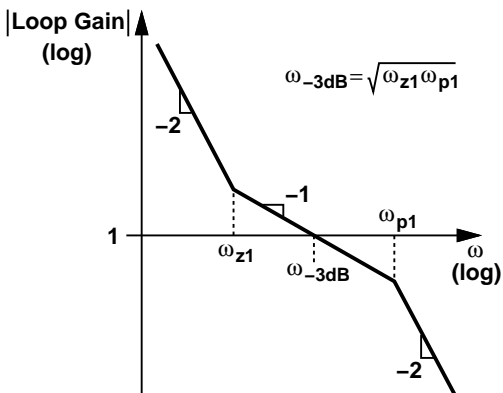
Figure 3.13: Magnitude plot of the loop gain.

where

$$b = 1 + \frac{C_1}{C_2}. \tag{3.3}$$

As depicted in Fig. 3.13, if the loop bandwidth is designed to be the geometric mean of the loop-filter zero, $\omega_{z1} = 1/(R_1 C_1)$, and the loop-filter pole, $\omega_{p1} = b/(R_1 C_1)$, i.e. $\omega_{-3\text{dB}} = \sqrt{b}/(R_1 C_1)$, the phase margin of the system equals

$$PM = \tan^{-1} \left[\frac{1}{2} \left(\sqrt{b} - \frac{1}{\sqrt{b}} \right) \right]. \tag{3.4}$$

The parameter b is usually chosen to be 16 or 25 to obtain phase margins of 62° and 67°, respectively.

3.3.2 Bang-bang (Alexander) Phase Detector

Another way of obtaining the relative phase information is to use the clock to sample the data at multiple points in the vicinity of expected transitions.

Figure 3.14 explains the principle of Alexander phase detection [27]. This technique is also called "early-late" detection [7]. The clock samples the data at three points, S_1, S_2 and S_3, by means of four flipflops. Note that the last flipflop on the bottom path may be replaced by a positive level-triggered latch.
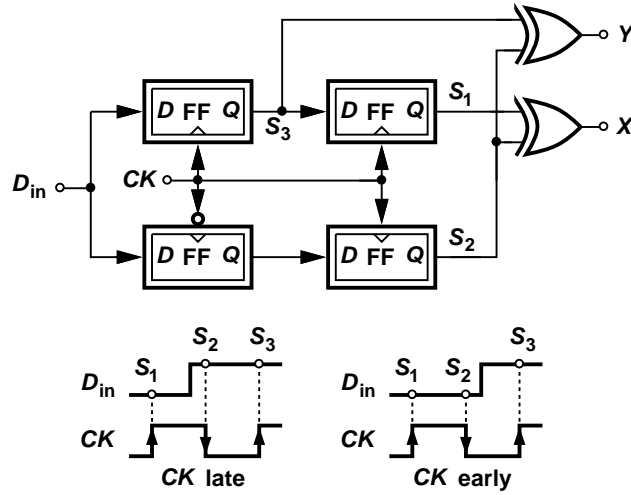
Figure 3.14: Alexander phase detector.

By taking the following XORs, $S_1 \oplus S_2$ and $S_2 \oplus S_3$, we determine whether the clock is early or late with respect to the data, as illustrated in Fig. 3.14: (a) If $S_1 \oplus S_2 = S_2 \oplus S_3$, no data transition is present, (b) if $S_1 \oplus S_2$ is high and $S_2 \oplus S_3$ is low, the clock is late, and (c) if $S_1 \oplus S_2$ is low and $S_2 \oplus S_3$ is high, the clock is early.
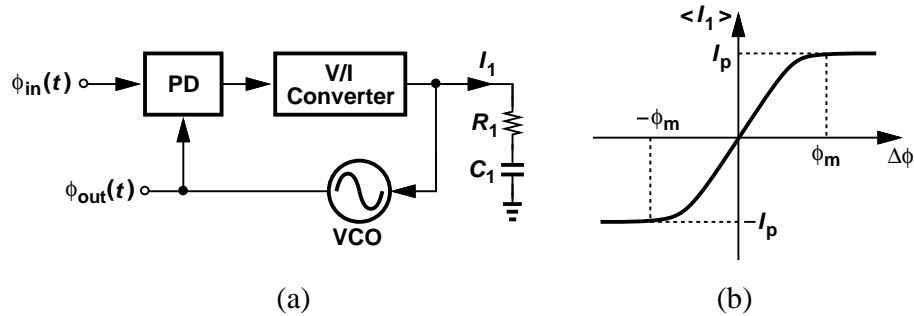
Figure 3.15: (a) Bang-bang CDR model and (b) the corresponding phase-detector characteristics.

Thus, this phase detector only determines whether the clock is early or late, leading to its bang-bang characteristics, thus resulting in a high gain region in the vicinity of zero phase difference. Consequently, a CDR employing the Alexan-

der phase detector locks such that S_2 is aligned with the data transitions. Note that the non-linear nature of the phase detector does not agree with the model in Fig. 3.12. Figure 3.15(a), depicts the model of a bang-bang CDR [28].[1] Figure 3.15(b) shows the corresponding phase-detector characteristics. The CDR's loop bandwidth is approximately given by [28]

$$\omega_{-3\text{dB}} = \frac{\pi K_{VCO} I_p R_1}{2\phi_{in,p}},\tag{3.5}$$

where $\phi_{in}(t) = \phi_{in,p} \cos \omega_\phi t$ models the input sinusoidal jitter.[2] The closed-loop jitter transfer can be approximated as [28]

$$\frac{\phi_{out,p}}{\phi_{in,p}}(s) = \frac{1}{1 + \dfrac{s}{\omega_{-3\text{dB}}}},\tag{3.6}$$

where $\phi_{out,p}$ is the peak value of $\phi_{out}(t)$.

Note that with the linear or bang-bang detectors, the loop gain is dependent on the density of data transitions as the control line of the VCO remains idle in the absence of data edges [7].

3.4 Half-Rate Phase Detectors

At high data rates, the CDR circuits may employ a VCO operating at half the data rate, as it is difficult to design oscillators at full rate that provide good phase noise performance. In addition, the clock buffers and routing may become power-hungry at full rate. Also, a half-rate operation might be preferred if it relaxes the speed requirements of some blocks in the receiver, such as phase detectors,

---

[1]Note that second capacitor in the loop filter is not included for simplicity of analysis.

[2]Equation (3.5) assumes that the phase detector operates every clock cycle. To account for the absence of data edges in the input, this expression may be multiplied by the activity factor, η, as we did for the linear-PD case.

frequency dividers, etc. Thus, it might be required to have a "half-rate" CDR that operates with a full-rate input data, but with a half-rate clock.

3.4.1 Half-Rate Linear Phase Detector

Figure 3.16 shows the implementation of a linear phase detector working at half rate [29]. The input full-rate stream, D_{in}, is first demultiplexed into two half-rate streams at A and B using two latches, L_1 and L_2. The XOR of these half-rate streams at A and B generates error pulses indicating the phase difference between the received data and the half-rate clock, as illustrated in the waveforms in Fig. 3.16. However, these pulses are present only when the full-rate input has transitions, and hence are data-dependent. To remove this data dependence from the phase-detector output, two more latches, L_3 and L_4, are added to produce reference pulses, similar to the full-rate linear PD in Section 3.3.1.
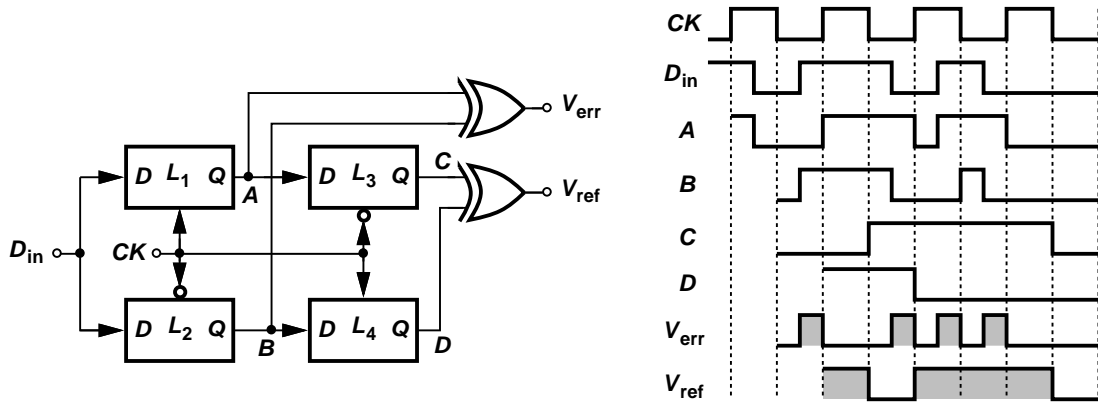
Figure 3.16: Half-rate linear phase detector.

If the clock samples the data at the center of its bit period, the error pulses would be exactly half as wide as the reference pulses. Hence, the half-rate phase detector output is taken as the average value of $2V_{err} - V_{ref}$. The CDR loop can be designed similar to the one with a full-rate linear phase detector.

3.4.2 Half-Rate Bang-Bang Phase Detector

Figure 3.17 depicts the implementation of a half-rate bang-bang (or "early-late") phase detector [30]. Three flipflops employ quadrature clocks to obtain three samples of the full-rate data near the transitions, as shown in Fig. 3.17. Similar to the full-rate version in Section 3.3.2, the following XORs, $S_1 \oplus S_2$ and $S_2 \oplus S_3$, are taken to determine whether the clock is early or late with respect to the data. The rising edge of CK_Q occurs in the vicinity of the data transitions under locked condition.
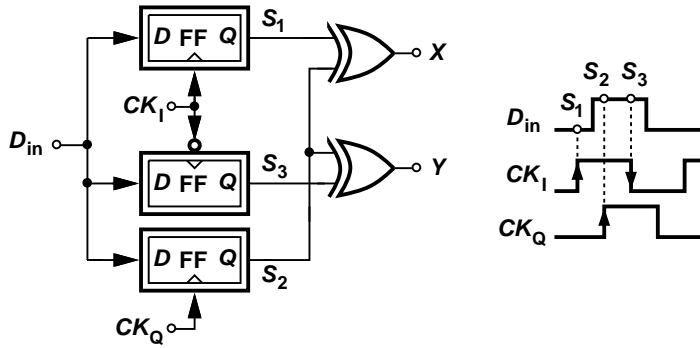
Figure 3.17: Half-rate bang-bang phase detector.

Note that this topology, unlike the linear half-rate detector, requires a quadrature VCO. For a given power consumption, the phase noise of quadrature VCOs is typically higher than a single oscillator [31][3], because: (a) The flicker noise of the coupling transistors degrade the phase noise at low frequency offsets, (b) the quality factor of the tank is reduced as oscillation departs from the resonance frequency [15]. Moreover, quadrature LC VCOs need two inductors, which can occupy a large area on chip. Also, at high speeds, it is more difficult to maintain perfectly quadrature phases (as opposed to complementary phases) in the clock

---

[3]In [31], the phase noise of the 60-GHz quadrature VCO is about 5 dB higher than a single oscillator at 1-MHz offset.

distribution network.

3.5 Jitter Tolerance in CDR circuits

We recognize that for a slowly varying jitter at the input, which lies within the loop bandwidth of the CDR, the CDR tracks the phase variations, thus ensuring that the data is always sampled in the middle of the eye. For input jitter frequencies which lie outside the loop bandwidth, the CDR cannot track the variations fully. Jitter tolerance specifies the maximum allowable jitter on the received signal that can be tracked by the CDR without increasing the bit error rate. The jitter tolerance specification is typically described in terms of a mask shown in Fig. 3.18 [32].[4] In order to meet the specification, the jitter tolerance curve of the CDR must lie above the mask.
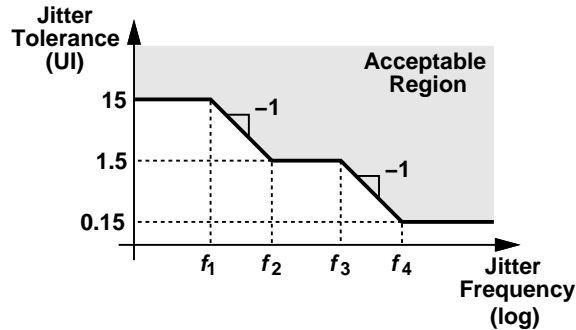
Figure 3.18: Example of jitter tolerance mask.

In order to measure the jitter tolerance at a given frequency, we keep increasing the peak-to-peak jitter at the input, until the BER begins to rise. This condition happens when the phase difference between the input excess phase, ϕ_{in}, and the recovered clock's excess phase, ϕ_{out}, approaches 0.5 UI. In actual

---

[4]For SONET OC192, $f_1 = 2.4$ kHz, $f_2 = 24$ kHz, $f_3 = 0.4$ MHz, and $f_4 = 4$ MHz, at nearly 10-Gb/s data rate.
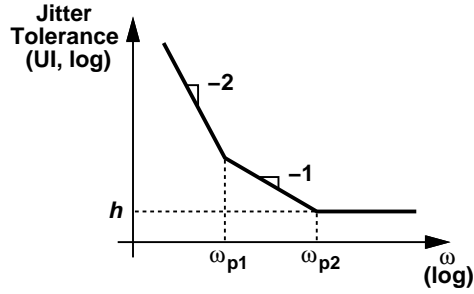
Figure 3.19: Jitter tolerance of a second-order CDR loop.

circuits, this value might be lower than 0.5 UI, and can be measured by plotting the bathtub curve. The bathtub measurement is discussed in Chapters 4 and 5. Let us assume that the peak horizontal eye opening in the bathtub curve is h UI. Then, the condition to avoid increasing the BER is [7]

$$\phi_{in} - \phi_{out} < h \tag{3.7}$$

$$\implies \phi_{in}[1 - H(s)] < h, \tag{3.8}$$

where $H(s) = \phi_{out}/\phi_{in}$ is the jitter transfer. Therefore,

$$\phi_{in} < \frac{h}{1 - H(s)}. \tag{3.9}$$

Hence, jitter tolerance can be expressed as

$$G_{JT}(s) = \frac{h}{1 - H(s)}. \tag{3.10}$$

For a second-order CDR loop [7],

$$H(s) = \frac{2\zeta\omega_0 s + \omega_0^2}{s^2 + 2\zeta\omega_0 s + \omega_0^2}. \tag{3.11}$$

Thus, from Eq. (3.10),

$$G_{JT}(s) = h\frac{s^2 + 2\zeta\omega_0 s + \omega_0^2}{s^2}. \tag{3.12}$$

Plotted in Fig. 3.19, ω_{p1} and ω_{p2} represent the poles of $H(s)$. Note that ω_{p2} represents the loop bandwidth of the CDR. Thus, the jitter tolerance is constant ($= h$ UI) for frequencies above ω_{p2}.

CHAPTER 4

A 32-Gb/s 9.3-mW CMOS Equalizer with 0.73-V Supply

This chapter describes the design of a full-rate equalizer operating at 32 Gb/s with a 0.73-V supply. Employing a CTLE and a one-tap DFE, the circuit draws upon two ideas, namely, nested inductors and latches with inductive feedforward, to achieve a power consumption of 9.3 mW in 45-nm CMOS technology.

4.1 Background

At 32 Gb/s, the unit interval (UI) of 31.25 ps poses critical challenges in the design of the DFE loop. Shown in Fig. 4.1 are a full-rate direct DFE, a half-rate direct DFE and a full-rate loop-unrolled DFE [33]. For a full-rate direct DFE, the constraint on the loop delay is given by

$$t_{CQ} + t_{setup} + t_{FB} < 1 \text{ UI}, \qquad (4.1)$$

where t_{CQ} is the clock-to-Q delay of the flipflop, t_{setup} is the setup time of the flipflop and t_{FB} is the "feedback delay" arising from the time constant at the summing node. One disadvantage of this topology is that the design of the clock buffer can be difficult and/or power-hungry at full rate.

A half-rate DFE still has the same timing constraint but it simplifies the design of the clock buffer. Unfortunately, the transconductances and latches
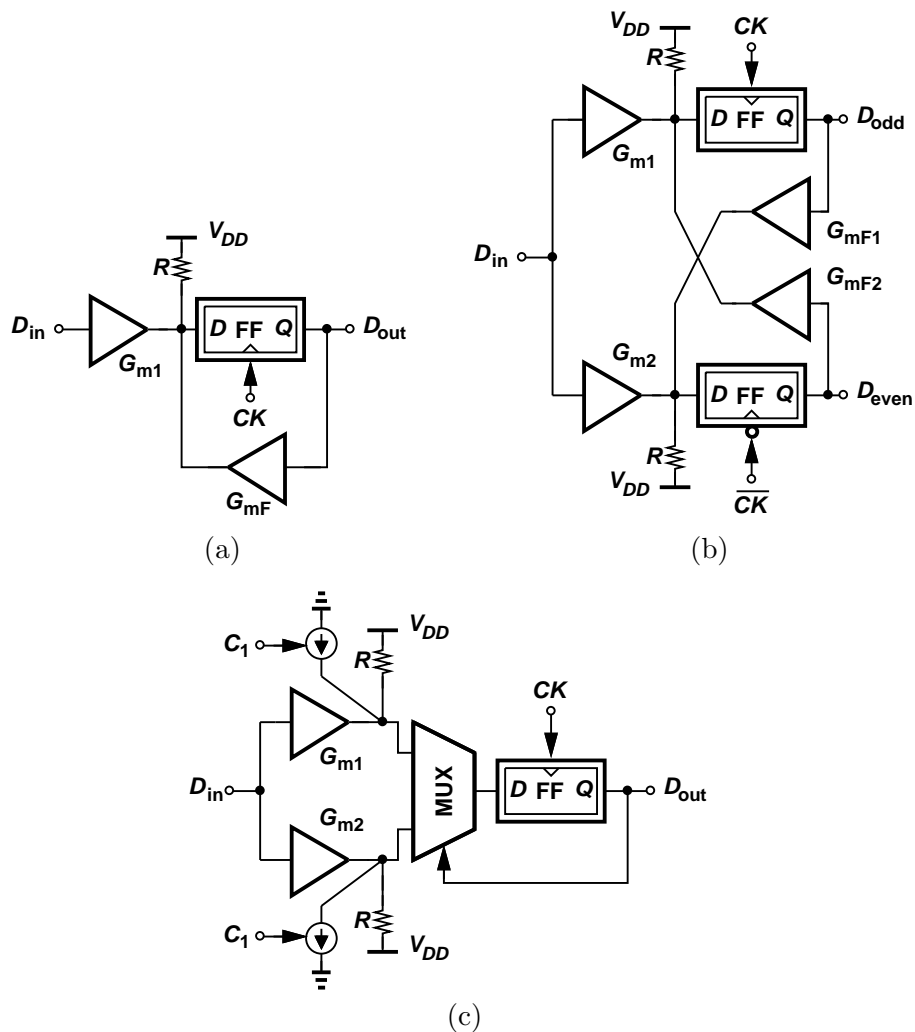
Figure 4.1: DFE architectures: (a) Direct full-rate DFE, (b) direct half-rate DFE, (c) unrolled full-rate DFE.

must still be as fast, and we have twice as many here. In other words, it makes sense to always start with a full-rate architecture, unless it turns out to be too difficult.

The timing constraint is roughly the same for direct DFE (full-rate or half-rate), and unrolled DFE [3]. The unrolled topology replaces the settling time at the summing junction with the propagation delay through the multiplexer, t_{MUX},

but it requires that the *data* perform the multiplexing operation and hence have sufficiently large voltage swings. This issue makes unrolled DFEs less attractive at low supply voltages.
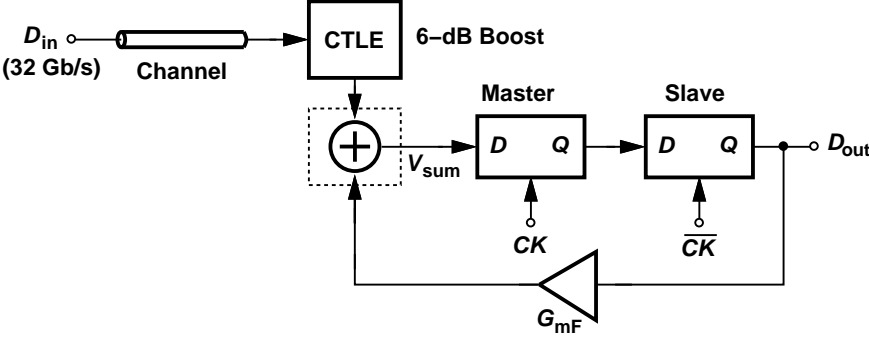
Figure 4.2: Full-rate data path.

In order to process a data rate of 32 Gb/s in 45-nm technology, the stages in the signal path shown in Fig. 4.2 must incorporate inductive peaking, potentially occupying a large area. Moreover, this approach does not adequately reduce the feedback delay, still producing a relatively narrow eye at the summing junction. Figure 4.3 shows the simulated eye diagram for a simple DFE loop using CML stages without and with inductive peaking (including layout parasitics) with a channel loss of 18 dB at 16 GHz, suggesting that new measures are necessary to ensure a more robust operation. The circuit techniques presented in this chapter address these issues.

It is worth noting that the speed of inductively-peaked CML latches can be improved if the value of their load resistors is decreased. This approach also reduces the voltage headroom consumption, facilitating operation with low supplies. However, latches using primarily inductive loads (with a small load resistance), as in [34], fail to retain a correct output in the presence of long runs because they allow their differential outputs to collapse to the common-mode level. Simulations suggest that a DFE loop using such stages suffers from enormous ISI and

(a)   (b)

(c)
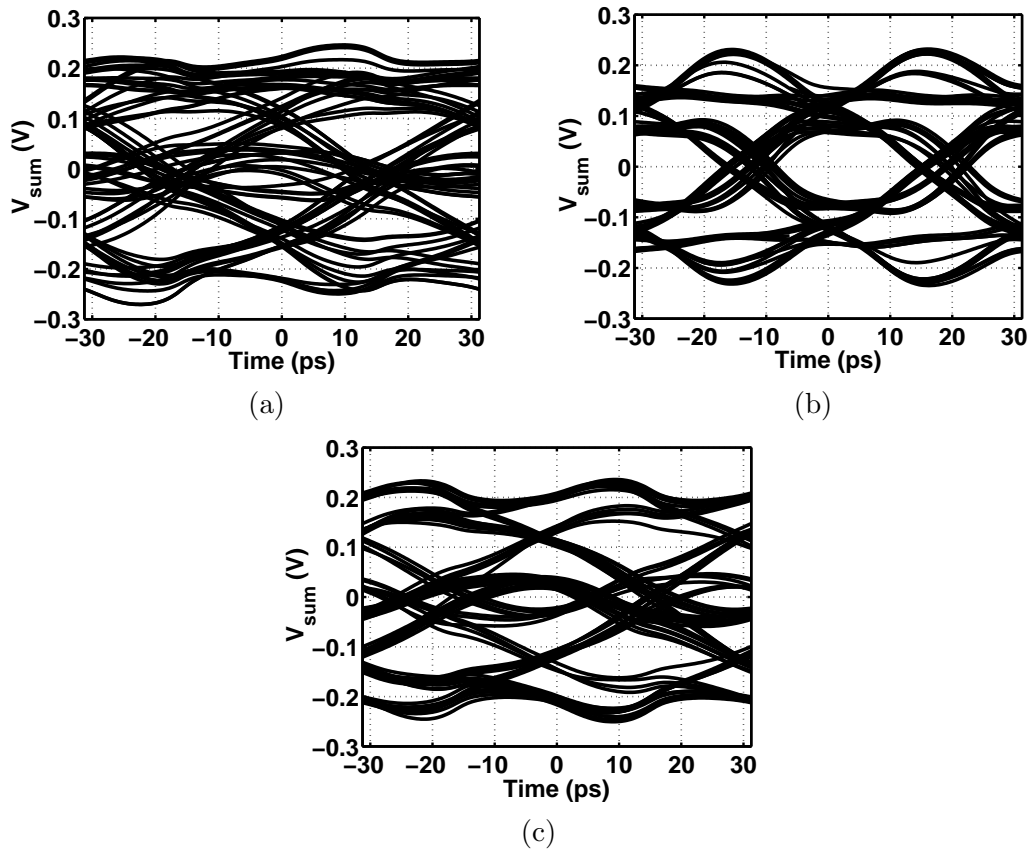
Figure 4.3: Eye diagram at the summing junction of DFE loop using CML stages (a) without inductive peaking, (b) with inductive peaking, (c) with an inductive load in latches (with a 30-Ω series resistance).

may even oscillate. Even if we add a small series resistance to the inductors, we observe that the latches fail to retain the correct output in the presence of long runs because of a very low dc gain, as shown in Fig. 4.3(c).

4.2   Equalizer Architecture

Figure 4.4 shows the proposed equalizer architecture. In order to save chip area, the design "nests" the load inductors used in the CTLE and the DFE summer, thus creating mutual coupling between the two stages. As explained in Sec-

tion 4.3, this effect can be exploited to further reduce the area.
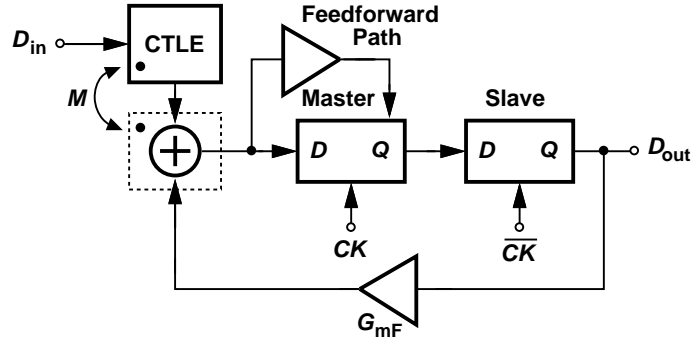
Figure 4.4: Proposed equalizer architecture.

The architecture of Fig. 4.4 also incorporates a feedforward path around the master latch so as to reduce the loop delay. Feedforward is particularly effective here as this latch senses quite smaller voltage swings than those it applies to the slave latch.

The equalizer's power dissipation can be reduced through the use of "linear scaling" [3]; i.e., the width and bias current of all transistors can be scaled down by a factor of α and the load resistors scaled up by the same factor. The scaling, however, eventually faces two issues because the inductors must also scale up by a factor of α: (1) The area penalty becomes significant, and (2) the large inductor parasitics degrade the speed. The design reported here makes a compromise between the power dissipation and these two drawbacks.

4.3   Design of Building Blocks

This section presents the circuit-level implementation of the equalizer building blocks in 45-nm CMOS technology. The inductors used in this work have been designed and simulated in Ansys's HFSS and ported into Cadence as S-parameter-

based models that are extracted for a frequency range of near DC to 60 GHz.

4.3.1   CTLE and DFE Input Stage

In a full-rate equalizer, the CTLE and the input stage of the DFE can be designed as one entity so as to improve the performance. In the prototype reported here, these two blocks consume about 44% of the total power, warranting careful optimization of their performance.

Figure 4.5(a) shows the realization of the two stages. In addition to the CTLE, the DFE input also includes a zero so as to raise the high frequency boost [35]. Located at 4 GHz and 4.7 GHz, respectively, these zeros provide a maximum boost of about 6 dB at 20 GHz [gray plot in Fig. 4.5(b)].

The first DFE tap is a differential pair whose current can be programmed. In order to vary the tap-coefficient, bits b_0-b_3 in Fig. 4.6 are set in binary fashion. The value of the tail current source can thus be varied from 0 to 1.5 mA in 16 steps of 100 µA. To turn a current source on or off, the gate is switched between the bias voltage or ground, respectively. The switches are of minimum size.

While essential to operation at 32 Gb/s, the two differential load inductors in Fig. 4.5(a) (along with inductors in latches) can occupy a large area as shown in Fig. 4.7(a). The two structures can be stacked as shown in Fig. 4.7(b), but the large capacitance between the two spirals would severely limit the bandwidth. Instead, we nest L_1 and L_2 as shown in Fig. 4.8(c), facing two issues. First, since $L_1 \approx L_2$, one inductor must be designed with a larger number of turns so as to fit in a smaller diameter. Second, the coupling factor (≈ 0.3) between L_1 and L_2 alters the behavior of both stages by both feedforward and feedback. This phenomenon can be studied with the aid of the simplified model depicted in Fig. 4.9. We wish to determine the transfer function from V_{in} to V_{out2}. We have

44
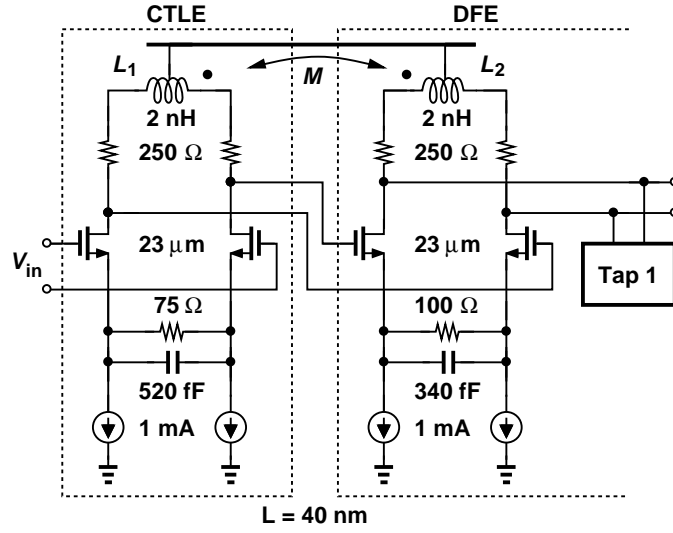
(a)
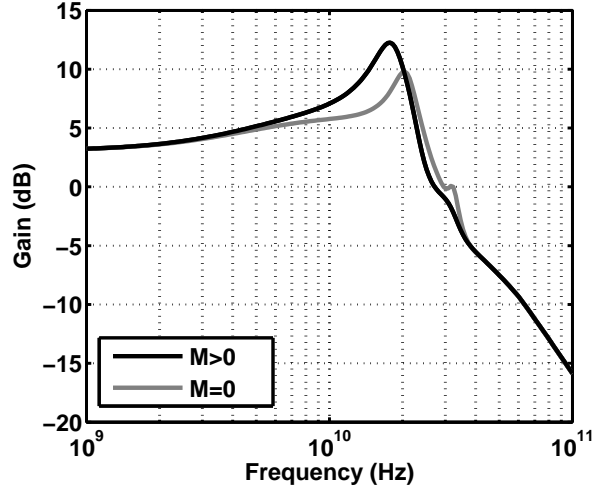
(b)

Figure 4.5: CTLE and input stage of the DFE: (a) Circuit diagram, (b) frequency response with and without mutual coupling.

$$V_{out1} = G_{m1}V_{in}(2R_1 + j\omega L_1) + j\omega M G_{m2}V_{out1} \tag{4.2}$$

$$V_{out2} = G_{m2}V_{out1}(2R_2 + j\omega L_1) + j\omega M G_{m1}V_{in}. \tag{4.3}$$

Further manipulating these equations and assuming $\omega^2 M^2 G_{m2}^2 \ll 1$ and $2R_1 \gg \omega^2 L_1 M G_{m2}$ for $\omega \leq 16$ GHz, we obtain,

45
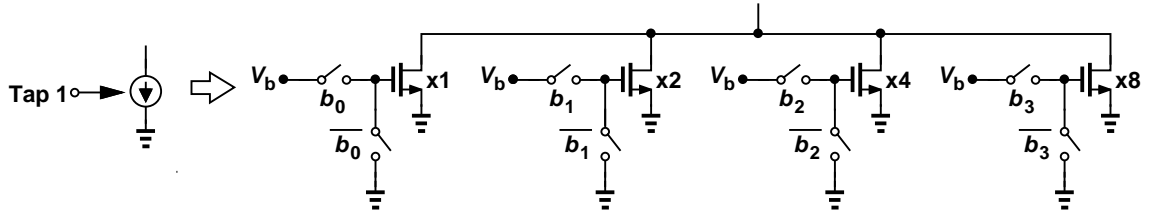
Figure 4.6: DFE tap control.
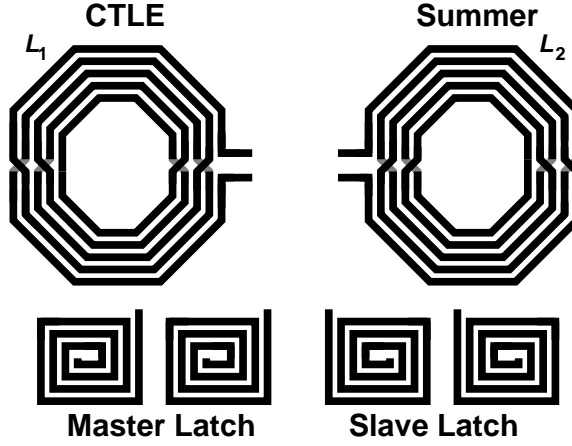
(a)
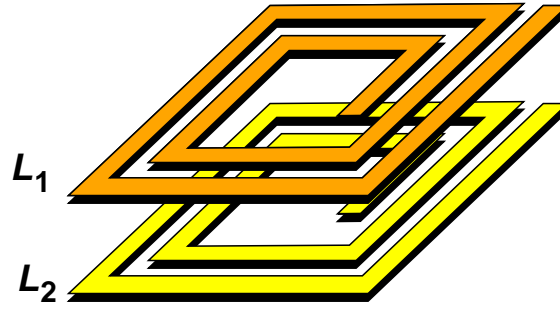
(b)

Figure 4.7: (a) Floor plan with differential load inductors of CTLE and DFE input-stage, and inductors in latches, (b) Stacking of L_1 and L_2.

$$\frac{V_{out2}}{V_{in}}(j\omega) \approx 4G_{m1}G_{m2}R_1R_2 +$$
$$j\omega G_{m1}G_{m2}(2R_1L_{eff,2} + 2R_2L_{eff,1}), \qquad (4.4)$$
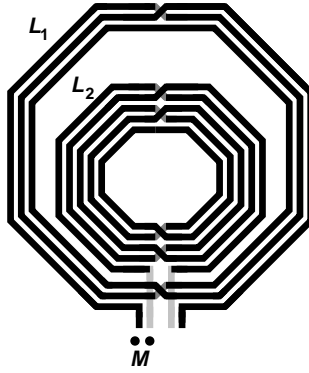
Figure 4.8: Nesting of L_1 and L_2.
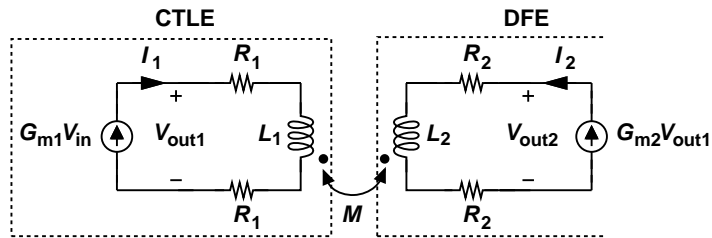
Figure 4.9: Simplified model to study nested inductors.

where $L_{eff,1} = L_1 + 2MG_{m2}R_1$, $L_{eff,2} = L_2 + M/(2G_{m2}R_1)$.

Thus, in the presence of mutual coupling, the zero caused by inductive peaking moves to a lower frequency and the magnitude of the boost increases. Simulation of the actual circuit with various capacitances confirms this result [black plot in Fig. 4.5(b)]. The apparent increase in the value of inductors further reduces the area.

In order to analyze the effect of nesting inductors on feedback G_m, we apply a step at its input and view the step response at the summing node as shown in Fig. 4.10(a). Figure 4.10(b) shows the step response with and without mutual coupling. In the presence of mutual coupling, the output voltage rises faster and has a bigger overshoot, indicating increase in the high-pass boost. This effect could slightly improve the rise and fall times but the result is not significant in

this design.
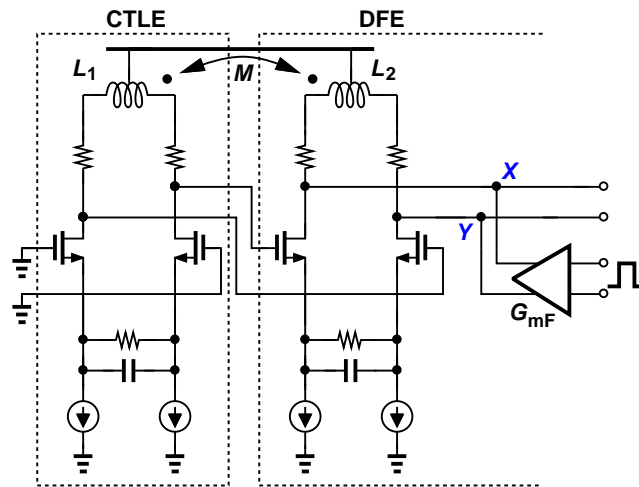
(a)
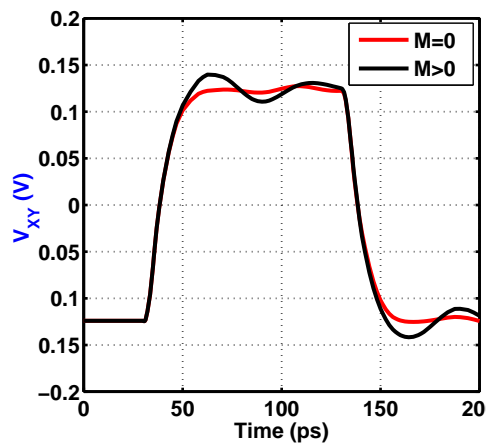
(b)

Figure 4.10: Effect of nesting on feedback G_m: (a) Circuit diagram, (b) step response.

4.3.2 Latch with Feedforward

We face three issues in our latch design: First, it is difficult to use rail-to-rail swings at these speeds, and hence CML stages are preferable. Second, as seen from Fig. 4.3, inductive peaking seems inevitable for CML stages. Third, if we

plan to use inductors, how do we accommodate so many inductors in our layout?

It is possible to improve the speed of latches by means of feedforward [36]. Particularly suited to low-voltage implementations is feedforward to the load inductors as it entails no headroom penalty. Shown in Fig. 4.11(a), the master latch employs the unclocked differential pair, M_3-M_4, as the feedforward path, allowing the data to propagate towards the output before the main pair, M_1-M_2, is clocked. The injection of this "early" signal to X and Y (rather than to P and Q) avoids additional IR drops, but it also creates a high-pass response. Fortunately, this effect is desirable in a DFE environment for the components that must be fed forward indeed lie at high frequencies.
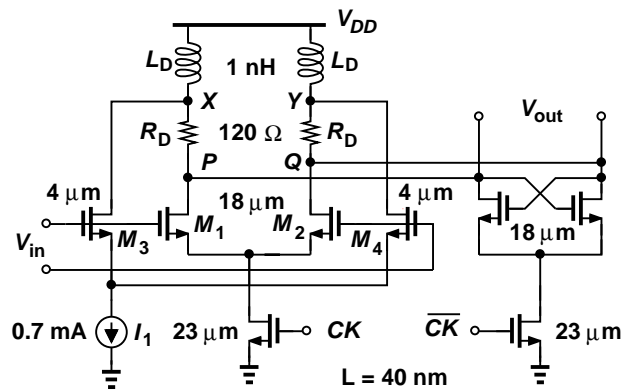
(a)
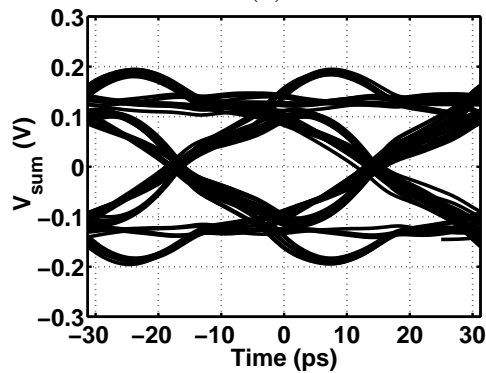
(b)

Figure 4.11: (a) Master latch with feedforward, and (b) simulated eye diagram at the summing junction of DFE loop using feedforward.

The high-pass behavior of the above feedforward path also affects the nature of the data processed by the master latch. This attribute can be analyzed by modeling the high-pass transfer function as ks and constructing the equivalent circuit shown in Fig. 4.12(a), where $A_0 \approx g_{m1,2}R_D$. Factoring out the composite transfer function, we arrive at the system shown in Fig. 4.12(b), recognizing that latch feedforward is equivalent to some boost in both the data path and the feedback path. This is an interesting departure as conventional DFEs assume a flat frequency response for the feedback G_m.
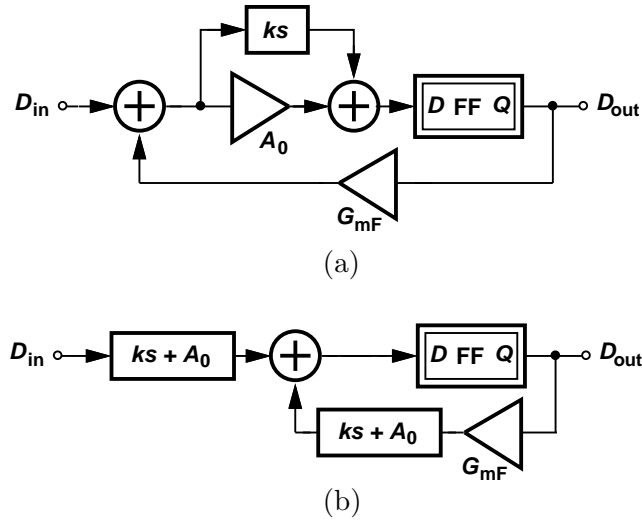
Figure 4.12: (a) Equivalent circuit for analyzing feedforward, and (b) modified equivalent circuit.

The relative strengths of the feedforward and main paths must be chosen carefully. If excessive, the former produces heavy ringing and hence intersymbol interference. Also, since the feedforward path remains on even during the regeneration phase, it can corrupt the stored bit by the new input bit. In this design, the ratio is about 1 to 5.

Figure 4.11(b) repeats the simulations leading to Fig. 4.3 but with the above feedforward method applied. We observe considerable reduction in the jitter.

The nesting of inductors can also be considered for those in the master and slave latches. However, since the two latches operate on successive bits, the coupling between the inductors would result in ISI. Hence, the two latches' inductors are realized as four independent, single-ended metal-5 to metal-9 stacked structures, each occupying an area of 25 μm × 25 μm, as shown in Fig. 4.13.
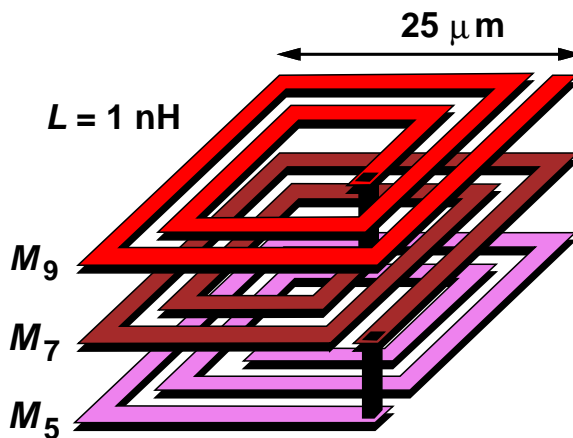
Figure 4.13: Implementation of an inductor as a stacked structure, used in latches.

4.4 Experimental Results

The equalizer has been fabricated in TSMC's 45-nm CMOS technology. Shown in Fig. 4.14, the active area of the die measures about 200 μm × 340 μm. The prototype has been mounted directly on a printed-circuit board and the high-speed signals are carried through probes. The circuit operates robustly from 1 Gb/s to 32 Gb/s with a supply voltage ranging from 1.2 V to 0.73 V, at which the results reported below have been measured. Such a low supply voltage is possible because all of the stages have at most two stacked transistors. The CTLE draws 1.46 mW, the summer 2.59 mW, and the two latches 5.27 mW.

Figure 4.15 shows the test setup to measure the BER. Four 8-Gb/s pseudo-
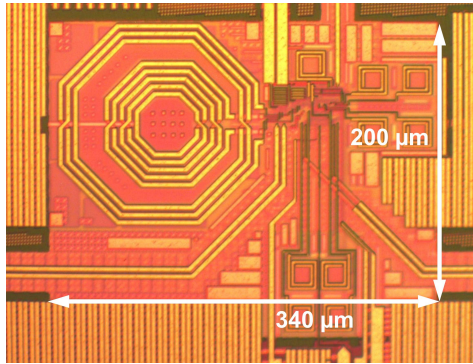
Figure 4.14: Equalizer die photograph.
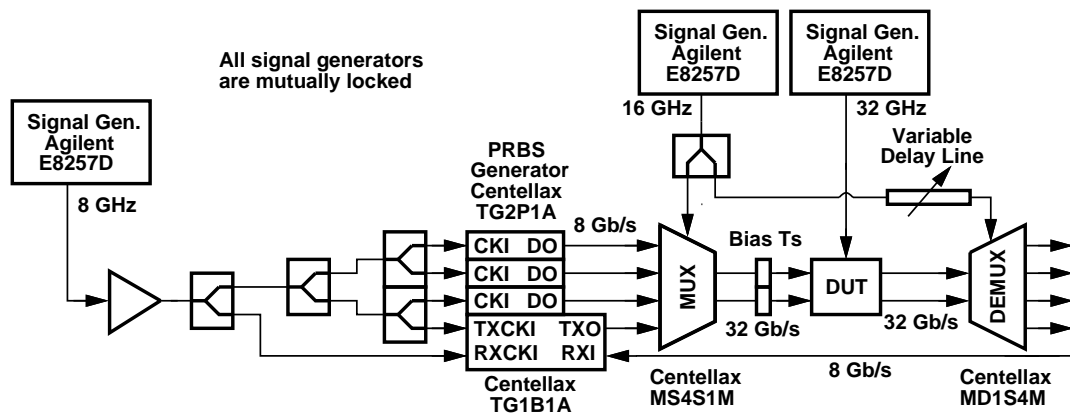
Figure 4.15: Test setup for measuring BER.

random-bit-sequence (PRBS) generators (three Centellax TG2P1A and one Cen-
tellax TG1B1A) are multiplexed (using Centellax MS4S1M) to form a 32-Gb/s
PRBS stream which is the input to the channel. The output of our chip at 32 Gb/s
is demultiplexed off chip (using Centellax MD1S4M) to recover the four 8-Gb/s
sequences, out of which one is sent to the BERT receiver (Centellax TG1B1A)
to measure the BER. Note that all the signal generators (Agilent E8257D) are
mutually locked by connecting their 10-MHz references.

Figure 4.16 plots the measured loss profile of the channel used in the equalizer
characterization. The channel exhibits a loss of 18 dB at 16 GHz and a deep notch
at 10 GHz. Figure 4.17 shows the eye diagrams at the output of the channel and

52
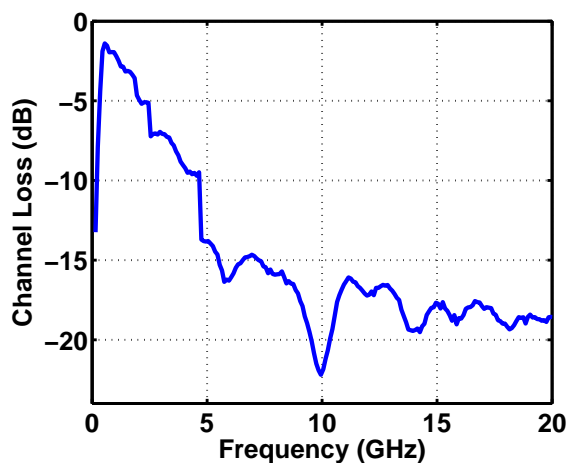
Figure 4.16: Measured frequency response of lossy channel.
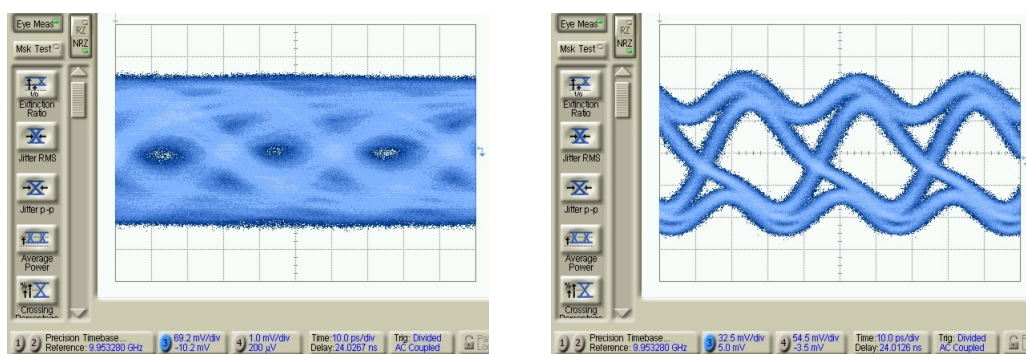
(a)   (b)

Figure 4.17: Eye diagrams at (a) channel output (10 ps/div., 69.2 mV/div.), (b) equalizer output (10 ps/div., 32.5 mV/div.).

at the output of the equalizer. Note that the PRBS generator itself has a peak-to-peak jitter of about 7 ps.

Figure 4.18 shows the bathtub curve for 32 Gb/s, measured by varying the generator's clock phase. We observe a horizontal eye opening of 0.44 UI for a BER < 10^{-12}. If the generator's jitter (≈ 0.22 UI) is discounted, the actual opening is larger.
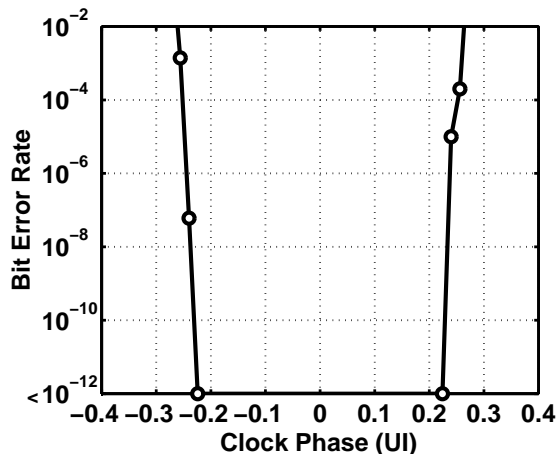
Figure 4.18: Measured bathtub curve.

Table 4.1: Performance summary and comparison to prior art

Reference	Hsieh VLSI 2009	Bulzacchelli ISSCC 2012	Toifl VLSI 2012	Kaviani CICC 2012	Jung JSSC Feb. 2015	This Work
Data Rate (Gb/s)	40	28	32	27	32	32
Architecture	1–tap DFE	CTLE + 15–tap DFE	CTLE + 15–tap DFE	1–tap DFE	CTLE + 2–tap DFE	CTLE + 1–tap DFE
DFE Clocking	Full–Rate	Half–Rate	Quarter–Rate	Quarter–Rate	Half–Rate	Full–Rate
Channel Loss @ Nyquist	15 dB	35 dB	36 dB	>10 dB	24 dB	18 dB
BER/ Eye Opening	$<10^{-11}$/ NA	$<10^{-9}$/ 35.6% UI	$<10^{-12}$/ 19% UI	$<10^{-9}$/ 11% UI	$<10^{-12}$/ 44% UI	$<10^{-12}$/ 44% UI
Supply (V)	1.2	1.05	1.15	1.1	1.0	0.73
Power (mW)	45	80*	97.6	11.1	5.8	9.3
Power Efficiency (pJ/bit)	1.125	2.857	3.05	0.411	0.232	0.29
Area (mm²)	0.05	0.81**	0.018	0.015	0.01	0.068
Technology	65–nm CMOS	32–nm SOI CMOS	32–nm SOI CMOS	40–nm CMOS	45–nm CMOS	45–nm CMOS

* Only for odd and even DFEs. Excludes CTLE, etc.
** Includes TX+RX+PLL.

Table 4.1 summarizes the performance of our prototype and recent state of the art. We achieve a power efficiency of 0.29 pJ/bit which is closer to [39] with a lower channel loss, but at a higher data rate. This work demonstrates the feasibility of high-speed full-rate DFEs with low power consumption.

CHAPTER 5

A 40-Gb/s 9.2-mW CMOS Equalizer

This chapter describes a 40-Gb/s equalizer that achieves a power efficiency of 0.23 mW/Gb/s. This performance is achieved through the use of a one-stage CTLE, a one-tap discrete-time linear equalizer (DTLE), a two-tap DFE, and two new latch topologies. Since in recent designs such as [39] and the design in Chapter 4, the CTLE draws significant power, this work introduces the DTLE as an efficient means of creating a high-frequency boost with only 0.3 mW.

5.1 Problem of CTLE
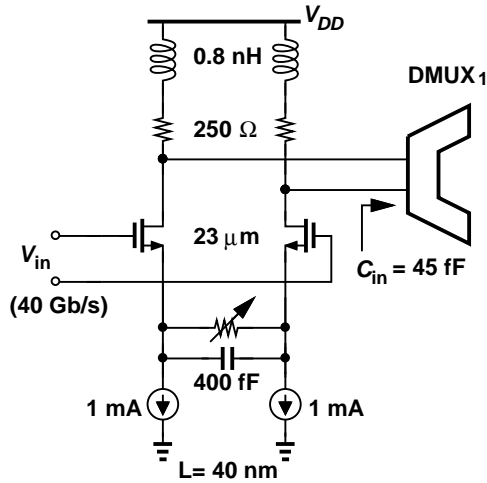
Figure 5.1: Implementation of a one-stage CTLE.

The front-end CTLE in Fig. 5.1, implemented as a capacitive degeneration amplifier, provides about 5.5 dB boost at Nyquist while consuming 2-mA current

from a 1-V supply, while driving a load of approximately 45 fF. The amount of boost can be programmed by adjusting the degeneration resistance, thus changing the low-frequency gain as shown in Fig. 5.2.
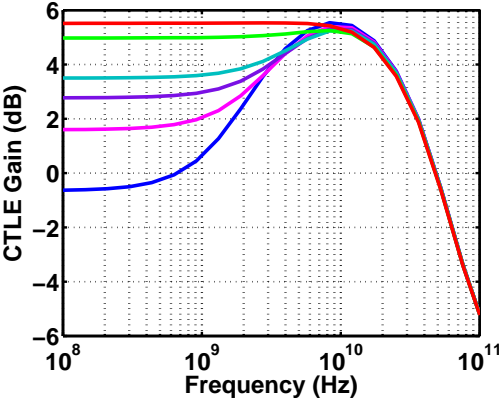
Figure 5.2: Frequency response of a one-stage CTLE.
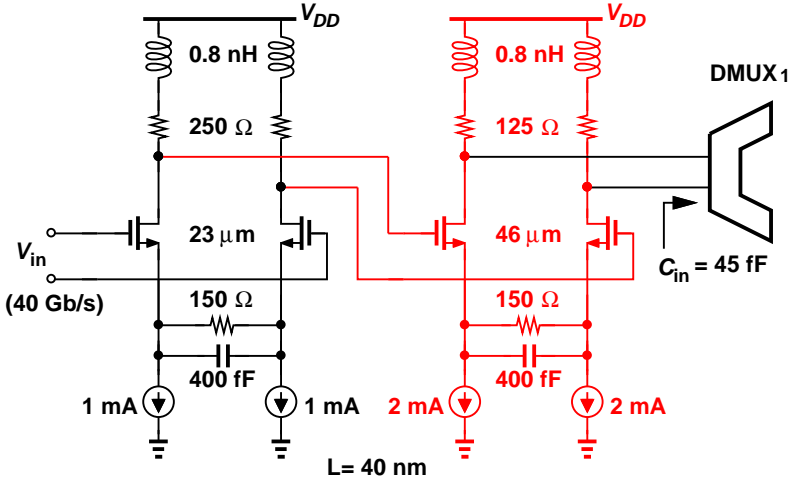
Figure 5.3: Implementation of a two-stage CTLE.

One approach to increase the boost is to add more stages to the CTLE. However, cascading stages inevitably reduces the bandwidth, thus demanding more power. Shown in Fig. 5.3 is a two-stage implementation which provides a boost of 11 dB at Nyquist, as shown in Fig. 5.4. The additional stage though, consumes about 4 mA thus increasing the power consumption by three times to
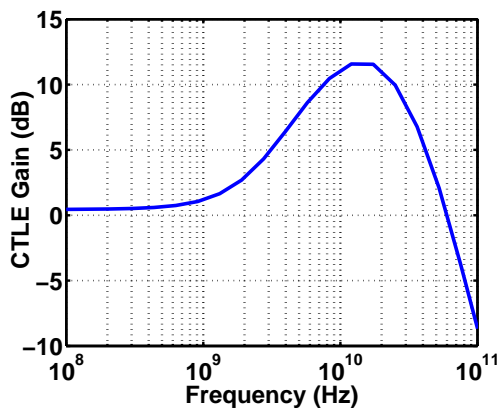
Figure 5.4: Frequency response of a two-stage CTLE.

obtain twice the boost with the same DC gain. This indicates a steep tradeoff between boost and power consumption. Note that a two-stage CTLE contains two zeros because of capacitive degeneration, one in each stage, and hence, the boost profile is twice as steep as that for a single-stage CTLE, which may be beneficial for certain channels. The number of zeros in the CTLE response may be optimally chosen according to the roll-off rate of the channel's magnitude response near Nyquist.

5.2 Discrete-time Linear Equalization

The steep tradeoff between the amount of boost and the power consumption of a CTLE mandates the necessity for other methods of creating linear equalization. The technique of linear equalization is not limited to continuous time, but can be performed in discrete time as well. A classic example is the discrete-time equalization used in transmitters where the data is discrete in time and discrete in amplitude [9]. Hence, equalization can be done using an FIR filter with a flipflop as a delay element, as shown in Fig. 5.5(a).

In receivers however, the data is continuous in amplitude and hence, an FIR
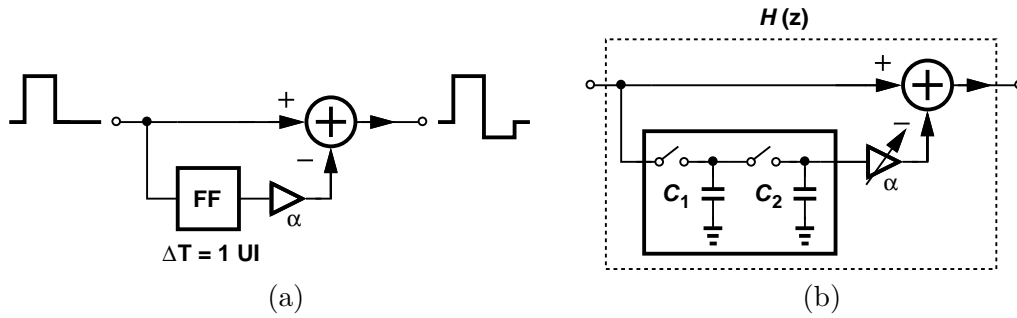
Figure 5.5: (a) Discrete-time linear equalization in transmitters, (b) discrete-time linear equalization in receivers using a master-slave sampling circuit.

filter would require a linear delay element so that the channel information is not lost before reaching the DFE. This linear delay element is implemented as a master-slave sampling circuit, as illustrated in Fig. 5.5(b), where the switches operate on complementary clock phases. This results in a transfer function $H(z) = 1 - \alpha z^{-1}$ and hence, a high frequency boost of $(1 + \alpha)/(1 - \alpha)$ with a DC gain of $1 - \alpha$. The tradeoff between the boost and the DC gain is apparent from the transfer function. The higher the boost, the lower is the DC gain, and vice versa.

But, we face two issues here. First, the master-slave sampling circuit needs to operate at full rate i.e. with a 40-GHz clock, which is difficult. Second, the charge sharing between capacitors C_1 and C_2 can cause significant ISI if their values are comparable. The former is discussed in the next section whereas the latter is addressed in Section 5.6.2.

5.3   Evolution of Architecture

In order to overcome the speed problem of the DTLE, it is implemented at half rate, as depicted in Fig. 5.6(a). Therefore, we need a demultiplexer, DMUX_1, at

(a)

(b)
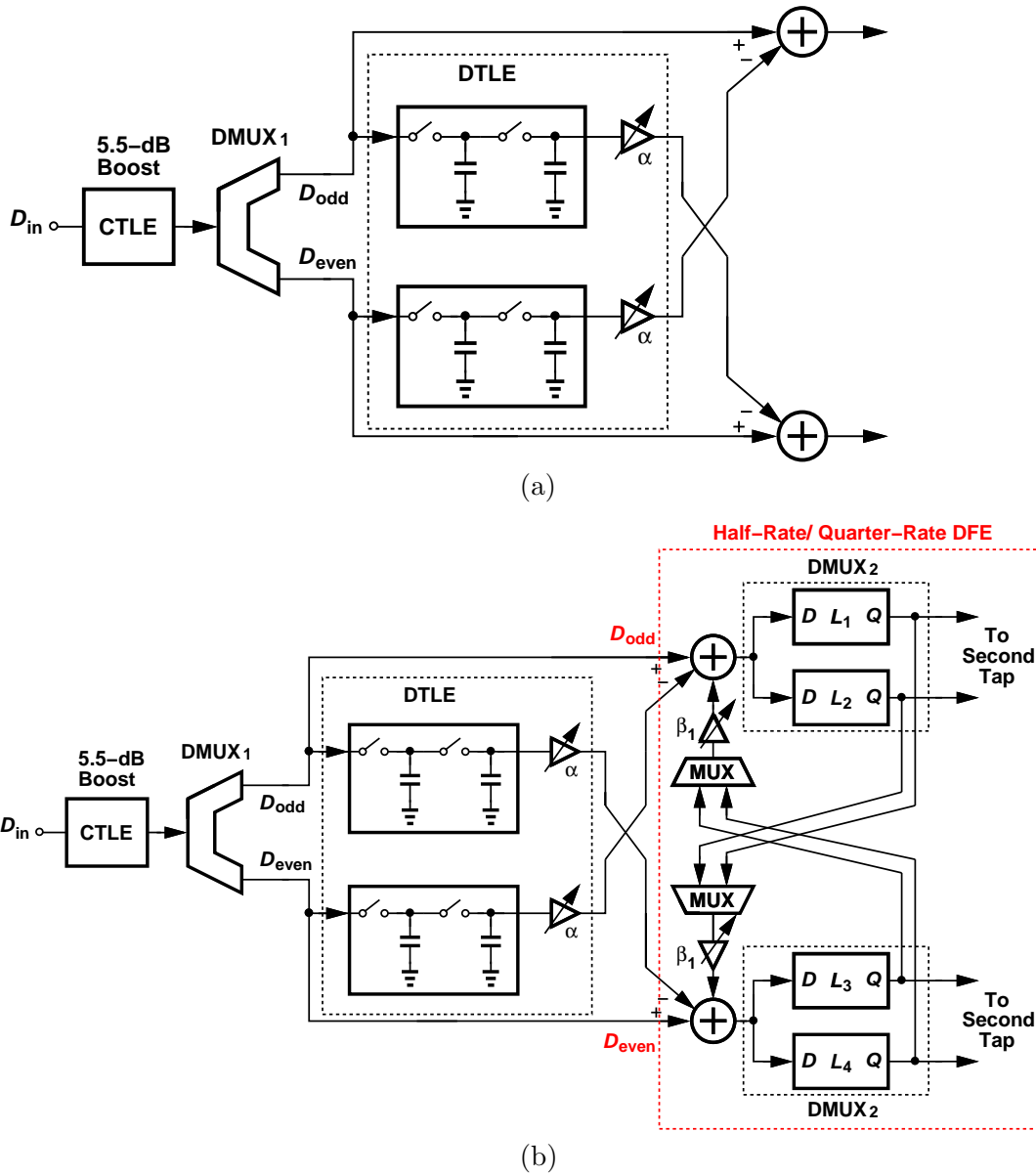
Figure 5.6: Evolution of equalizer architecture.

the input which decomposes the full-rate input stream at 40 Gb/s into two half-rate streams, D_{odd} and D_{even}, at 20 Gb/s. DMUX$_1$ must be linear enough so that the channel information is not lost before going into the DFE. Since there are two summers already present for the one-tap DTLE, a DFE can be naturally added to this structure by merging the summers, as shown in Fig. 5.6(b). The DFE itself

receives D_{odd} and D_{even}, sums them with the first tap, and demultiplexes the result, generating 10-Gb/s outputs. These outputs are pairwise multiplexed and injected back into the summing junction. This is called a half-rate/quarter-rate DFE [39].

5.4 Charge-Steering Circuits

The use of charge steering can be sketched back to the early 1990s, in regenerative BiCMOS comparators [40, 41]. This work uses charge-steering techniques extensively in designing demultiplexers, DTLE and DFE.

5.4.1 Return-to-Zero Latch

Shown in Fig. 5.7(a) is a charge-steering return-to-zero (RZ) latch [42]. When CK is low, the output nodes are precharged to V_{DD} and the tail node is discharged to ground. When CK goes high, X and Y discharge differentially based on the input and the operation ceases when the tail capacitor charges to a value high enough such that the input pair turns off. This circuit can provide amplification and latching. For a small differential input, V_{in}, the voltage gain is relatively independent of the input common-mode level, V_{CM}, and is given by [42]

$$A_v \approx 2\frac{C_T}{C_D}.$$ (5.1)

For moderate to large input swings, the differential output voltage depends on V_{CM} and is equal to [43]

$$V_{out} = \frac{C_T}{C_D}\frac{(V_{CM} - V_{TH})^2 + \dfrac{3V_{in}^2}{4}}{V_{CM} - V_{TH} + \dfrac{V_{in}}{2}},$$ (5.2)

where V_{TH} denotes the threshold voltages of the input transistors. The gain of this circuit can be increased by adding a regenerative cross-coupled PMOS pair, as shown in Fig. 5.7(b) [39].

(a)   (b)
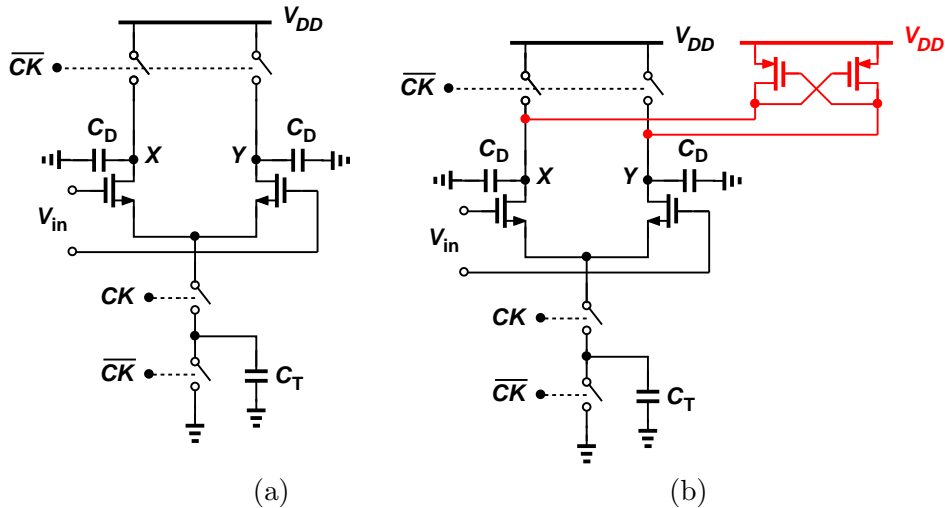
Figure 5.7: Charge-steering RZ latch (a) without, and (b) with the cross-coupled PMOS pair.

5.4.2   Non-Return-to-Zero Latch

Shown in Fig. 5.8 is a charge-steering NRZ latch [42]. When CK is low, the inputs are sampled on the parasitic capacitances of nodes X and Y. When CK goes high, the cross-coupled NMOS pair provides regeneration and consequently increases the output differential swing. Since this circuit does not need any reset phase, the outputs are in NRZ form.

The NRZ latch of Fig. 5.8 can provide voltage gain. If the transistors operate in weak to moderate inversion, the gain can be approximated as [42]

$$\frac{V_{XY\infty}}{V_{XY0}} = \exp\left(\frac{C_T}{C_D}\frac{V_{CM} - V_{GS}}{2\zeta V_T}\right) \tag{5.3}$$

where the left-hand side is the ratio of the final and initial voltages, V_{GS} is

assumed relatively constant, and ζ denotes the subthreshold nonideality factor and is given by $1 + C_d/C_{ox}$, where C_d is the capacitance of the depletion layer under the gate oxide.
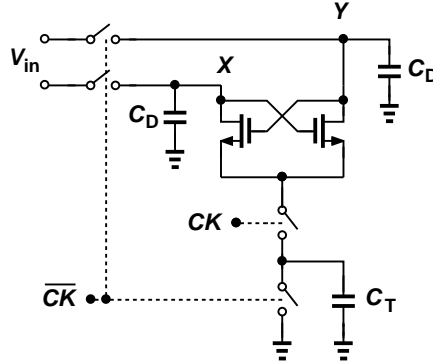
Figure 5.8: NRZ charge-steering latch.

These charge-steering circuits provide moderate differential swings of a few hundred millivolts without consuming any static power. This helps improve speed and reduces power consumption ($\approx fCV_{DD}V_{swing}$).

5.5   DFE Timing Considerations

For a full-rate DFE using charge-steering stages, if we reset the summing node, as shown in Fig. 5.9(a), the timing constraint of the circuit is given by

$$t_{CQ} + t_{setup} < 0.5 \text{ UI}, \tag{5.4}$$

which is difficult to meet at 40 Gb/s where 1 UI = 25 ps. In this design, the half-rate DFE in Fig. 5.9(b) is employed instead. When CK is low, the odd summer is reset, the odd flipflop is in regenerating mode, and hence the odd data from the previous bit can be fed to the even summer to form the first DFE tap. Therefore, its timing constraint can be written as

$$t_{CQ} + t_{setup} < 1 \text{ UI}, \tag{5.5}$$

62
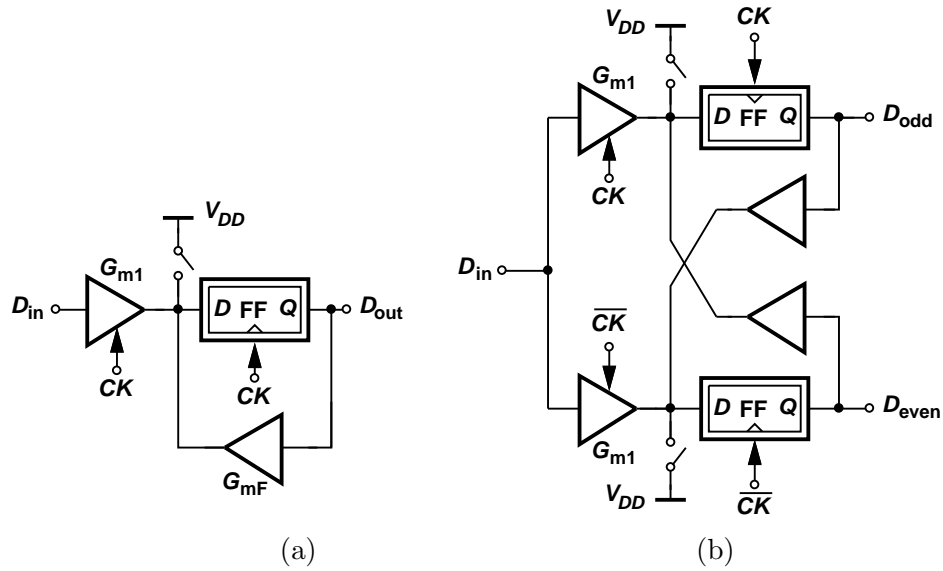
Figure 5.9: Charge-steering DFE architectures: (a) Full-rate DFE, (b) half-rate DFE.

which is feasible. It must also be noted that the setup time of charge-steering
RZ latches is lower than the CML latch because of the inherent reset phase [39].
This makes it easier to meet this timing with charge-steering circuits.

5.6 Building Blocks

This section presents the circuit-level implementation of the equalizer building-
blocks in 45-nm CMOS technology.

5.6.1 1-to-2 Demultiplexer

As depicted in Fig. 5.10, DMUX$_1$ is implemented using two charge-steering NRZ
latches with complementary clock phases. The gain of these latches can be in-
creased by increasing the value of the tail capacitor. However, as discussed ear-
lier, this demultiplexer needs to be linear enough and hence, we want to have
moderate gains for these regeneration pairs. In this design, the tail capacitor is

chosen to be 6 fF which provides a gain of about 6 dB, as seen from its input-output characteristics shown in Fig. 5.11. The simulated 1-dB gain-compression point is 120 mV$_p$ which suffices for our purposes. A detailed analysis of linearity requirements for equalizers is described in [39]. Nonlinearity causes significant additional ISI if the main cursor at the receiver input approaches 1.5 times the 1-dB compression point of the receiver.
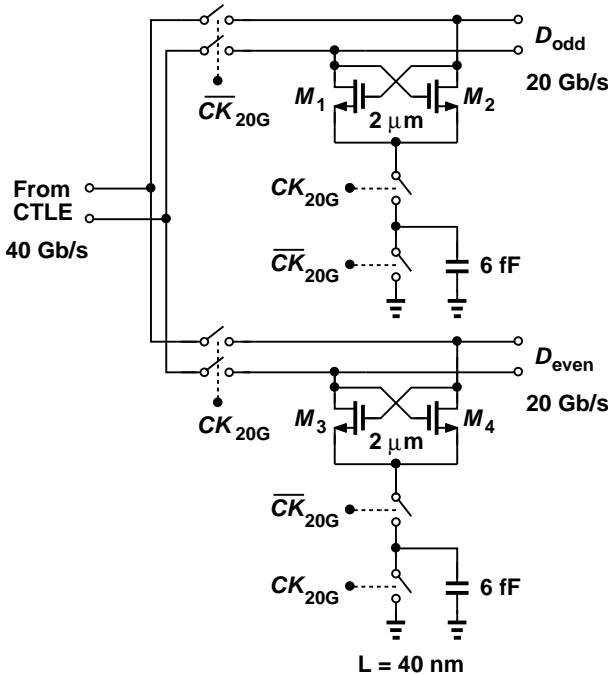
Figure 5.10: Implementation of DMUX$_1$.

5.6.2 Discrete-Time Linear Equalizer

The output of the demultiplexer needs to drive the DTLE. As shown in Fig. 5.6(a), the DTLE needs a master-slave sampling circuit as a linear delay element. Since there is one level of sampling in DMUX$_1$, it is possible to merge DMUX$_1$ and the DTLE by adding just one more stage of sampling switches, as illustrated in Fig. 5.12. When CK is low, the data is sampled on nodes X and Y. When CK
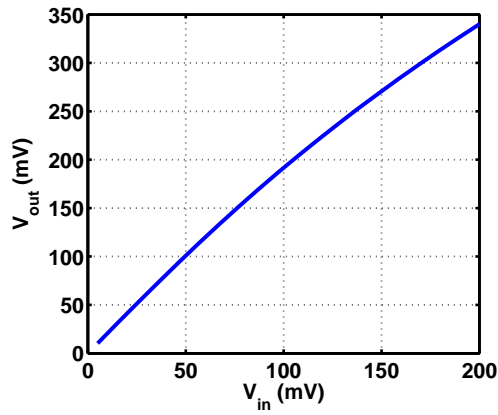
Figure 5.11: Voltage transfer characteristics of DMUX$_1$.
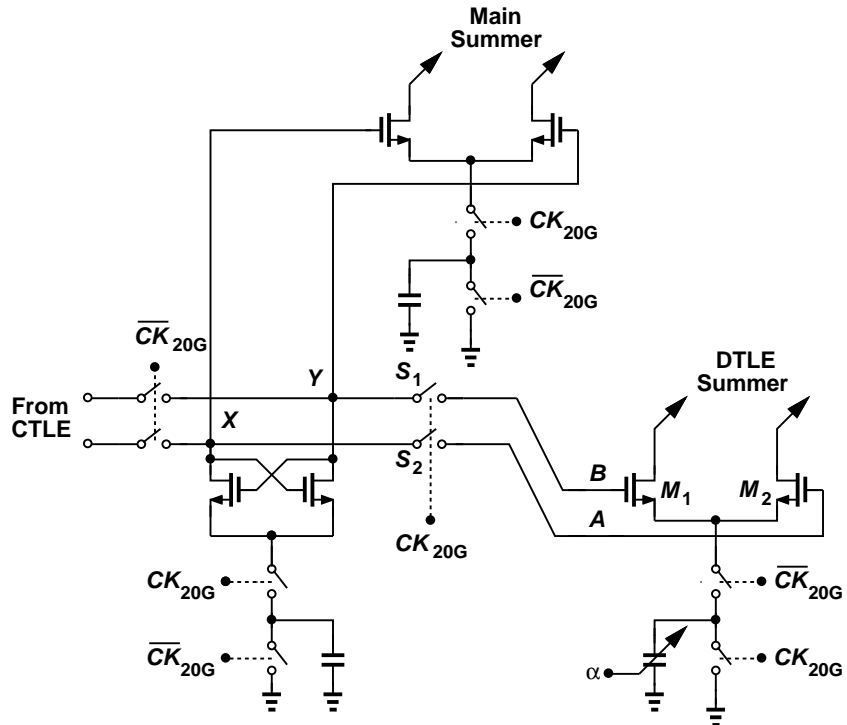
Figure 5.12: One half-rate path showing the implementation of DTLE.

goes high, this latched data is sampled on to the parasitic capacitances of nodes A and B. When CK goes low again, the data at A and B is amplified by the differential pair and fed back to the opposite summer, thus effectively creating a 1-UI delay, thereby resulting in a transfer function $1 - \alpha z^{-1}$.

The value of α can be programmed by adjusting the tail capacitor of the charge-steering differential pair in Fig. 5.12, thus adjusting its transconductance. In this design, α can be programmed from 0 to 0.3, which results in a maximum boost of 5.4 dB at a DC gain of -3 dB. Since DMUX$_1$ provides a gain of 6 dB, the overall DC gain of the combined DMUX$_1$ + DTLE is 3 dB. The two differential pairs in the DTLE draw only 0.3 mW.

In order to minimize the effect of charge sharing in Fig 5.12, the capacitances at A and B must be designed to be much smaller than those at X and Y. It must be noted that DMUX$_1$ not only drives the DTLE, but also the main summer, as shown in Fig. 5.12. Hence, $C_{X,Y}/C_{A,B} \approx 5$ in this design, resulting in little charge sharing. Also, transistors M_1-M_2 are made small by using PMOS devices for switches S_1 and S_2, such that when they turn off, they raise the common-mode voltage at A and B, thus resulting in more overdrive for M_1-M_2.

It is interesting to note that charge sharing in the DTLE has a useful side effect. To understand this, let us assume C_A is initially discharged in Fig. 5.12. When CK_{20G} goes high for the very first time, C_X dumps charge on to C_A, and if the initial voltage stored on C_X is V_1, the final voltage at the end of the first cycle, V_{F1}, across C_A, is given by,

$$V_{F1} = \frac{C_X V_1}{C_X + C_A}. \tag{5.6}$$

In the next cycle (corresponding to the third bit period because of half-rate operation), let us assume the initial voltage stored on C_X is V_3. The final voltage, V_{F3}, across C_A, after the circuit is clocked again, can be written using charge conservation as,

$$(C_X + C_A)V_{F3} = C_X V_3 + C_A V_{F1} \tag{5.7}$$

$$\implies V_{F3} = \frac{C_X}{C_X + C_A}\left[V_3 + \frac{V_1 C_A}{C_X + C_A}\right]. \tag{5.8}$$

Continuing these calculations,

$$V_{F5} = \frac{C_X}{C_X + C_A}\left[V_5 + \frac{V_3 C_A}{C_X + C_A} + \frac{V_1 C_A^2}{(C_X + C_A)^2}\right], \qquad (5.9)$$

and so on. Consequently, the transfer function of the DTLE can be written as

$$H(z) = 1 - \alpha\frac{C_X}{C_X + C_A}\left[z^{-1} + \frac{V_3 C_A}{C_X + C_A}z^{-3} + \frac{V_1 C_A^2}{(C_X + C_A)^2}z^{-5} + ...\right]. \qquad (5.10)$$

Thus, the one-tap DTLE in the presence of charge sharing manifests itself as a multi-tap FIR filter, with all even coefficients zero. This transfer function is of the form

$$H(z) = 1 - \alpha_1 z^{-1} - \alpha_3 z^{-3} - \alpha_5 z^{-5} - ... \qquad (5.11)$$

where,

$$
\begin{aligned}
\alpha_1 &= \alpha\frac{1}{1 + C_A/C_X}, \\
\frac{\alpha_{n+2}}{\alpha_n} &= \frac{C_A/C_X}{1 + C_A/C_X}, \text{for odd integer values of } n, \text{and} \\
\alpha_n &= 0, \text{for even integer values of } n. \qquad (5.12)
\end{aligned}
$$

Thus, by choosing α and the ratio $C_{A,B}/C_{X,Y}$, we can set α_1 and α_3 independently. All other coefficients are automatically set based on Eq. (5.12). Thus, a one-tap DTLE can be extended to a multi-tap version by employing charge sharing.

Figure 5.13 compares the eye diagram at the summing junction without and with the DTLE, with the DFE turned off. It is seen that the presence of the DTLE improves the eye opening at the summing junction.

5.6.2.1 Noise in DTLE

It is important to address the additional noise contributed by the DTLE. Since, the second capacitor $C_{A,B}$ is designed to be small to avoid significant charge
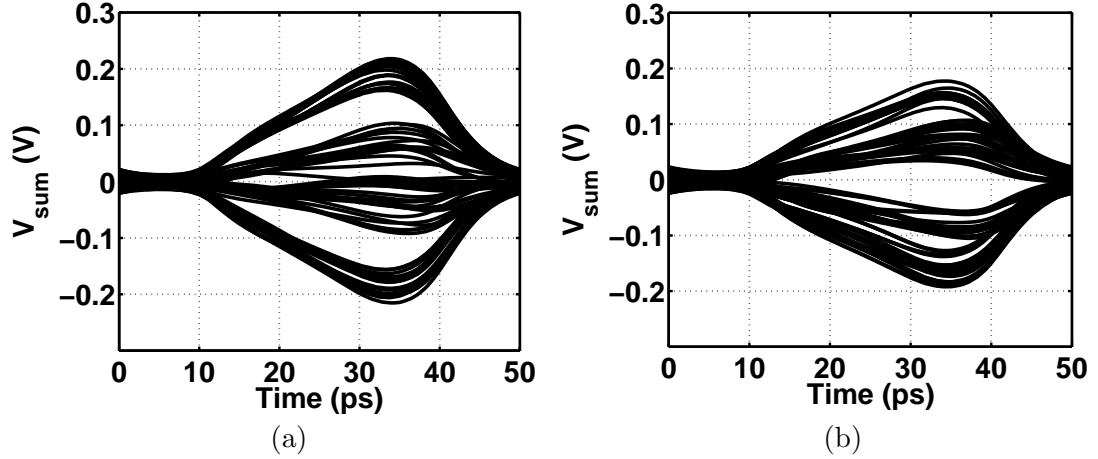
Figure 5.13: Eye diagrams at the summing junction with DFE off (a) without DTLE, (b) with DTLE.

sharing, its kT/C noise can be high, thus requiring careful study [44]. In order to perform this analysis, a simplified diagram of the DTLE, with a transfer function $1 - \alpha z^{-1}$, is shown in Fig. 5.14. Considering the noise of only the passive samplers, and neglecting the noise contribution from other MOSFETs, the total mean-square noise voltage at the summing node because of the DMUX+DTLE alone can be written as,

$$\overline{V_{n,sum}^2} = \frac{4kT}{C_X} + (\text{total rms noise across } C_A)^2 \alpha^2. \qquad (5.13)$$

Here, the factor 4 comes from the fact that the regenerative NMOS pair of DMUX_1 provides a voltage gain of 2 (or 6 dB).

The mean-square noise across C_A consists of two components: (1) The noise of S_2 given by kT/C_A, and (2) the $4kT/C_X$ noise sampled on C_A due to charge sharing approximately given by $\dfrac{4kT}{C_X} \times \left(\dfrac{C_X}{C_X + C_A}\right)^2$ (assuming zero on-resistance of the switches). Thus, from Eq. (5.13),

$$\overline{V_{n,sum}^2} \approx \frac{4kT}{C_X} + \left[\frac{kT}{C_A} + \frac{4kT}{C_X}\left(\frac{C_X}{C_X + C_A}\right)^2\right]\alpha^2. \qquad (5.14)$$

Note that the first term, $4kT/C_X$, would be present even without the DTLE,
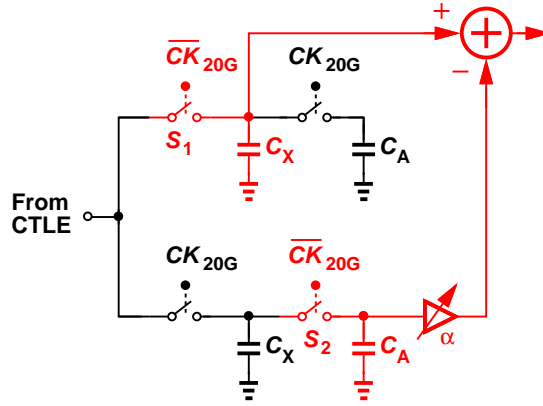
Figure 5.14: Simplified diagram of the DTLE.

and hence the DTLE contributes only the remaining terms in Eq. (5.14). In this design, $C_X = 5C_A$ and $\alpha_{\max} = 0.3$, thus,

$$\overline{V_{n,sum}^2} \approx \frac{4kT}{C_X} + \frac{0.7kT}{C_X}, \tag{5.15}$$

which is an increase of only 8.4% in the rms noise voltage at the summing node caused by adding the DTLE to DMUX_1.[1] With the inclusion of the summer noise, and other noise sources, this penalty would be even smaller.
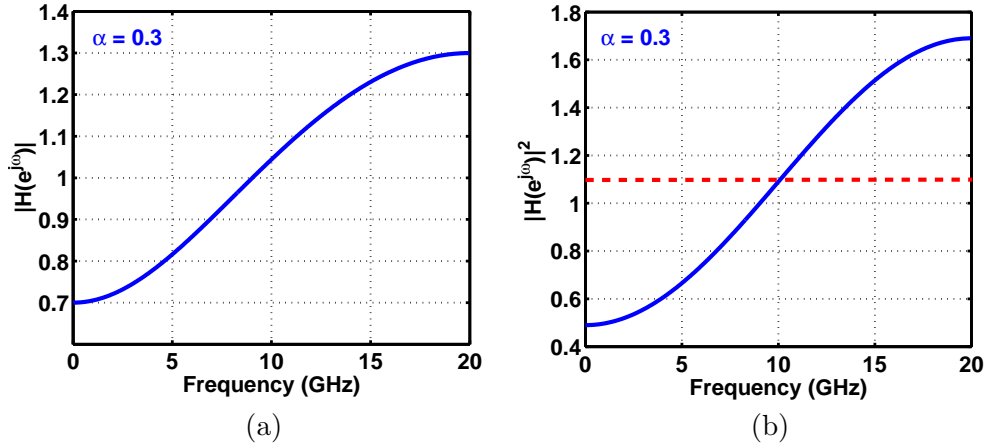
Figure 5.15: DTLE: (a) Magnitude response, (b) square of the magnitude.

---

[1]The right-hand side in Eq. (5.15) must be multiplied by a factor 2 to account for the differential signal path.

Does DTLE significantly amplify the noise of the previous stages? To understand this, the magnitude of DTLE transfer function $H(z) = 1 - \alpha z^{-1}$ is plotted in Fig. 5.15(a) for $\alpha = 0.3$. The power spectral density (PSD) of noise at the DTLE output can be obtained by multiplying the PSD of the noise at the input of the DTLE by $|H(e^{j\omega})|^2$. Since $|H(e^{j\omega})|^2$ in Fig. 5.15(b), for $\alpha = 0.3$, is symmetric about 1.1, any white-noise power at the input will get amplified by a factor 1.1, when integrated at the output of the DTLE. Thus, the DTLE amplifies the input white-noise power by only 10% for $\alpha = 0.3$. This number is lower for smaller values of α. This can be quickly verified by putting $\alpha = 0$, which results is an all-pass response of magnitude 1.

5.6.3 Decision-Feedback Equalizer

The DFE is implemented as a half-rate/quarter-rate architecture as shown in Fig. 5.6(b). Latches L_1-L_4 may be implemented as the RZ latch of Fig. 5.7(a). This circuit can achieve a large gain from input to nodes P and Q by using a large tail capacitor but at the cost of a common-mode degradation at the output. The common-mode degradation might make it difficult to drive the following stage. The following stage will have a much lower input common-mode resulting in a very low gain for that stage. This problem can be alleviated by adding the cross-coupled PMOS pair, as shown in Fig. 5.7(b). However, the PMOS pair turns on only when one of the outputs falls to one PMOS threshold below the supply. Following this, the PMOS pair will take time to regenerate one of the outputs to V_{DD} as shown in Fig. 5.16. However, a 25-ps UI is not sufficient for these PMOS devices to turn on and regenerate appreciably, thus resulting a low output differential swing as shown in Fig. 5.17(a) and a reduced output common-mode. When this latch is used in the DFE loop, the eye diagram at the summing

junction is as shown in Fig. 5.17(b). How do we improve the output swing of the latch without degrading the output common-mode significantly?
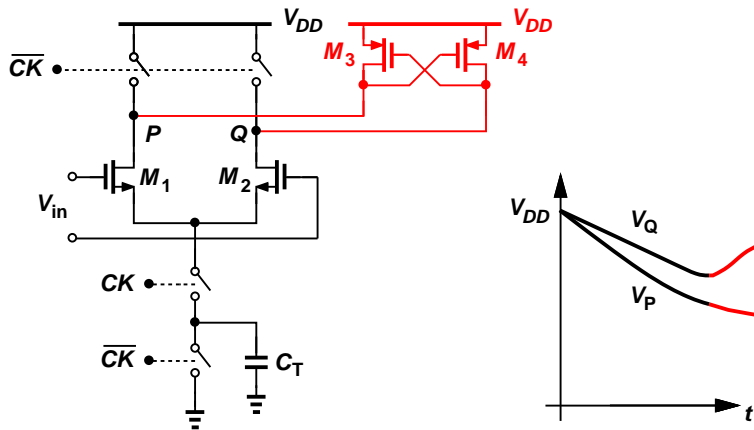
Figure 5.16: Operation of RZ charge-steering latch with a cross-coupled PMOS pair.
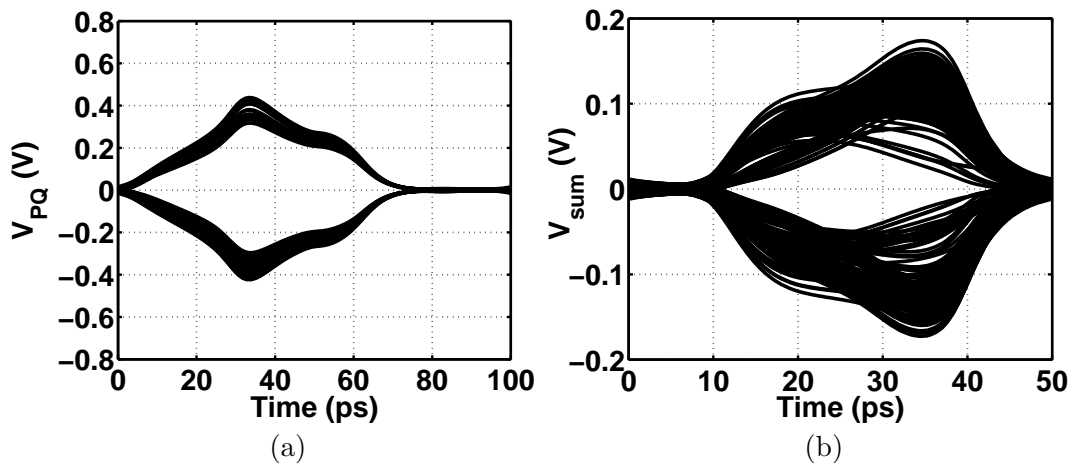
Figure 5.17: (a) Eye diagram at the output of the old charge-steering latch with a cross-coupled PMOS pair, and, (b) corresponding eye diagram at the summing junction.

To solve this problem, we add cascode transistors M_5-M_6 in the signal path as shown in Fig. 5.18. Initially when CK_{10G} is low, all the nodes P, Q, X and Y are reset to V_{DD}. When CK_{10G} goes high, M_1 and M_2 draw a differential current from X and Y, and M_5-M_6 are off. M_5 and M_6 remain off until either V_X or V_Y falls to about $V_{DD} - V_{TH}$. At this point, the large voltage difference between
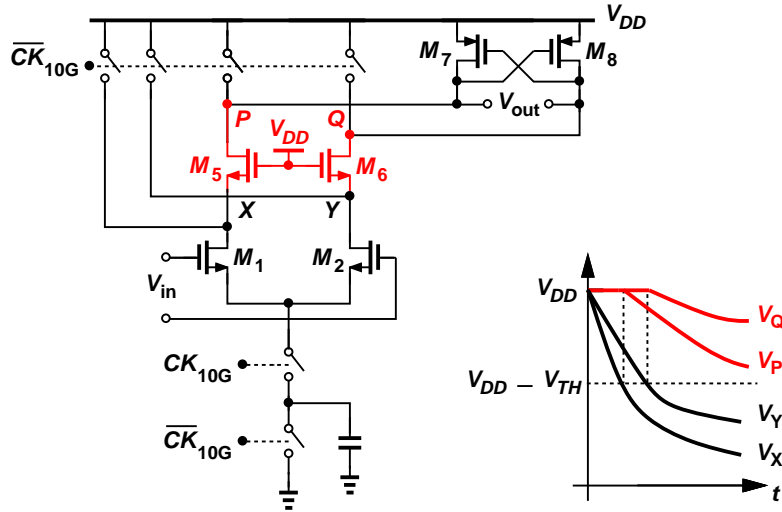
Figure 5.18: Adding a cascode pair to the charge-steering latch.

V_X and V_Y allows only M_5 or M_6 to turn on and transfer the amplification to P or Q, minimizing the common-mode degradation at these nodes, as shown in Fig. 5.18. In other words, M_5 and M_6 isolate P and Q from the large common-mode drop inevitably imposed by the need for a high differential gain. The cascode transistors also isolate the outputs from X and Y, inherently reducing the capacitance at these nodes and hence improving the gain from V_{in} to V_{XY}, effectively improving the gain to the output.

The gain from V_{in} to V_{XY} can be improved further by adding an NMOS cross-coupled pair M_3-M_4, as shown in Fig. 5.19. Note that M_3-M_4 have their own tail capacitor which is chosen to be about one-tenth the main tail capacitor. Instead, M_3-M_4 could be connected to the main tail capacitor by sharing the source terminals of transistors M_1-M_4. However, in that case, M_3-M_4 could potentially "steal" charge from the main differential pair at the start of the operation by just reducing the common-mode at X and Y, without providing any differential gain, thus resulting in an overall reduction in the output swing. By providing a separate tail capacitor for M_3-M_4, we make sure that it provides regeneration only
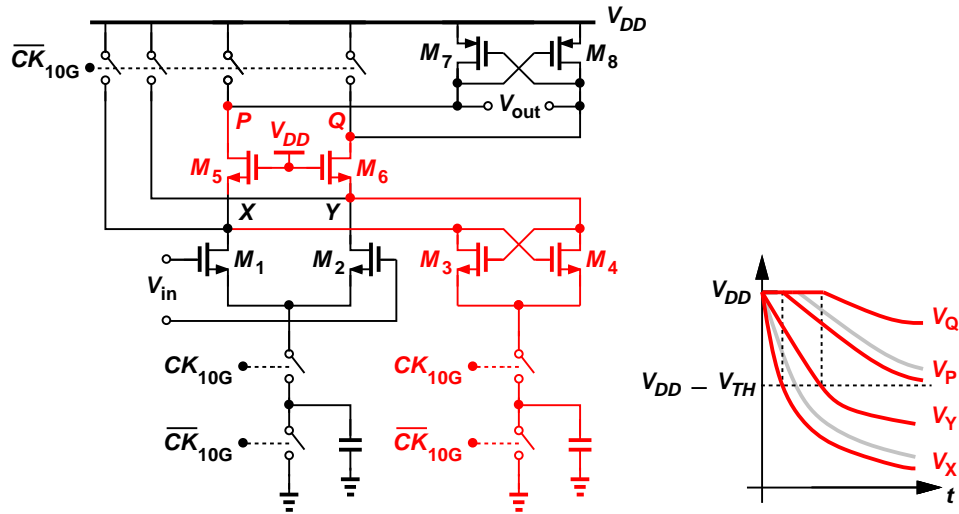
Figure 5.19: Adding a cross-coupled NMOS pair to the cascode charge-steering latch.
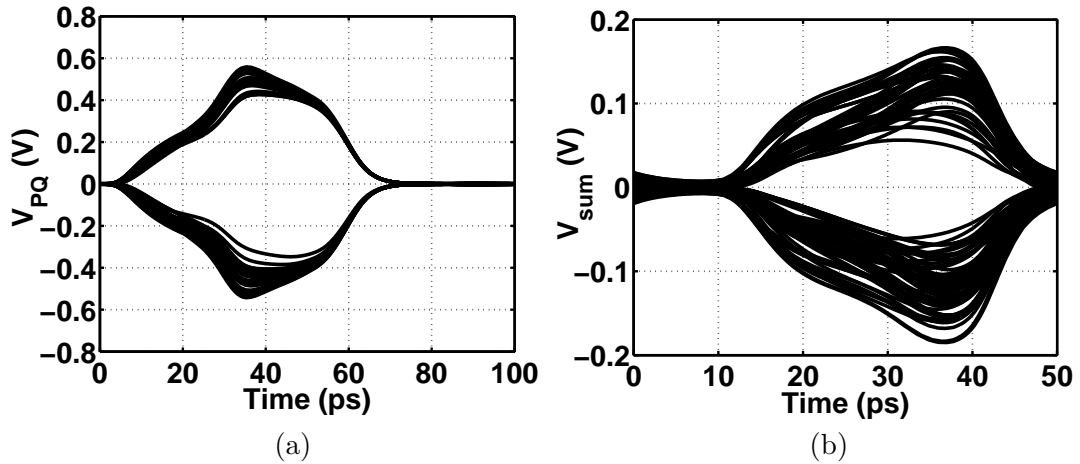
Figure 5.20: (a) Eye diagram at the output of the cascode charge-steering latch, and, (b) corresponding eye diagram at the summing junction.

for a finite time at the start of the operation. The eye diagrams at the output of this cascode charge-steering latch and the corresponding summing junction are shown in Fig. 5.20, indicating a taller and wider eye at the summing junction than in Fig. 5.17(b).

Note that it is not required to connect the gates of transistors M_5-M_6 to V_{DD}. These transistors can be cross-coupled as shown in Fig. 5.21 to further aid our

73
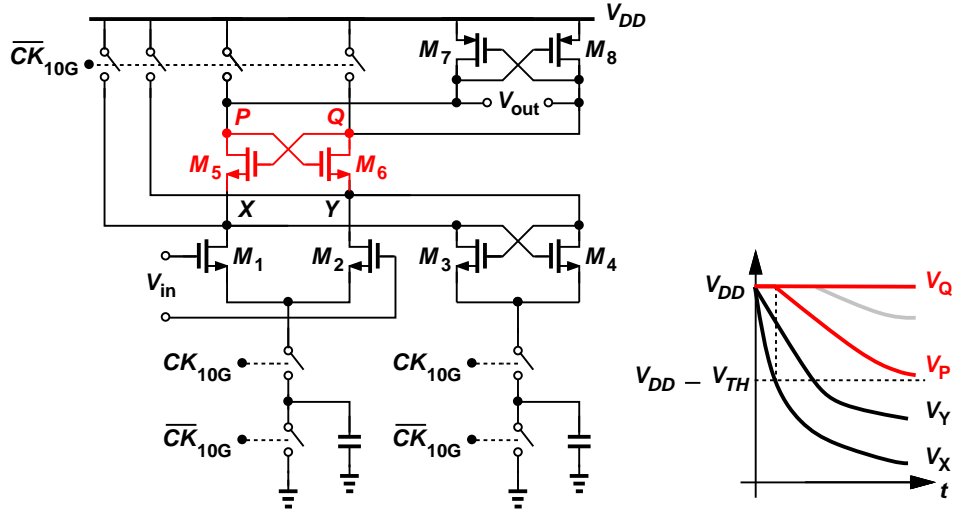
Figure 5.21: Improved charge-steering latch with two cross-coupled NMOS pairs.

operation. When either X or Y goes below $V_{DD} - V_{TH}$, one of the transistors, M_5 or M_6, turns on, and one of the nodes, P or Q, starts to discharge. This delays the turn-on time of the other transistor. Say, P begins to discharge as M_5 turns on first. Then, V_Y must go one threshold below V_P for M_6 to turn on. This improves the output differential swing by delaying the turn-on of M_6. Consequently, the latch provides two and a half times the output swing of the topology in [39], as seen in Fig. 5.22(a). The corresponding eye diagram at the summing junction [Fig. 5.22(b)] is also improved significantly as a result, with a vertical eye opening of about 150 mV$_{\text{pp}}$.[2]

We can point out two key distinctions between the topology of Fig. 5.21 and the StrongARM latch [45, 46]: (1) Our circuit operates with a finite tail charge, producing moderate (rather than rail-to-tail) swings ($\approx$ 500 mV$_{\text{pp}}$, single-ended)

---

[2]From simulations, the total integrated noise at the summer output is equal to 1.15 mV$_{\text{rms}}$ for both minimum and maximum CTLE peaking conditions. Not considering offsets and latch sensitivity, the minimum eye opening required at the summing junction for a BER $< 10^{-12}$ is only 16.1 mV$_{\text{pp}}$.
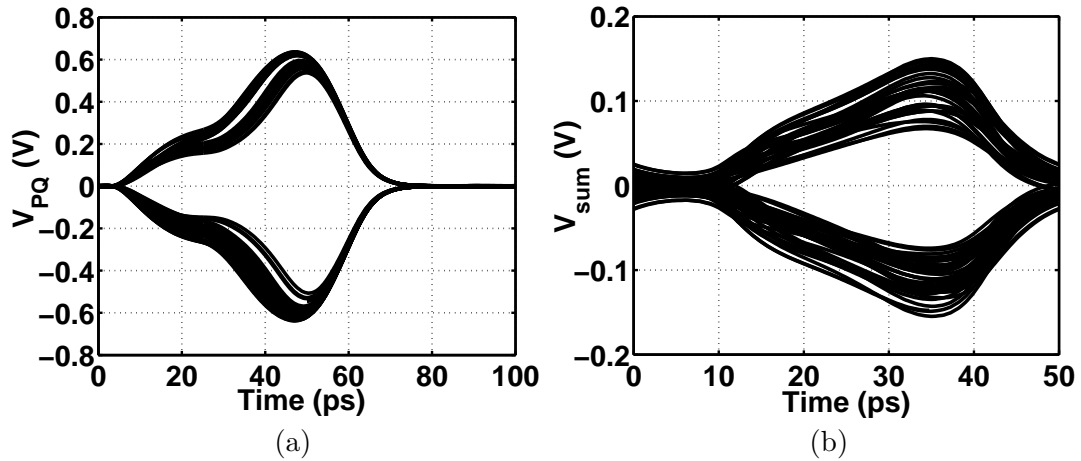
Figure 5.22: (a) Eye diagram at the output of the improved charge-steering latch with two cross-coupled NMOS pairs, and, (b) corresponding eye diagram at the summing junction.

at X, Y, P and Q, improving the speed, and reducing the power consumption ($\approx fCV_{DD}V_{swing}$ for each node), and (2) the additional gain provided by M_3-M_4 also enhances the speed of the latch.

5.6.4   Drawback of Summer Output Resetting

The input to the latches L_1-L_4 are driven by the summer. Since, the half-rate summer drives the quarter-rate latches, the inputs of these latches reset after half the period of their amplification phase (0.5×2 UI). Considering the old latches in Fig. 5.7, let us assume that the inputs are such that $V_X > V_Y$, at the point at which the outputs of the summer is reset. Now, both the inputs of the latch are pulled to V_{DD}. Consequently, the tail capacitor can charge further until the voltage across it reaches $V_{DD} - V_{TH}$. Both M_1 and M_2 have the same V_{GS}, but since $V_X > V_Y$, M_1 carries more current than M_2, thus discharging X faster than Y. Hence, the output differential swing starts reducing as seen in the eye diagram in Fig. 5.17(a). This effect is alleviated by introducing cascode devices in

Fig. 5.19 and cross-coupled devices in Fig. 5.21, since the outputs are effectively shielded from these intermediate nodes, X and Y.

In order to compare the performance of the new latch in Fig. 5.21 and old latch in Fig. 5.7, an RZ PRBS input of 100 mV$_\text{p}$ differential is applied to these latches, with a pulse width of 25 ps, and a reset period with both its inputs connected to V_{DD}, for 25 ps, emulating the actual summer. The latches drive a load of 20 fF, single-ended. The tail capacitance is swept from 20 fF to 150 fF, and the output differential swing and the output common mode at the end of the latching operation are plotted in Fig. 5.23. For the new latch in Fig. 5.21, the tail capacitor of the NMOS cross-coupled pair M_3-M_4 is chosen to be of one-tenth the value of the main tail capacitor.
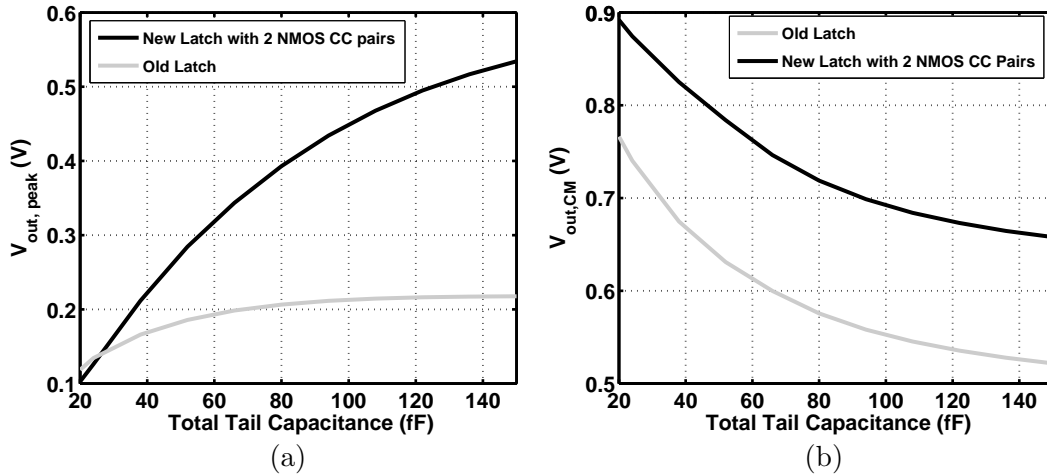
Figure 5.23: Comparison of old and new charge-steering latches in terms of (a) output differential swing, (b) output common mode.

Figure 5.23 indicates that the new latch provides a higher output swing for a given tail-node capacitance i.e. for a given power consumption, with a reduced common-mode degradation at its output. The improvement in the output swing is more significant at higher values of tail capacitance.

5.6.5 Effect of DTLE in the Presence of DFE

In order to study the combined effect of the DTLE and the DFE on the overall system, we compare the eye diagrams at the summing junction with and without DTLE, in the presence of DFE (Fig. 5.24). It can be seen that the DFE along with the DTLE improves the eye diagram at the summing junction significantly.
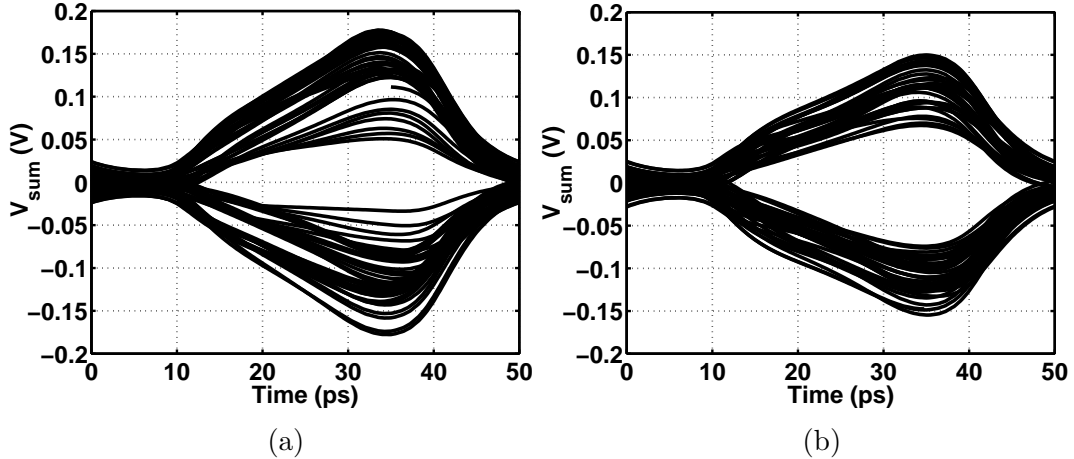
Figure 5.24: Eye diagrams at the summing junction with DFE on (a) without DTLE, (b) with DTLE.

5.6.6 Feedback Tap Control and Vernier Charge Delivery

In order to program the tap-coefficients, we adjust the tail capacitance of feedback/feedforward transconductances, thus changing their gains. In this design, the tail capacitor of the DTLE can be varied from 0 to 32 fF in steps of 2 fF, the DFE tap-1 from 0 to 50 fF in steps of 1 fF, and DFE tap-2 upto 20 fF in steps of 1 fF. Each tap's transconductance is also provided with an enable signal to turn off the tap as shown in Fig. 5.25(a).

The minimum value of a tap-coefficient is limited by the parasitic capacitance at the tail node which can be of the order of a few femtofarads. However, we might need a value lower than that. In order to overcome this difficulty, we incorporate
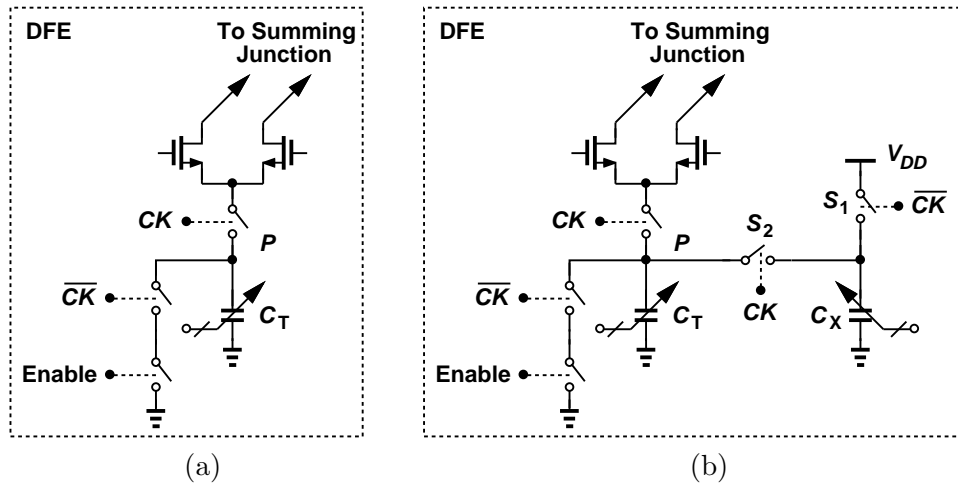
Figure 5.25: (a) Feedback tap control, (b) vernier charge delivery.

a vernier technique by which we precharge the tail capacitor, C_T, using another capacitance, C_X, in Fig. 5.25(b). When CK is low, C_T is completely discharged to ground, and C_X is precharged to V_{DD}. When CK goes high, S_2 turns on, resulting in charge-sharing between C_T and C_X. Therefore, by selecting the relative values of C_T and C_X, C_T can be programmed to have any initial voltage from 0 to V_{DD}. In this design, for the feedback transconductance of the second DFE tap, C_X can be varied from 0 to 10 fF in steps of 1 fF.

5.6.7 Half-Rate Path

Figure 5.26 shows one half-rate path. The multiplexer required in the feedback path of the DFE is integrated into the feedback G_m thus avoiding any additional delay from the multiplexing operation.

5.6.8 Overall Architecture

The overall architecture shown in Fig. 5.27 consists of a one-stage CTLE, a one-tap DTLE and a two-tap DFE. The second tap is shown in gray. The overall
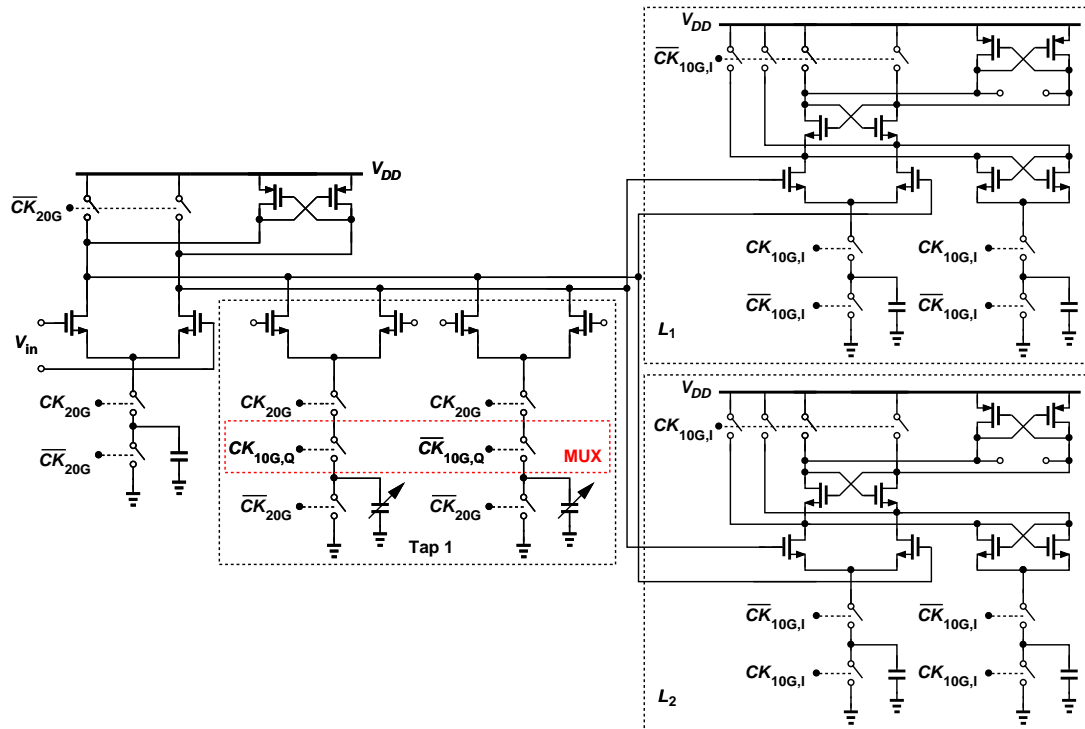
Figure 5.26: One half-rate path.

equalizer takes a 40-Gb/s input and deserializes it into four streams of 10 Gb/s each while performing equalization. A clock divider is used to generate I and Q phases of the 10-GHz clock to drive the quarter-rate latches L_1-L_8. Since, the outputs of these latches are in RZ form, these are converted to NRZ form using an on-chip RZ-NRZ conversion circuitry [42].

5.7   Experimental Results

The equalizer has been fabricated in TSMC's 45-nm digital CMOS process. Shown in Fig. 5.28, the active area of the die measures 100 μm × 200 μm. The equalizer draws 9.2 mW from a 1-V supply, with 2 mW consumed by the CTLE, 3.3 mW by the DTLE + summers + latches, 0.526 mW by the RZ-NRZ conversion, and 3.4 mW by the divide-by-2 circuit. All the measured results re-
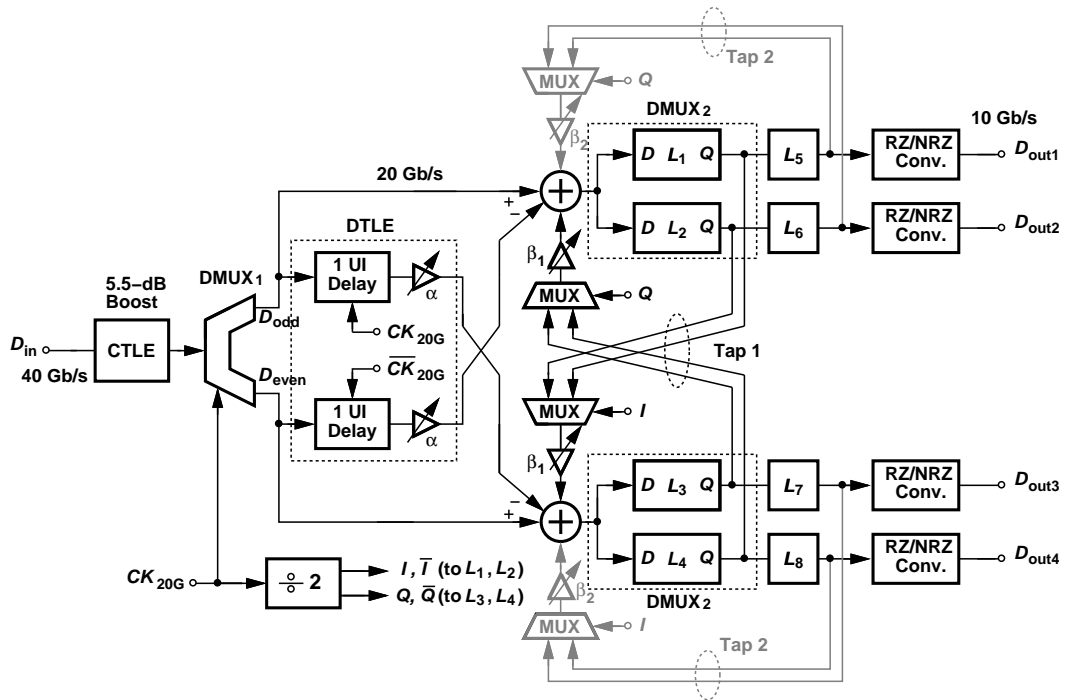
Figure 5.27: Proposed equalizer architecture.

ported here are for a channel loss of 20 dB at 20 GHz for a 40-Gb/s data rate [black plot in Fig. 5.28(b)], unless mentioned otherwise.

Figure 5.29 shows the setup to measure BER. The chip is directly mounted on a printed-circuit board, but the high-speed lines are carried through probes. An RF generator (Agilent E8257D) drives four PRBS generators (three Centellax TG2P1A and one Centellax TG1B1A). The 10-Gb/s outputs of each of the PRBS generators are multiplexed to form a 40-Gb/s stream. The 20-GHz clock for the multiplexer is generated from another signal generator. Our chip receives a 20-GHz differential clock from a third signal generator with a Balun (HL9402) shown on the bottom of Fig. 5.29. All three generators are mutually locked by connecting their 10-MHz references. One quarter-rate output of our chip is then returned to the BERT receiver (Centellax TG1B1-A) for BER measurement. In order to plot the bathtub curve, the internal phase of the signal generator connected to
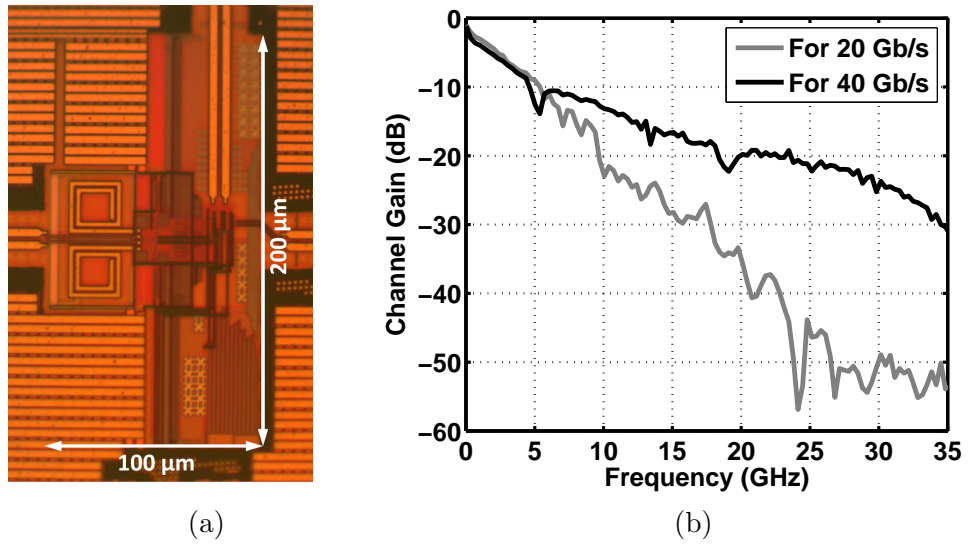
Figure 5.28: (a) Equalizer die photograph, (b) measured frequency response of lossy channels used for 20-Gb/s and 40-Gb/s data rates.
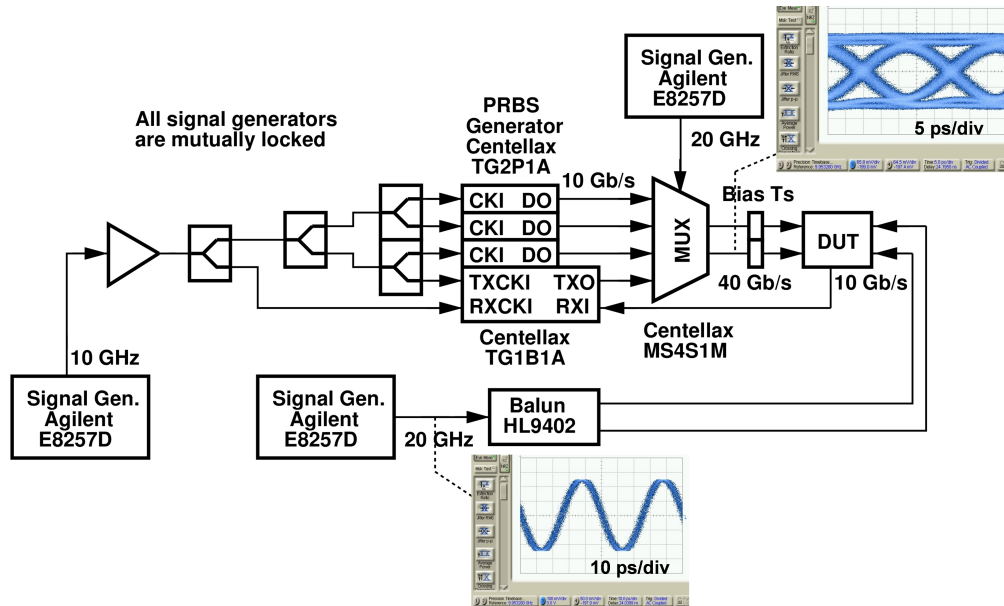
Figure 5.29: Test setup to measure BER.

our chip is varied and the BER is monitored.

Figure 5.30 shows the eye diagrams of the received 40-Gb/s data and one of the quarter-rate outputs of the equalizer. Figure 5.31 plots the bathtub curve for 40-
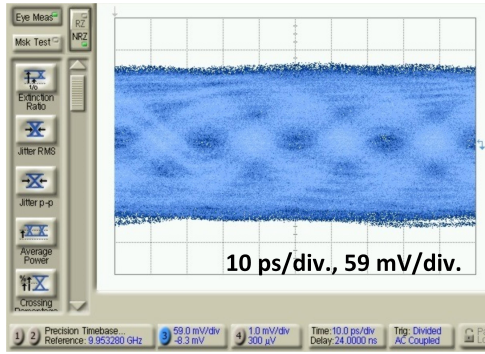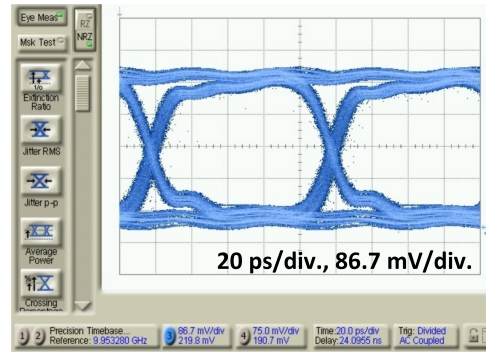
Figure 5.30: Measured eye diagram of (a) equalizer input at 40 Gb/s, and (b) equalized and demultiplexed output data at 10 Gb/s.
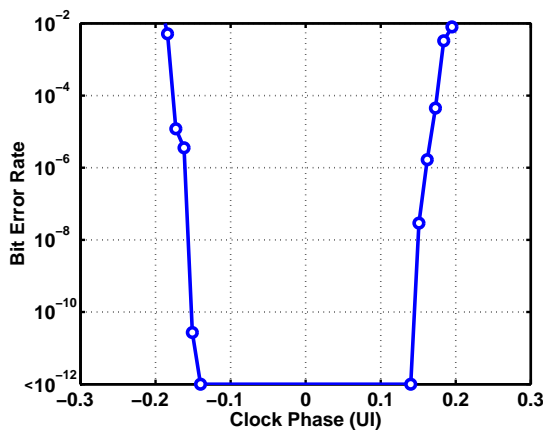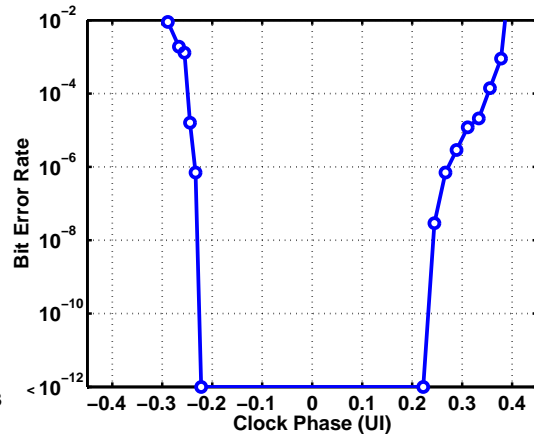
Figure 5.31: Measured bathtub curves at (a) 40 Gb/s, and (b) 20 Gb/s, with a 20-dB channel loss.

Gb/s and 20-Gb/s data rates, with nearly 20-dB channel loss at Nyquist in both cases. The measured frequency response of the channel used for 20-Gb/s data rate is shown in Fig. 5.28(b), in gray. At 40 Gb/s, a horizontal eye opening of 0.28 UI with a BER $< 10^{-12}$ is observed. We should remark that the PRBS generator output jitter is equal to 8 ps$_{pp}$ (0.32 UI) as seen in Fig. 5.29. Even though this generator and the DFE clock are mutually locked, the PRBS generator jitter

substantially degrades the measured bathtub curve width. This is because the locking occurs by sharing the 10-MHz references of the generators, which creates little correlation between their jitters. At 20 Gb/s, in Fig. 5.31(b), the horizontal eye opening is 0.44 UI, which illustrates the scaling property of charge-steering circuits. Since these circuits operate based on finite-time charging, they can easily scale to lower data rates with a linear reduction in their power consumption.

Table 5.1: Performance summary and comparison to prior art

Reference	Hsieh VLSI 2009	Toifl VLSI 2012	Lu ISSCC 2013	Manian CICC 2014	This Work
Data Rate (Gb/s)	40	32	66	32	40
Architecture	1–tap DFE	CTLE + 15–tap DFE	3–tap DFE	CTLE + 1–tap DFE	CTLE + DTLE + 2–tap DFE
DFE Clocking	Full–Rate	Quarter–Rate	Half–Rate	Full–Rate	Half–Rate/ Quarter–Rate
Channel Loss @ Nyquist	15 dB	36 dB	NA	18 dB	20 dB
BER/ Eye Opening	$<10^{-11}$/ NA	$<10^{-12}$/ 19% UI	$<10^{-12}$/ 60% UI	$<10^{-12}$/ 44% UI	$<10^{-12}$/ 28% UI
Supply (V)	1.2	1.15	1.2	0.73	1.0
Power (mW)	45	97.6	46	9.3	9.2
Power Efficiency (pJ/bit)	1.125	3.05	0.697	0.29	0.23
Area (mm^2)	0.05	0.018	0.00165	0.068	0.02
Technology	65–nm CMOS	32–nm SOI CMOS	65–nm CMOS	45–nm CMOS	45–nm CMOS

Table 5.1 summarizes the performance of state-of-the-art equalizers in the range of 32 Gb/s to 66 Gb/s. Our 40-Gb/s equalizer compensates a channel loss of 20 dB at Nyquist while achieving a horizontal eye opening of 0.28 UI with a BER $< 10^{-12}$. This amounts to a power efficiency of 0.23 pJ/bit which is better than other equalizers in the vicinity of 40 Gb/s in Table 5.1.

CHAPTER 6

A 40-Gb/s 14-mW CMOS Wireline Receiver

Reaching a power efficiency of 1 mW/Gb/s has proven difficult for wireline transceivers operating at tens of gigabits per second. At 40 Gb/s, recent receivers consume from 150 mW [4] to 1 W [5]. This chapter describes a receiver that achieves a tenfold reduction in power and an efficiency of 0.35 mW/Gb/s.

6.1 Minimalist Approach

An innovative aspect of the proposed receiver is our "minimalist" approach, which recognizes that every additional stage in the data or clock path consumes more power and limits the bandwidth. The minimalist mentality avoids multiple stages in the front-end CTLE, quadrature oscillators[1] in the CDR circuit, clock or data buffers, or phase interpolation[2]. Moreover, we share building blocks among different functions so as to reduce the number of current paths between V_{DD} and ground. Using charge-steering techniques extensively, the receiver contains only a few static bias currents adding up to about 6 mA. The minimalist approach also leads to a small footprint, about 110 μm × 175 μm, for the entire receiver,

---

[1]As described in Chapter 3, quadrature oscillators typically have a higher phase noise for a given power consumption as compared to a single oscillator.

[2]A phase-interpolation-based CDR, such as [48], typically consumes more power than an analog-PLL-based CDR such as the one described in this chapter.

making it possible to design a multi-lane system in a small area and with short interconnects.

6.2 Conceptual Receiver Architecture

Figure 6.1 conceptually depicts the receiver architecture, where overlapping boundaries indicate hardware sharing between the functions. The receiver consists of a CTLE, a CDR circuit, a DFE, a DTLE, and a 1-to-4 deserializer. Employing a single differential pair, the CTLE consumes 2 mW but provides only 5.5 dB of boost at the Nyquist rate; another 5.4 dB is created by the DTLE at a cost of 0.3 mW, as described in Chapter 5. The CDR runs at half rate and shares latches with the DFE and the deserializer. The CDR output clock frequency is also divided by 2 to generate quadrature phases at 10 GHz necessary for the second-rank latches in the DFE and the deserializer. Figure 6.1 exemplifies how minimalist design produces "growing" returns: the compact architecture contains no high-speed data interconnects longer than 25 μm, avoiding the need for buffers, inductive peaking (except for the CTLE), etc.

While our three principles of minimalist design, charge steering, and hardware sharing in Fig. 6.1 are attractive for a tenfold reduction in power, they also present their own challenges: (a) Full-rate operation becomes very difficult, (b) half-rate operation doubles the load capacitance that the CDR and DFE present to the CTLE, (c) the merged CDR/DFE topology in [49] is difficult to be implemented at 40 Gb/s because of its full-rate architecture and the resulting power-hungry clock buffer design, (d) the merged CDR/DFE topology in [4] cannot operate with charge steering or without quadrature oscillator phases, (e) the bufferless VCO frequency shifts if the DFE is turned off during lock acquisition, and (f) the CDR phase detector (PD) cannot be placed after the (charge-steering) DFE
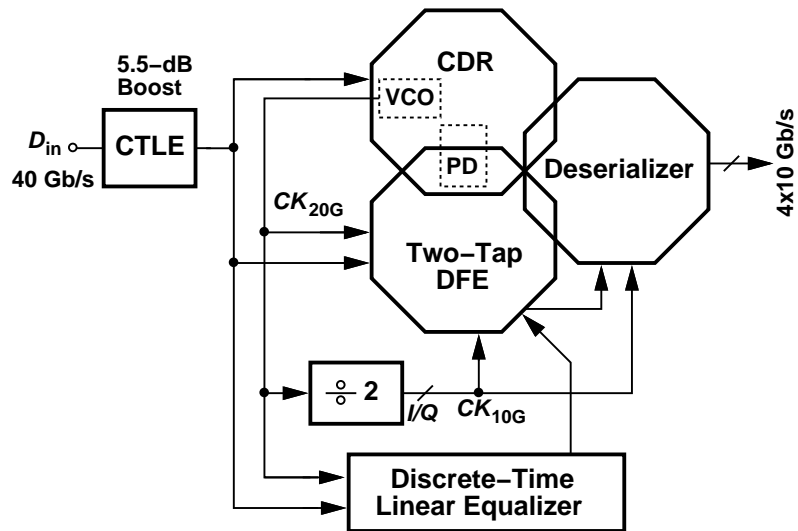
Figure 6.1: Conceptual receiver architecture.

summer because the summer output is precharged to V_{DD} for half of the clock cycle. These considerations require that the receiver be specifically architected to accommodate our three design principles.

6.3 Proposed Phase Detector

The sixth issue, namely, the return-to-zero nature of the summer output, complicates hardware sharing between the CDR and the DFE. Noting that a half-rate charge-steering phase detector must take samples of the full-rate data by means of six latches [42], we seek interfaces within the DFE that can provide such samples. The proposed circuit is shown in Fig. 6.2, where the blocks in gray belong to the DFE. For half-rate operation, DMUX_1 and the summing circuits also operate as sampling elements. Thus, only two more latches, L_a and L_b, are necessary for unambiguous phase detection. As shown in Fig. 6.2, the XOR of D_{odd} and D_{even} provides the phase error information, V_{err}, between the full-rate data and the 20-GHz clock but with dependence on the data pattern. On the other hand,

the XOR of samples X_1, X_2, Y_1, and Y_2 generates a fixed-width reference pulse on V_{ref} for every data transition, eliminating the data dependence from the final output, I_{out}. Note that the G_m stage measures the areas under V_{err} and V_{ref} and need not operate at high speeds.
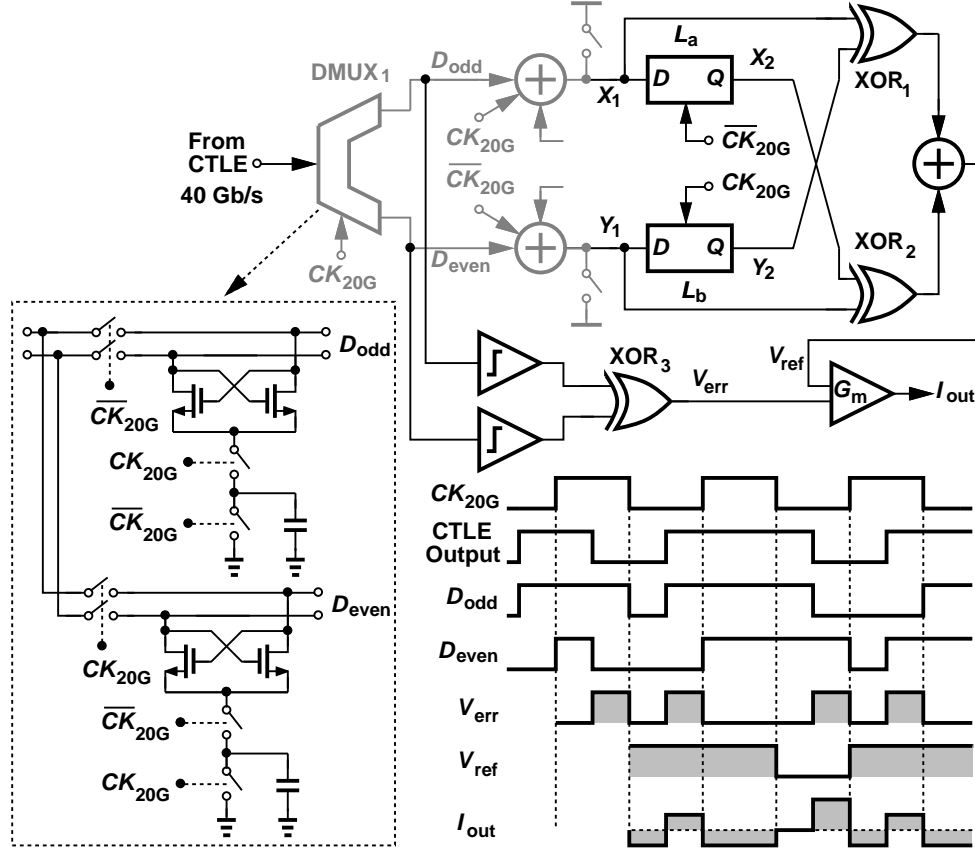
Figure 6.2: Proposed phase detector.

The PD topology of Fig. 6.2 faces an issue arising from the CTLE's limited boost factor. Since a lossy channel yields heavily attenuated 1010 swings in D_{odd} and D_{even}, and since DMUX$_1$ must be sufficiently linear for the equalization operation, XOR$_3$ can produce an output inconsistent with those of XOR$_1$ and XOR$_2$, which are driven by large data swings. To resolve this difficulty, two limiters realized as single differential pairs precede XOR$_3$. Owing to the voltage

gain provided by DMUX$_1$ ($\approx$ 6 dB), each differential pair can act as a limiting amplifier at 20 Gb/s with a tail current of 0.2 mA.

In steady-state operation, the CDR is locked and the DFE (described below) produces properly-equalized data at the summer outputs, X_1 and Y_1, in Fig. 6.2. In the presence of a lossy channel, therefore, the BER of X_1 and Y_1 (or X_2 and Y_2) is much lower than that of D_{odd} and D_{even}. This points to some inconsistency between the output of XOR$_3$ and those of XOR$_1$ and XOR$_2$, which ultimately manifests itself as a static phase error within the CDR loop. Nevertheless, so long as the BER in D_{odd} and D_{even} is less than about 10^{-2}, the resulting phase error is negligible.

6.4   Proposed Receiver Architecture

Figure 6.3 shows the proposed half-rate/quarter-rate CDR and DFE details. Highlighted in gray, the CDR loop consists of DMUX$_1$, the summers, latches L_a and L_b, the XOR gates, the loop filter and the VCO. The DFE comprises DMUX$_1$, the summers and the feedback taps formed by DMUX$_{2,3}$ and MUX$_{1-4}$. The DTLE injects delayed and scaled copies of D_{odd} and D_{even} into the summers, realizing a transfer function given by $1 - \alpha z^{-1}$, as described in Chapter 5, and hence a boost factor of $(1 + \alpha)/(1 - \alpha)$.

The intertwined CDR and DFE loops in Fig. 6.3 can fight and fail to converge. The operation thus begins by setting the DFE and DTLE tap coefficients to zero and the CTLE boost factor to its maximum value. The gray path detects the phase error between the data and the VCO output, delivering a proportional current to the loop filter and driving the oscillator toward 20 GHz. Despite the heavy intersymbol interference at the CTLE output, the CDR locks, as shown
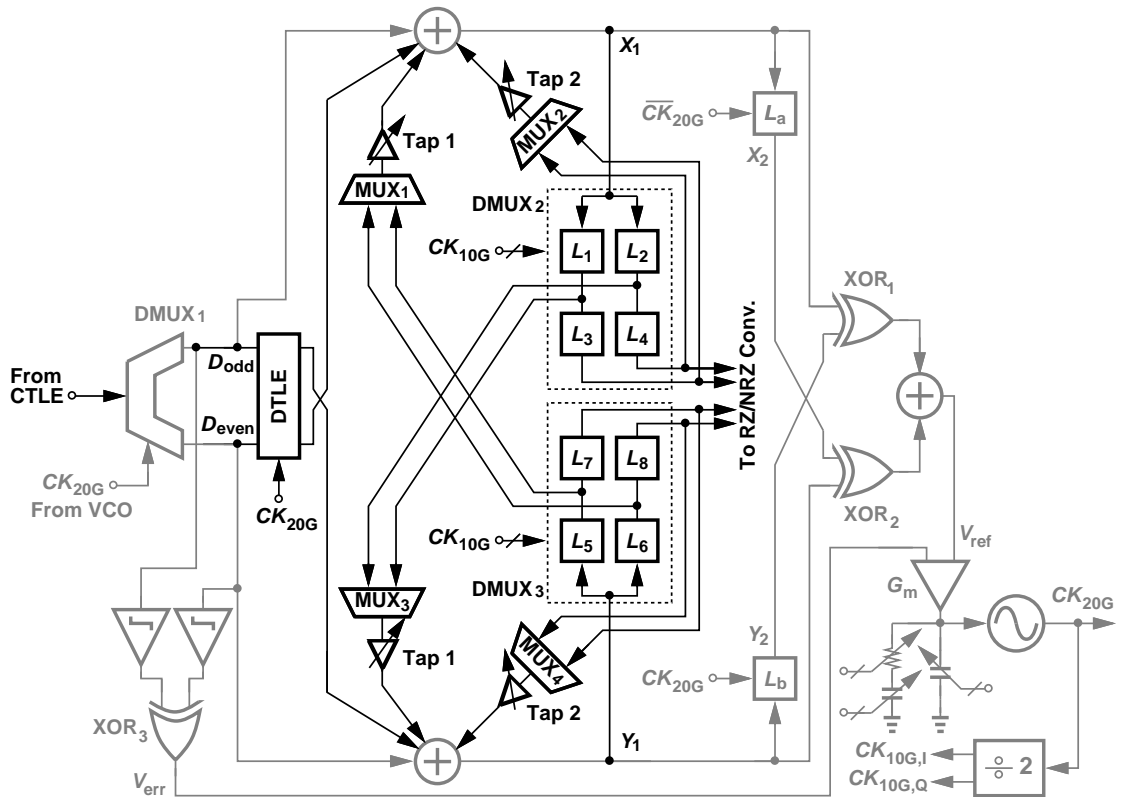
Figure 6.3: Proposed receiver architecture with half-rate/quarter-rate CDR and DFE.

in Fig. 6.4 because data patterns having several consecutive ONEs or ZEROs make full transitions and provide sufficient phase information. Moreover, with about 10.6 dB of voltage gain through $DMUX_1$ and the limiters, the transitions presented to XOR_3 become sharp enough to produce proper phase error. The CDR takes approximately 100 ns to lock, after which the CTLE boost and the DFE and DTLE tap coefficients are adapted to complete the equalization. (In this prototype, the adaptation is done manually through a serial bus.)

If the DFE and DTLE tap coefficients are correctly set initially, the CDR still locks in about 100 ns with an open eye at the summing junction as shown in Fig. 6.5. This means that the CDR can lock with any value of the DFE and DTLE coefficients. In order for the automatic coefficient-adaptation algorithm
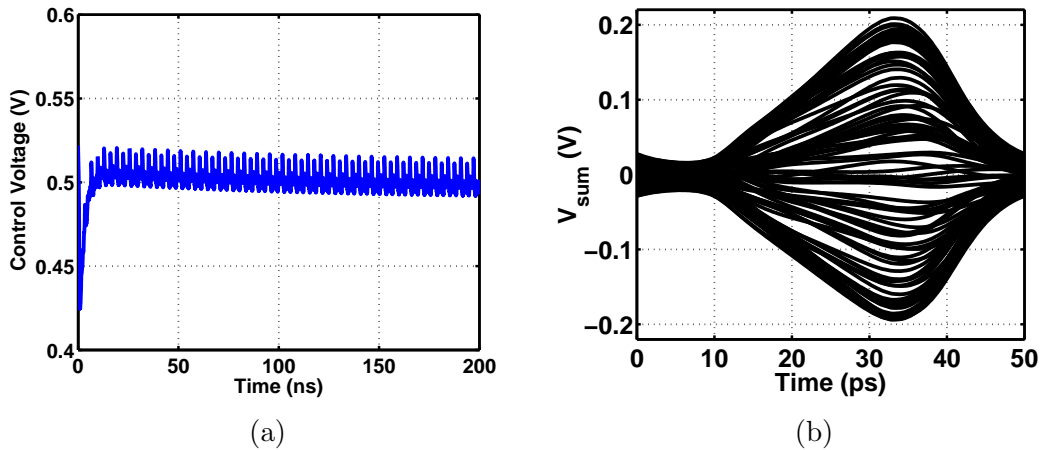
Figure 6.4: DFE/DTLE tap-coefficients set to zero: (a) Control voltage transient, (b) eye diagram at the summing junction after the CDR has locked.

to work, the updating process must be slow enough such that the CDR is allowed to lock after every update.
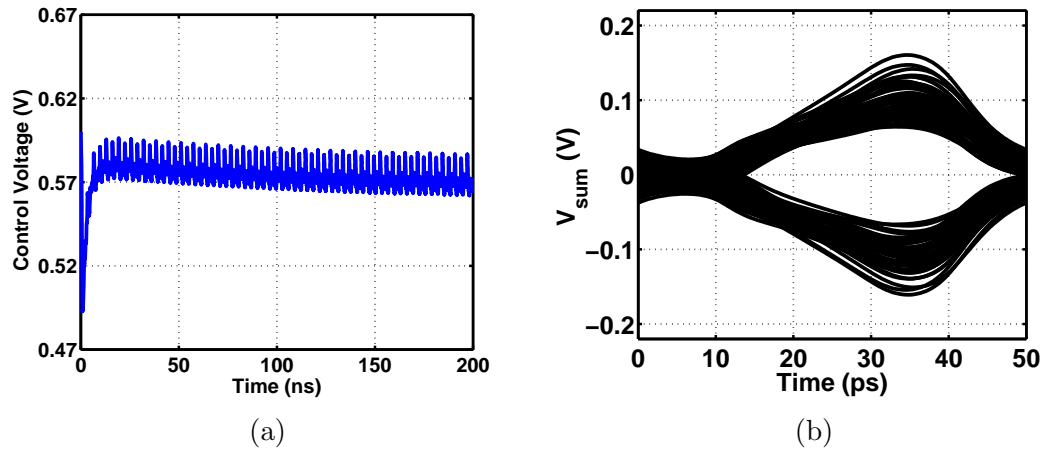
Figure 6.5: DFE/DTLE tap-coefficients set correctly initially: (a) Control voltage transient, (b) eye diagram at the summing junction after the CDR has locked.

The DFE in Fig. 6.3 demultiplexes the half-rate data at X_1 (or Y_1) by another factor of 2 using L_1-L_4 (or L_5-L_8), multiplexes the results and injects a fraction thereof into the summers, thus realizing the first and second feedback taps.

6.5 Building Blocks

The CTLE, DMUX$_1$, the DTLE and DFE latches are designed similar to those in Chapter 5. This section presents the circuit-level implementations of the CDR building blocks in 45-nm CMOS technology. Latches L_a and L_b are designed as the charge-steering latch with NMOS cross-coupled pairs as in Fig. 5.21.

6.5.1 Voltage Controlled Oscillator (VCO)

In the receiver of Fig. 6.3, the VCO drives DMUX$_1$, the summers, L_a and L_b, the DTLE, and the divide-by-2 circuit, which, along with the interconnects, present a single-ended capacitance of about 180 fF. If driven by buffers, two such capacitances (for CK and $\overline{CK}$) would lead to a power consumption, $2fCV_{DD}^2$, of 7.2 mW, as shown in Fig. 6.6, which is more than half of the overall receivers budget (unless the buffers utilize inductive loads). More fundamentally, it seems more efficient to burn any power that we have to in the VCO rather than in the buffers.
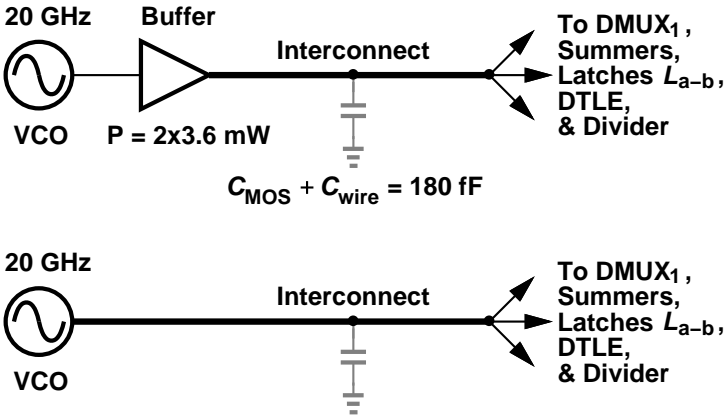
Figure 6.6: VCO: To buffer or not to buffer?

In this work, the VCO employs a 0.4-nH differential inductor with complemen-
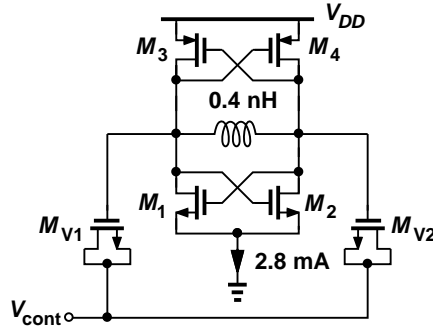
Figure 6.7: VCO implementation.

tary cross-coupled pairs, as shown in Fig. 6.7. The bias current, I_{SS}, is dictated by the voltage swing given by

$$V_{swing} = \frac{4}{\pi} I_{SS} R_P, \tag{6.1}$$

where R_P is the effective parallel resistance of the LC tank. In this design, I_{SS} is chosen to be 2.8 mA so as to deliver nearly rail-to-rail swings without buffers.
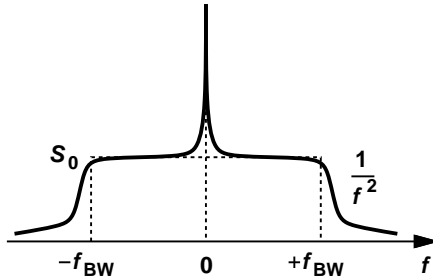
Figure 6.8: Locked phase noise profile of the VCO.

Is the phase noise of this VCO good enough? The locked phase noise profile can be approximated by the curve shown in Fig. 6.8, and the integrated rms jitter t_σ can be calculated using the formula

$$t_\sigma \approx \frac{T_{CK}}{2\pi} \sqrt{4 S_0 f_{BW}}. \tag{6.2}$$

For a jitter specification of 0.5 ps$_{\text{rms}}$, we can find the phase noise in the plateau to be -103 dBc/Hz whereas the phase noise of the designed oscillator is around

-127 dBc/Hz from simulations. This means that the bias current is dictated by the voltage swing rather than the phase noise.

The size of the varactors in Fig. 6.7 is chosen to be 1 μm/200 nm in order to achieve a K_{VCO} of approximately 1 GHz/V. The simulated VCO tuning curves are shown in Fig. 6.9. To achieve coarse tuning, four single-ended 10 fF capacitors are digitally switched in or out of the VCO.
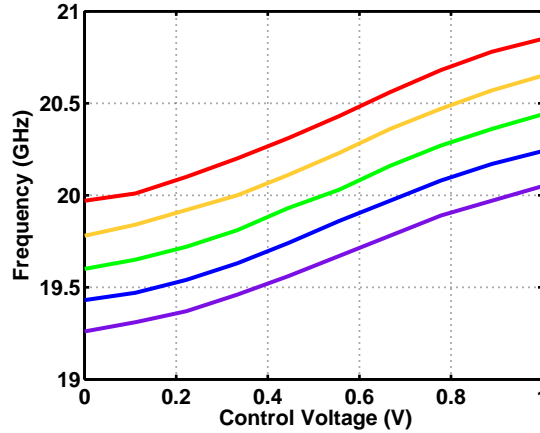
Figure 6.9: VCO tuning curves.

6.5.2   Vernier Charge Delivery

The vernier charge delivery described in Chapter 5 (Fig. 6.10) is also used in this design. It is important to note that the bufferless VCO frequency shifts when the DFE/DTLE tap coefficients are changed. In Fig. 6.10, if C_T is programmed to change the tap coefficient, the total capacitance seen by CK or $\overline{CK}$ changes slightly. To avoid changing this capacitance by a large ratio, a vernier technique is used, whereby capacitor C_X, precharged to V_{DD} is allowed to charge-share with C_X, thus lending the capability of changing the tap coefficient without having to change C_X by a large value.

The issue matters only during startup or during DFE adaptation. In steady

state, the taps are frozen.
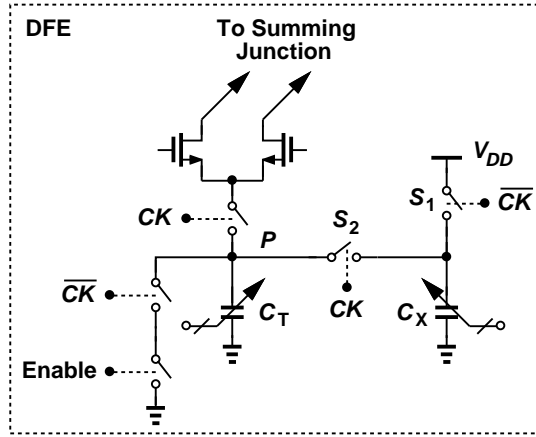
Figure 6.10: Vernier charge delivery.

6.5.3 XOR and V/I Converter
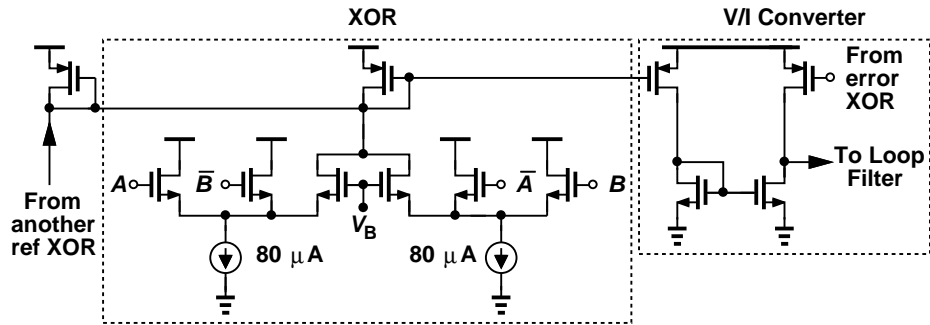
Figure 6.11: XOR and V/I converter.

The symmetric XOR [50] and the G_m stage (V/I converter) for the phase de-tector are implemented as shown in Fig. 6.11. Each XOR consumes only 160 μA, and need not have a high output bandwidth as the G_m stage measures only the average values of V_{err} and V_{ref}. The addition of the output of the two XORs in Fig. 6.2 is performed by adding their output currents as shown in Fig. 6.11.

Since the output currents of error and reference XORs can be directly sub-

tracted, the V/I converter circuit is implemented as a current mirror arrangement of Fig. 6.11, thus getting rid of the tail current source in the V/I converter of [29]. This increases the available swing on the control line of the VCO by one overdrive.

6.5.4 Programmable CDR Loop Filter
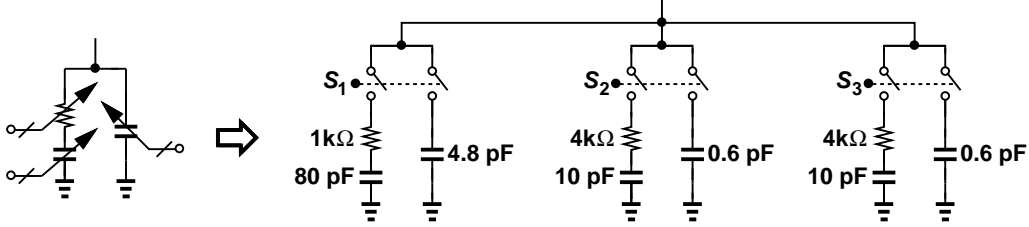
Figure 6.12: Programmable CDR loop filter.
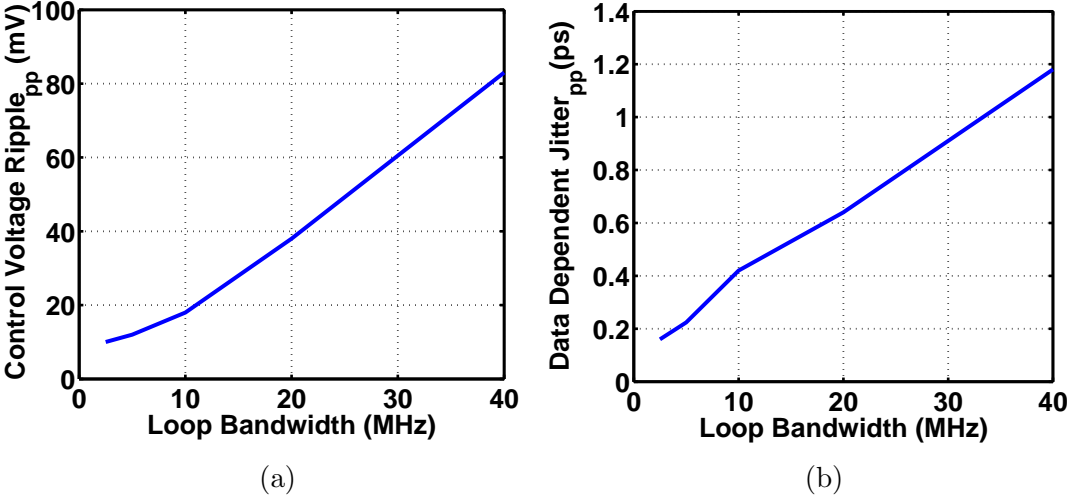
Figure 6.13: Effect of loop bandwidth on control voltage ripple and data-dependent jitter.

The CDR loop bandwidth can be varied from 4 MHz to 20 MHz by means of programmable loop filter components, shown in Fig. 6.12. The loop filter has been implemented on chip in this design. It is desirable to maximize the

bandwidth so as to both suppress the VCO phase noise and improve the jitter tolerance, but the upper bound is dictated by the data-dependent jitter resulting from the ripple on the control voltage. Figure 6.13 shows the effect of increasing the loop bandwidth on the peak-to-peak ripple on the control voltage and the data-dependent jitter. With a 20-MHz bandwidth, this jitter is still negligible compared to the VCO contribution.

6.6 Experimental Results
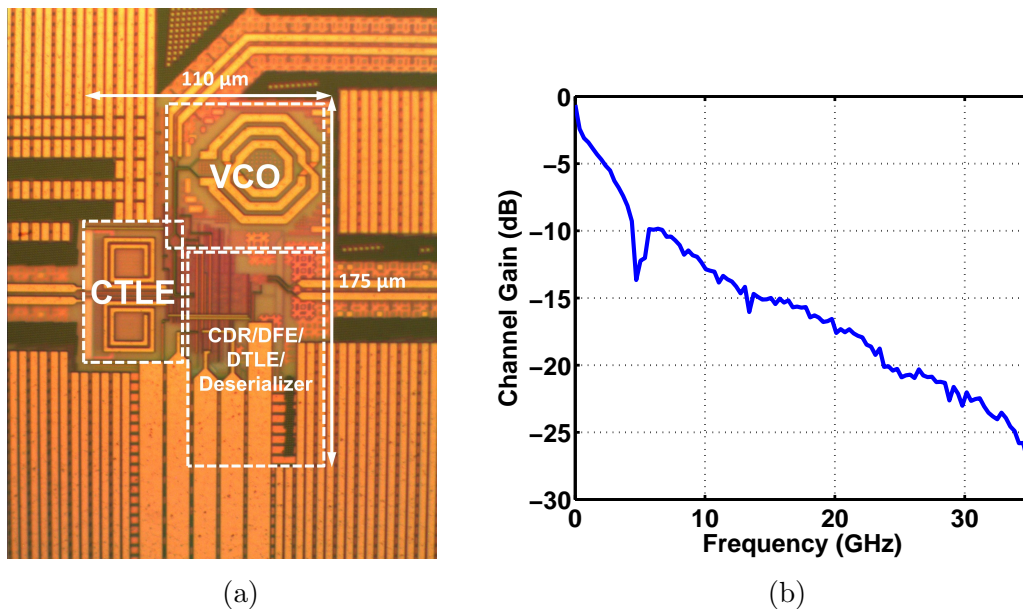
Figure 6.14: (a) Receiver die photograph, (b) measured channel frequency response (additional 1-dB insertion loss of probes).

The overall receiver has been fabricated in TSMC's 45-nm digital CMOS technology and tested with a 1-V supply on a probe station. Figure 6.14 (a) shows the die photograph. Of the total power of 14 mW, 2 mW are consumed by the CTLE, 5.3 mW by the PD, the summers, the DTLE, and the latches, 2.8 mW by the VCO, 3.4 mW by the divide-by-2 circuit, and 0.53 mW by the

RZ-NRZ conversion stages in the 10-Gb/s data paths. All measurements have been carried out with a channel loss [Fig. 6.14(b)] of 18.6 dB at Nyquist with a PRBS length of $2^7 - 1$.
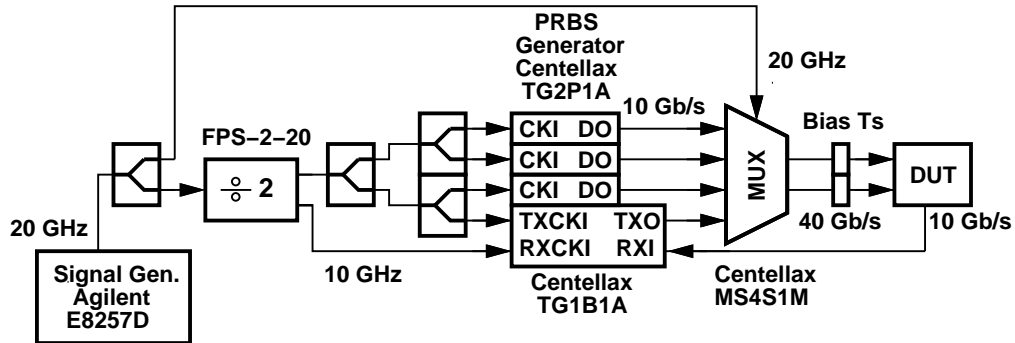
Figure 6.15: Test setup to measure BER.

| (a) | (b) |

Figure 6.16: Measured eye diagram of (a) equalizer input at 40 Gb/s (10 ps/div., 61 mV/div.), and (b) equalized and demultiplexed output data at 10 Gb/s (20 ps/div., 97.6 mV/div.).

The test setup to measure BER is shown in Fig. 6.15. This is similar to the test setup in Chapter 5, except for a divide-by-2 circuit (FPS-2-20) used to provide the 10-GHz clocks to the PRBS generators/receiver. This is required to synchronize the MUX clock with the PRBS generators' so as to produce accurate phase modulation required to measure jitter transfer and jitter tolerance curves.
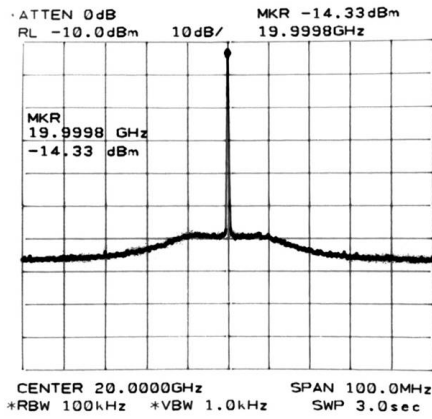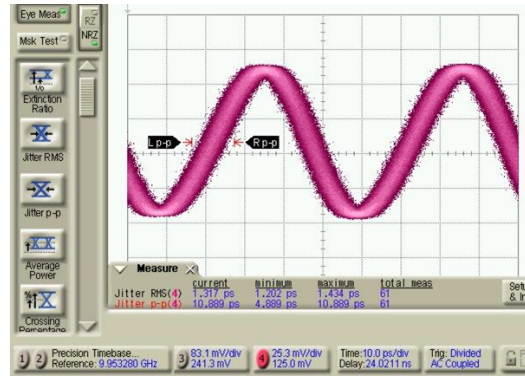
(a)   (b)

Figure 6.17: Recovered clock: (a) Spectrum, (b) waveform (10 ps/div., 25.3 mV/div.).
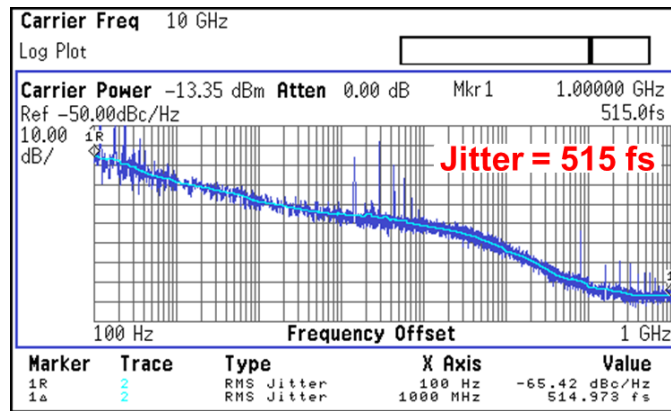
Figure 6.18: Phase noise of recovered clock at 10 GHz.

Figure 6.16 shows the measured channel output at 40-Gb/s and the recovered data eye at 10-Gb/s. The recovered clock's spectrum and waveform is shown in Fig. 6.17. The recovered clock is divided by 2 to measure its phase noise at 10 GHz, shown in Fig. 6.18. The 10-GHz clock exhibits an rms jitter of 0.515 ps, when integrated from 100 Hz to 1 GHz.

Figure 6.19 shows the test setup to measure jitter transfer [51]. In order to create a phase modulation, two RF signals at frequencies $\omega_c = 20$ GHz and $\omega_c + \omega_m$, generated using Agilent E8257D's, are summed to create the spectrum
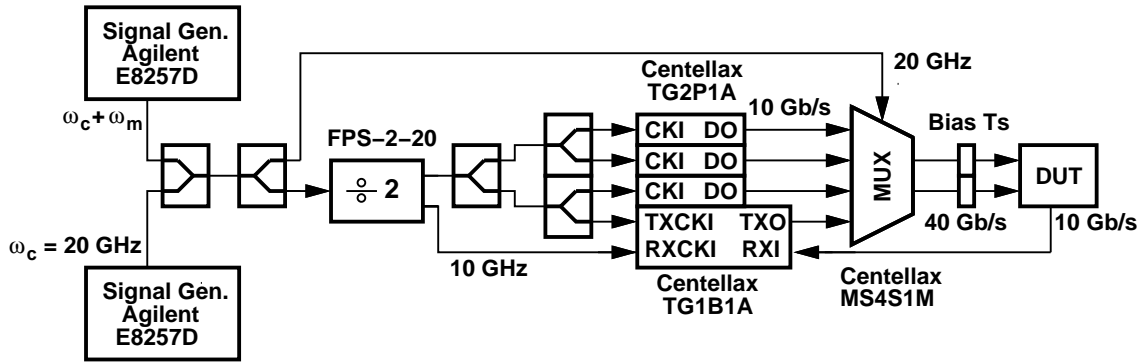
Figure 6.19: Test setup to measure jitter transfer.
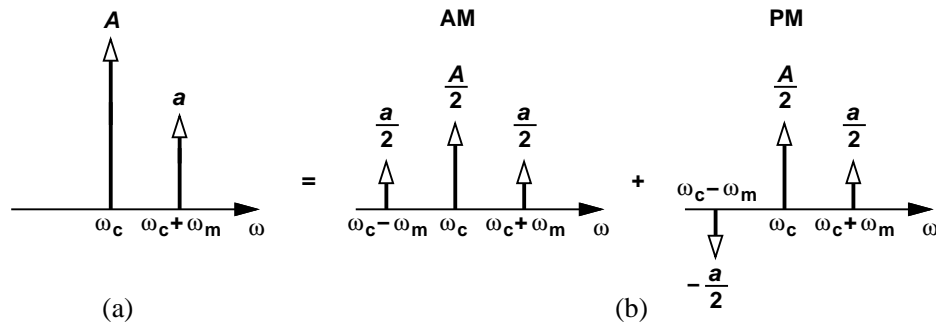
Figure 6.20: Equivalence of a single sideband to the sum of AM and PM.

shown in Fig. 6.20(a). This spectrum can be viewed as a combination of amplitude modulation (AM) and phase modulation (PM) components, as depicted in Fig. 6.20(b). The limiting operation of the 4:1 MUX and the dividers eliminate AM, preserving only PM. Thus, we can generate arbitrarily high jitter frequencies, ω_m, with low jitter amplitudes. It is not required to know the actual amplitude of the applied input jitter. We know that at jitter frequencies much lower than the loop bandwidth, the jitter transfer is 0 dB. The output jitter measured for one such frequency is chosen as reference and the jitter components for all higher frequencies are measured relative to it.

Figure 6.21 shows the test setup to measure jitter tolerance [51]. An RF generator (jitter at ω_m) is connected to the PM input of another RF generator
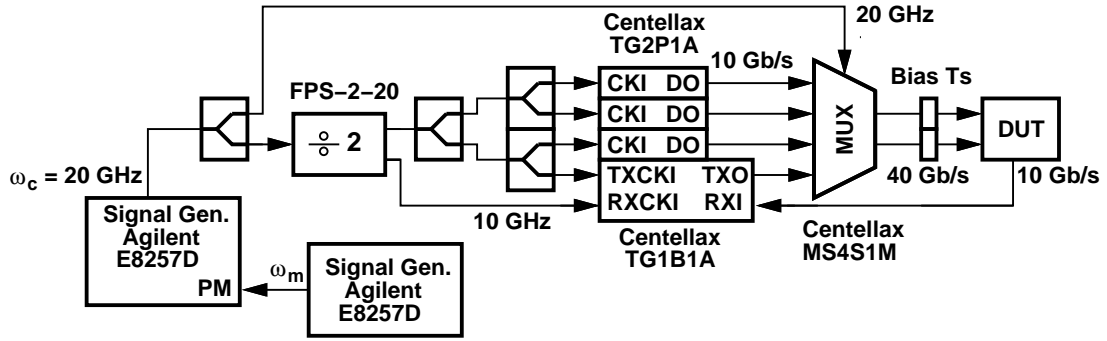
Figure 6.21: Test setup to measure jitter tolerance.

(clock at ω_c). The jitter amplitude, at a frequency ω_m, is increased slowly until the BER rises above 10^{-12}, to obtain the jitter tolerance at ω_m.
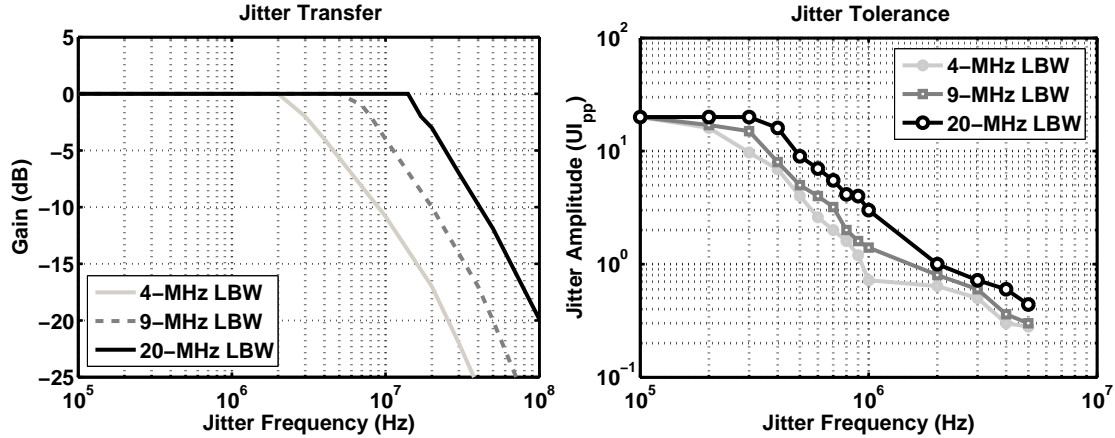
Figure 6.22: Measured jitter transfer and tolerance curves.

Figure 6.22 plots the measured jitter transfer and tolerance for three different CDR loop bandwidths, revealing a tolerance as high as 0.45 UI$_{\text{pp}}$ at 5 MHz, for a BER $< 10^{-12}$. It is important to note that the PRBS generator and the external clock contribute about 8 ps$_{\text{pp}}$ which negatively impacts the jitter tolerance measurement.

Table 6.1 summarizes the performance of our prototype and compares it with that of the state of the art. We achieve a tenfold reduction in power consumption

Table 6.1: Performance summary and comparison to prior art

Reference	Hsieh VLSI 2011	Chen JSSC Mar. 2012	Raghavan JSSC Dec. 2013	This Work
Data Rate (Gb/s)	40	40	40	40
Supply (V)	1.2 for DFE/CDR, 1.5 for CTLE	1.6	1	1
Channel Loss at Nyquist (dB)	23.5	19	>21	18.6
Bit Error Rate	$<10^{-12}$	$<10^{-12}$	$<10^{-12}$	$<10^{-12}$
Power (mW)	150	520	1050[†]	14
Power Efficiency (pJ/bit)	3.75	13	26.25	0.35
Recovered Clock Jitter (ps)	6.8 pp	0.319 rms	-	0.515 rms
Jitter Tolerance	-	≈ 0.65 UI_{pp} at 10 MHz	0.95 UI_{pp} at 10 MHz[‡]	0.45 UI_{pp} at 5 MHz
Area (mm^2)	0.278	1.1475*	3.9*	0.019
Technology	65–nm CMOS	65–nm CMOS	40–nm CMOS	45–nm CMOS

* Includes pads   **[†] Includes SFI–5.2 TX; 350 mW for line–side RX**
[‡] Measured for BER $=10^{-9}$

as compared to prior art through the use of a minimalist approach in our design, hardware sharing, and charge-steering techniques.

CHAPTER 7

Conclusion

This work describes a number of architecture and circuit techniques that make it feasible for wireline receivers to achieve an efficiency of 0.35 mW/Gb/s at 40 Gb/s.

A full-rate equalizer architecture employing inductor nesting to save area and latch feedforward to improve the speed, has been presented in Chapter 4. It has been shown that the inductor coupling due to nesting could be exploited to shrink the area further. Also, it has been recognized that latch feedforward is equivalent to a high-frequency boost in the feedback tap. Such techniques afford operation at 32 Gb/s with 9.3 mW from a 0.73-V supply, producing an eye opening of 0.44 UI.

A discrete-time half-rate/quarter-rate equalizer architecture, extensively employing charge-steering techniques, has been described in Chapter 5. The concept and implementation of a DTLE has been introduced as an efficient means of creating a high-frequency boost of about 5.4 dB with only 0.3 mW. In addition, two new charge steering latch topologies with improved swing have been developed to afford operation at 40 Gb/s. This 40-Gb/s equalizer consumes 11.2 mW from a 1-V supply, while providing an eye opening of 0.28 UI.

A 40-Gb/s wireline receiver that achieves a tenfold reduction in power consumption as compared to prior art, has been introduced in Chapter 6. Designed

on the principles of minimalist approach, hardware sharing, and charge steering, this receiver consists of a one-stage CTLE, a half-rate CDR, a half-rate/quarter-rate DFE, a half-rate DTLE, and a 1-to-4 deserializer. This receiver achieves a BER $< 10^{-12}$ with a recovered clock jitter of 0.515 ps_{rms}, a jitter tolerance of 0.45 UI_{pp} at 5 MHz, while consuming only 14 mW from a 1-V supply.

The techniques introduced in this work can be extended to multi-level signaling, such as PAM-4. Charge-steering circuits may be used in the multiplexing stages preceding the final driver in serializing transmitters, to reduce its power consumption at high data rates. These techniques may also be utilized in building the front-end analog-to-digital converter (ADC) in an ADC-based wireline receiver.

References

[1] S. Chandramouli, "A novel analog decision-feedback equalizer in CMOS for serial 10-Gb/sec data transmission systems," Ph.D. dissertation, Georgia Institute of Technology, 2007.

[2] M. Pozzoni *et al.*, "A 12Gb/s 39dB loss-recovery unclocked-DFE receiver with bi-dimensional equalization," *IEEE ISSCC Dig. Tech. Papers*, pp. 164-165, Feb. 2010.

[3] S. Ibrahim and B. Razavi, "Low-power CMOS equalizer design for 20-Gb/s systems," *IEEE J. Solid-State Circuits*, vol. 46, no. 6, pp. 1321-1336, June 2011.

[4] C.-L. Hsieh and S.-I. Liu, "A 40Gb/s adaptive receiver with linear equalizer and merged DFE/CDR," *Symposium on VLSI Circuits Dig. of Tech. Papers*, pp. 208-209, June 2011.

[5] B. Raghavan *et al.*, "A Sub-2 W 39.8–44.6 Gb/s transmitter and receiver chipset with SFI-5.2 interface in 40 nm CMOS," *IEEE J. Solid-State Circuits*, vol. 48, no. 12, pp. 3219-3228, Dec. 2013.

[6] M.-S. Chen *et al.*, "A fully-integrated 40-Gb/s transceiver in 65-nm CMOS technology," *IEEE J. Solid-State Circuits*, vol. 47, no. 3, pp. 627-640, March 2012.

[7] B. Razavi, *Design of Integrated Circuits for Optical Communications*, McGraw-Hill, 2003.

[8] 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, Sept. 2007.

[9] M.-S. Chen and C.-K. K. Yang, "A 50-64 Gb/s serializing transmitter with a 4-tap, LC-ladder-filter-based FFE in 65 nm CMOS technology," *IEEE J. Solid-State Circuits*, vol. 50, no. 8, pp. 1903-1916, Aug. 2015.

[10] A. Momtaz and M. M. Green, "An 80 mW 40 Gb/s 7-tap T/2-spaced feedforward equalizer in 65 nm CMOS," *IEEE J. Solid-State Circuits*, vol. 45, no. 3, pp. 629-639, March 2010.

[11] T.-C. Lee and B. Razavi, "A 125-MHz CMOS mixed-signal equalizer for gigabit ethernet on copper wire," *Proc. IEEE Custom Integrated Circuits Conference (CICC)*, pp. 131-134, May 2001.

[12] J. E. Jaussi *et al.*, "8-Gb/s source-synchronous I/O link with adaptive receiver equalization, offset cancellation, and clock de-skew," *IEEE J. Solid-State Circuits*, vol. 40, no. 1, pp. 80-88, Jan. 2005.

[13] Y. Tomita *et al.*, "A 10-Gb/s receiver with series equalizer and on-chip ISI monitor in 0.11-µm CMOS," *IEEE J. Solid-State Circuits*, vol. 40, no. 4, pp. 986-993, April 2005.

[14] P. Staric and E. Margan, *Wideband Amplifiers*, Springer, 2015.

[15] B. Razavi, *RF Microelectronics*, Prentice Hall, 2011.

[16] S. Galal and B. Razavi, "Broadband ESD protection circuits in CMOS technology," *IEEE J. Solid-State Circuits*, vol. 38, no. 12, pp. 2334-2340, Dec. 2003.

[17] J. Paramesh and D.J. Allstot, "Analysis of the bridged T-coil circuit using the extra-element theorem," *IEEE Trans. Circuits and Systems - Part II*, vol. 53, no. 12, pp. 1408-1412, Dec. 2006.

[18] S. Galal and B. Razavi, "10-Gb/s limiting amplifier and laser/modulator driver in 0.18-µm CMOS technology," *IEEE J. Solid-State Circuits*, vol. 38, no. 12, pp. 2138-2146, Dec. 2003.

[19] D. D. Falconer, "Adaptive equalization of channel nonlinearities in QAM data transmission systems," *Bell System Technical Journal*, vol. 57, no. 7, pp. 2589-2611, Sept. 1978.

[20] B. Kim *et al.*, "A 10-Gb/s compact low-power serial I/O with DFE-IIR equalization in 65-nm CMOS," *IEEE J. Solid-State Circuits*, vol. 44, no. 12, pp. 3526-3538, Dec. 2009.

[21] S. Shahramian and A. Chan Carusone, "A 0.41 pJ/bit 10 Gb/s hybrid 2 IIR and 1 discrete-time DFE tap in 28 nm-LP CMOS," *IEEE J. Solid-State Circuits*, vol. 50, no. 7, pp. 1722-1735, July 2015.

[22] B. Razavi, *Monolithic Phase-Locked Loops and Clock Recovery Circuits: Theory and Design*, IEEE Press, 1996.

[23] S. K. Shanmugam, *Digital and Analog Communication Systems*, New York: Wiley & Sons, 1979.

[24] J. Lee *et al.*, "Design of 56 Gb/s NRZ and PAM4 serdes transceivers in CMOS technologies," *IEEE J. Solid-State Circuits*, vol. 50, no. 9, pp. 2061-2073, Sept. 2015.

[25] C. R. Hogge, "A self correcting clock recovery circuit," *IEEE Trans. on Electron Devices*, vol. 32, no. 12, pp. 2704-2706, Dec. 1985.

[26] T. H. Lee and J. F. Bulzacchelli, "A 155-MHz clock recovery delay- and phase-locked loop," *IEEE J. Solid-State Circuits*, vol. 27, no. 12, pp. 1736-1746, Dec. 1992.

[27] J. D. H. Alexander, "Clock recovery from random binary data," *Electronics Letters*, vol. 11, pp. 541-542, Oct. 1975.

[28] J. Lee, K. S. Kundert and B. Razavi, "Analysis and modeling of bang-bang clock and data recovery circuits," *IEEE J. Solid-State Circuits*, vol. 39, no. 9, pp. 1571-1580, Sept. 2004.

[29] J. Savoj and B. Razavi, "A 10-Gb/s CMOS clock and data recovery circuit with a half-rate linear phase detector," *IEEE J. Solid-State Circuits*, vol. 36, no. 5, pp. 761-767, May 2001.

[30] M. Ramezani and C. A. T. Salama, "Analysis of a half-rate bang-bang phase-locked-loop," *IEEE Trans. Circuits and Systems - Part II*, vol. 49, no. 7, pp. 505-509, July 2002.

[31] B. Razavi, "Design of millimeter-wave CMOS radios: A tutorial," *IEEE Trans. Circuits and Systems - Part I*, vol. 56, no. 1, pp. 4-16, Jan. 2009.

[32] Y. M. Greshishchev and P. Schvan, "SiGe clock and data recovery IC with linear-type PLL for 10-Gb/s SONET application," *IEEE J. Solid-State Circuits*, vol. 35, no. 9, pp. 1353-1359, Sept. 2000.

[33] V. Stojanovic *et al.*, "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, April 2005.

[34] C.-L. Hsieh and S.-I. Liu, "A 40Gb/s decision feedback equalizer using back-gate feedback technique," *Symposium on VLSI Circuits Dig. of Tech. Papers*, pp. 218-219, June 2009.

[35] J. Bulzacchelli *et al.*, "A 28Gb/s 4-tap FFE/15-tap DFE serial link transceiver in 32nm SOI CMOS technology," *IEEE ISSCC Dig. Tech. Papers*, pp. 324-326, Feb. 2012.

[36] B. Razavi, "The role of PLLs in future wireline transmitters," *IEEE Trans. Circuits and Systems - Part I*, vol. 56, no. 8, pp. 1786-1793, Aug. 2009.

[37] T. Toifl *et al.*, "A 3.1mW/Gbps 30Gbps quarter-rate triple-speculation 15-tap SC-DFE RX data path in 32nm CMOS," *Symposium on VLSI Circuits Dig. of Tech. Papers*, pp. 102-103, June 2012.

[38] K. Kaviani *et al.*, "A 27-Gb/s, 0.41-mW/Gb/s 1-tap predictive decision feedback equalizer in 40-nm low-power CMOS," *Proc. IEEE Custom Integrated Circuits Conference (CICC)*, pp. 1-4, Sept. 2012.

[39] J. W. Jung and B. Razavi, "A 25Gb/s 5.8mW CMOS equalizer," *IEEE J. Solid-State Circuits*, vol. 50, no.2, pp. 515-526, Feb. 2015.

[40] P. J. Lim and B. A. Wooley, "An 8-bit 200-MHz BiCMOS Comparators," *IEEE J. Solid-State Circuits*, vol. 25, no. 1, pp. 192-199, Feb. 1990.

[41] B. Razavi and B. A. Wooley, "Design techniques for high-speed, high-resolution comparators", *IEEE Journal of Solid-State Circuits*, vol. 27, no. 12, Dec. 1992.

[42] J. W. Jung and B. Razavi, "A 25-Gb/s 5-mW CMOS CDR/deserializer," *IEEE J. Solid-State Circuits*, vol. 48, no. 3, pp. 684-697, March 2013.

[43] B. Razavi, "Charge steering: A low-power design paradigm," *Proc. IEEE Custom Integrated Circuits Conference (CICC)*, pp. 1-8, Sept. 2013.

[44] B. Razavi, *Design of Analog CMOS Integrated Circuits*, McGraw-Hill, 2001.

[45] J. Montanaro *et al.*, "A 160-MHz, 32-b, 0.5-W CMOS RISC microprocessor," *IEEE J. Solid-State Circuits*, vol. 31, no. 11, pp. 1703-1714, Nov. 1996.

[46] Y.-T. Wang and B. Razavi, "An 8-Bit 150-MHz CMOS A/D converter," *IEEE J. Solid-State Circuits*, vol. 35, no. 3, pp. 308-317, March 2000.

[47] Y. Lu, E. Alon, "Design techniques for a 66 Gb/s 46 mW 3-tap decision feedback equalizer in 65 nm CMOS," *IEEE J. Solid-State Circuits*, vol. 48, no. 12, pp. 3243-3257, Dec. 2013.

[48] R. Kreienkamp *et al.*, "A 10-Gb/s CMOS clock and data recovery circuit with an analog phase interpolator," *IEEE J. Solid-State Circuits*, vol. 40, no. 3, pp. 736-743, March 2005.

[49] L. Li and M. Green, "Power optimization of an 11.75-Gb/s combined decision feedback equalizer and clock data recovery circuit in 0.18-µm CMOS," *IEEE Trans. Circuits and Systems - Part I*, vol. 58, no. 3, pp. 441-450, March 2011.

[50] B. Razavi, Y. Ota and R. G. Swartz, "Design Techniques for Low-Voltage High-Speed Digital Bipolar Circuits," *IEEE J. Solid-State Circuits*, vol. 29, no. 3, pp. 332-339, March 1994.

[51] J. W. Jung, "A 25-Gb/s 5-mW CDR/deserializer in 65-nm technology," Ph.D. dissertation, University of California, Los Angeles, 2012.