

UCLA

UCLA Electronic Theses and Dissertations

Title

Low-Power Wireline Transmitter Design

Permalink

<https://escholarship.org/uc/item/71c3v8kh>

Author

Chang, Yikun

Publication Date

2018

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
Los Angeles

Low-Power Wireline Transmitter Design

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Electrical Engineering

by

Yikun Chang

2018

© Copyright by
Yikun Chang
2018

ABSTRACT OF THE DISSERTATION

Low-Power Wireline Transmitter Design

by

Yikun Chang

Doctor of Philosophy in Electrical Engineering

University of California, Los Angeles, 2018

Professor Behzad Razavi, Chair

With the recent surge in the demand for high data rates, communication over copper media faces new challenges. First, the limited bandwidth removes so much of the signal's high-frequency energy that equalization and detection become very difficult. Second, the greater data rates in serial links inevitably translate to high power consumption. State-of-the-art transmitters operating in the range of tens of gigabits per second draw hundreds of milliwatts, underscoring the need for new circuit and architecture techniques that can ease the trade-off with speed.

The first part of this research introduces a 40-Gb/s non-return-to-zero transmitter that improves the power efficiency by a factor of 2.28. This is accomplished through removing power-hungry retimers in transmitter front end, merging the output driver and the final multiplexer stage, proposing a current-integrating multiplexer and "latchless" feedforward equalization path. Implemented in 45-nm CMOS technology, the transmitter provides 7.4-dB boosting and draws 32 mW at 40 Gb/s.

The second part of this research studies the design of an 80-Gb/s PAM4 transmitter that achieves nearly six-fold improvement in power efficiency with respect to state of the art. With a two-fold reduction in bandwidth occupancy

compared to non-return-to-zero data, the PAM4 format allows significant speed improvement but also introduces other issues such as skew and linearity. The design introduces a number of novel ideas so as to achieve both a very high data rate and much lower power consumption compared to state of the art. In particular, the design proposes a “latchless” serializer architecture, a charge-steering multiplexer, and a high-speed divide-by-two circuit that directly generates outputs with a 25% duty cycle. These techniques culminate in the 80-Gb/s PAM4 transmitter, including an on-chip phase-locked loop, that draws only 44 mW in 45-nm CMOS technology.

The dissertation of Yikun Chang is approved.

Chee Wei Wong

Gregory J. Pottie

Danijela Cabric

Behzad Razavi, Committee Chair

University of California, Los Angeles

2018

To my parents ...

TABLE OF CONTENTS

1	Data Formats	1
1.1	Time Domain	1
1.2	Spectrum	3
1.3	Nyquist Frequency	4
1.4	Effect of Noise on Bit Error Rate	5
1.4.1	BER of NRZ Data	5
1.4.2	BER of PAM4 Data	7
1.5	Jitter Due To Additive Noise	10
1.6	Signal-to-Noise Ratio	12
1.7	Ratio of Level Mismatch	14
2	Issues In Wireline Transmitter	16
2.1	Termination	16
2.2	Output Driver	18
2.2.1	CML Driver	19
2.2.2	SST Driver	20
2.3	Output Swing	23
2.4	Skew Between MSB And LSB Paths	24
2.5	Timing Of Serializer	26
2.6	Jitter On Output Data	27
2.6.1	Duty Cycle = 50%	28

2.6.2	Duty Cycle = 25%	30
3	Equalization In Transmitter	33
3.1	Pre-Emphasis	33
3.2	Inductive Peaking	37
3.2.1	Inductive Shunt Peaking	38
3.2.2	Inductive Series Peaking	40
3.2.3	T-Coil Peaking	42
4	A 40-Gb/s NRZ Transmitter	46
4.1	Design Considerations	46
4.2	Transmitter Architecture	48
4.3	Integrating MUX	49
4.4	Main and FFE Multiplexers/Drivers	51
4.5	Experimental Results	53
5	An 80-Gb/s PAM4 Transmitter	57
5.1	Background	57
5.2	Transmitter Architecture	58
5.3	Serializer Design	60
5.3.1	CMOS MUX	60
5.3.2	Charge-Steering MUX	63
5.3.3	Direct 4-to-1 MUX	69
5.4	Output Driver/DAC	73

5.5	Clock Generation	78
5.6	PLL Design	83
5.7	Floor Plan	86
5.8	Experimental Results	89
6	Conclusion	95
	References	97

LIST OF FIGURES

1.1	Transient waveforms of (a) NRZ signaling, and (b) PAM4 signaling.	2
1.2	Comparison of NRZ and PAM4 spectra.	3
1.3	PDF of (a) noiseless NRZ signal, and (b) noisy NRZ signal.	6
1.4	PDF of noiseless PAM4 signal.	7
1.5	PDF of noisy PAM4 signal under (a) binary code, and (b) Grey code.	7
1.6	BER of NRZ and PAM4 signalings.	10
1.7	Effect of additive noise on jitter.	11
1.8	Eye diagrams of (a) PAM4 data, and (b) NRZ data under the same swing and at the same data rate. (1 UI is referred to PAM4 data.)	13
1.9	Eye diagram with nonlinear PAM4 levels.	14
1.10	A PAM4 linearity test sequence.	15
2.1	Signal generation, propagation and reflection on T line with load.	16
2.2	Double termination.	18
2.3	CML driver.	19
2.4	SST driver.	21
2.5	Equivalent circuit of PAM4 SST driver when (a) MSB and LSB are equal, and (b) MSB and LSB are opposite.	21
2.6	Comparison of PAM4 CML and SST drivers' outputs on (a) vertical eye opening, and (b) horizontal eye opening.	22
2.7	Noise model of transceiver front end and channel.	23

2.8	PAM4 output eye diagrams with (a) LSB lags MSB by 0.2 UI, and (b) LSB leads MSB by 0.2 UI.	25
2.9	Conventional structure and timing of serializer.	26
2.10	Conventional configuration with retimer.	28
2.11	A 2-to-1 CML MUX.	28
2.12	DCD of 50%-duty-cycle clocks.	29
2.13	Skew of 50%-duty-cycle clocks.	30
2.14	DCD of 25%-duty-cycle clocks.	31
2.15	Skew of 25%-duty-cycle clocks.	32
3.1	Transmitted and received pulses.	34
3.2	(a) Transmitted waveform, and (b) received waveform with equal- ization.	35
3.3	Reason for pre-cursors.	35
3.4	A two-tap pre-emphasis.	36
3.5	Magnitude of de-emphasis transfer functions.	37
3.6	Evolution of inductive peaking: (a) only resistor as load, (b) add a capacitor, (c) add a switch in series with resistor, and (d) replace switch with an inductor.	39
3.7	Configuration of shunt peaking.	40
3.8	Shunt peaking: (a) frequency responses, and (b) transient waveforms.	41
3.9	(a) Configuration of series peaking, and (b) equivalent circuit.	41
3.10	Configuration of two cases: (a) $C_1 < C_2$, and (b) $C_1 > C_2$	42
3.11	Series peaking: (a) frequency responses, and (b) transient waveforms.	43

3.12	Configuration of T-coil peaking.	43
3.13	Tcoil peaking: (a) frequency responses, and (b) transient waveforms.	44
3.14	(a) Normalized magnitude and (b) phase of input impedance with T-coil peaking.	45
3.15	T-coil peaking (a) for input pad, and (b) for output pad.	45
4.1	(a) Full-rate front end with retimer and divider, (b) half-rate front end, (c) half-rate front end with combined 2-to-1 MUX and driver, and (d) quarter-rate front end with combined 4-to-1 MUX and driver.	47
4.2	Proposed NRZ transmitter architecture.	48
4.3	Driving the 4-to-1 MUX by (a) a CML stage, or (b) an integrating stage.	50
4.4	Proposed 2-to-1 integrating MUX.	50
4.5	Main and FFE data paths.	52
4.6	Interface between the integrating MUX and the main and FFE drivers/MUXes.	53
4.7	TX die photograph.	53
4.8	(a) Measured spectrum of 20-GHz clock, and (b) phase noise profile of 10-GHz clock.	54
4.9	Measured eye diagrams with (a) FFE off, and (b) four FFE slices on.	55
4.10	Measured spectrum of single-ended output delivering 0101 sequence.	55
5.1	Proposed transmitter architecture.	59

5.2	(a) Conventional three-latch MUX cell, and (b) simplified MUX cell.	61
5.3	Proposed timing scheme to remove latches by applying I and Q clocks.	62
5.4	(a) CMOS selector used in this work, and (b) simulated output eye diagram of the last stage of CMOS MUX.	63
5.5	Simple charge-steering 2-to-1 MUX.	64
5.6	(a) Proposed charge-steering MUX, and (b) role of PMOS pull-up device in suppressing the effect of kickback noise.	66
5.7	Simulated output waveforms of charge-steering MUX with and without PMOS differential pairs.	67
5.8	Timing diagram of charge-steering MUX with 25% duty-cycle clocks.	67
5.9	(a) Timing diagram of charge-steering MUX with 25% duty-cycle clocks, (b) four charge-steering MUXes with idealized waveforms, and (c) rotation of SEL_1 - SEL_4 in four charge-steering MUXes to accommodate the clock delay.	68
5.10	(a) Binary-tree 4-to-1 MUX, and (b) direct 4-to-1 CML MUX. . .	70
5.11	Dependence of height and width of PAM4 middle eye upon duty cycle.	72
5.12	(a) Output eye for 20% duty cycle, and (c) output eye for 37.5% duty cycle.	72
5.13	Direct 4-to-1 MUX single-ended output eye-diagrams (a) without and (b) with PMOS differential pairs, and transmitter differential output eye-diagrams (c) without and (d) with PMOS differential pairs.	73
5.14	Topology of the PAM4 CML output driver/DAC.	74

5.15	Equivalent circuit of single-ended current-steering DAC.	75
5.16	(a) Output of 7-bit single-ended current steering DAC, and (b) INL.	76
5.17	Equivalent circuit of differential current-steering DAC.	76
5.18	(a) Output of 7-bit differential current steering DAC, and (b) INL.	77
5.19	Comparison of INL between single-ended and differential PAM4 CML output driver.	77
5.20	(a) Divider topology to generate 25%-duty-cycle clocks directly [26], and (b) divider's waveforms.	79
5.21	(a) Latch topology to remove static current of M_a and M_b , and (b) M_c and M_d driven by CK to reduce transition delay of falling edge on V_{X1} and V_{Y1}	80
5.22	(a) Proposed latch topology with stacked NMOS devices, (b) sim- ulated waveforms of the divider outputs, and (c) the outputs after three inverters.	81
5.23	Relation of duty cycle upon transistor ratio.	82
5.24	(a) Divide-by-2 stage to generate eight-phase clocks, and (b) C ² MOS latch used in the divider.	83
5.25	(a) PLL with master-slave sampling filter, and (b) settling behavior of VCO control voltage.	85
5.26	(a) VCO implementation, and (b) simulated frequency tuning. . .	86
5.27	(a) Modular-based scaling between the first and the second ranks, and (b) modular placement in CMOS MUX array.	87
5.28	Placement of inductors in layout floor plan.	88
5.29	Die photograph.	89

5.30	Output eye diagram in NRZ mode at 40 Gb/s.	90
5.31	PAM4 output eye diagrams at (a) 40 Gb/s, and (b) 80 Gb/s.	91
5.32	Comparison between simulated and measured waveform.	91
5.33	RLM test sequence.	92
5.34	(a) Spectrum of 20-GHz clock, phase noise profile of 20-GHz clock divided by two externally, and (c) relation of jitter and integrating range of 20-GHz clock divided by two externally.. . . .	93
5.35	Measured spectrum of single-ended output delivering 20-GHz 0101 NRZ sequence.	94

LIST OF TABLES

4.1	PERFORMANCE SUMMARY.	56
5.1	POWER BREAKDOWN.	90
5.2	PERFORMANCE SUMMARY.	94

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to Professor Behzad Razavi, who sets a good example to academic researchers. His intelligence, knowledgeability and diligence inspired me so much. As an experienced advisor, he not only guided my research on the right path, but also shared his good working habits with me. Every time when I ran too far away from the topic, he stopped me in time and asked me to focus on the most critical and fundamental problems. He taught me to manage time, handle multi-tasks, write papers, polish presentation skills and develop proper plans, which are valuable not only for my Ph.D study but also for my future life.

I would like to thank Professors Danijela Cabric, Gregory J. Pottie, and Chee Wei Wong, for their time to serve on my committee.

I would like to thank my lab mates that I had overlap with. In particular, I would like to express my gratitude to the two senior fellows, Long Kong, who taught me tapeout procedure and gave me many suggestions on planning Ph.D study, and Abishek Manian, who taught me linux, HFSS, layout and testing techniques. I would also like to thank Atharav for his useful discussions and suggestions.

I would like to thank my Alma Mater, Peking University. She opened a door for me to a colorful world full of brilliant and versatile people. She redefined me and also let me aware of the definition of a meaningful life is not unique. The experience there motivated me to go abroad and explore a even bigger world.

I would like to thank my friends, Fengqi Zhang, Bilin Wang, Xiaoning Wang and Jieqiong Du. Especially, I would like to say thank you to Fenqi Zhang, who

comforted me during the hard and lonely time when I just arrived in America.

I would like to thank my boyfriend, Yifan Ding. I was fortunate to meet this kindhearted and considerate gentleman at UCLA. It was him that trusted and encouraged, endured with and accompanied me all along the journey. Without him, I would not get through the toughest time.

Lastly, but most importantly, I would like to thank my family for their support. I am extremely grateful to my parents, who provided me with the best chances to broaden my horizon and interests even in a small city, supported me to study aboard far away from them although I am their only child, and guided me but never pushed me. I also extremely appreciate my grandparents, for their utmost care during my growing up, and for their upright and tenacious personality that inspired me.

VITA

- 1990 Born, Luoyang, Henan, China.
- 2013 B.S. (Microelectronics), Peking University, Beijing, China.
- 2014 Intern, TSMC, San Jose, USA.
- 2015 M.S. (Electrical Engineering), University of California, Los Angeles, USA.
- 2015 Intern, Broadcom, Irvine, USA.
- 2015 Ph.D. Candidate (Electrical Engineering), University of California, Los Angeles, USA.

PUBLICATIONS

Y. Chang, A. Manian, L. Kong, and B. Razavi, “A 32-mW 40-Gb/s CMOS NRZ Transmitter,” accepted by *IEEE CICC*, Apr. 2018.

Y. Chang, A. Manian, L. Kong, and B. Razavi, “An 80-Gb/s 44-mW Wireline PAM4 Transmitter,” accepted by *IEEE J. Solid-State Circuits*, 2018.

CHAPTER 1

Data Formats

For a long time, non-return-to-zero (NRZ) signaling has dominated wireline communications because it is simple to equalize and detect. However, as the demand for data throughput keeps increasing, even the equalization and detection of NRZ signal become very difficult because the limited channel bandwidth attenuates high-frequency components so much. It is in this spirit that, after an initial appearance in the 2000s [1, 2, 3], PAM4 signaling has been resurrected.

1.1 Time Domain

Figure 1.1 depicts the transient waveforms of NRZ and PAM4 signals. To transmit a random binary sequence consisting of logical ONEs and ZEROs, NRZ signaling exploits high and low levels to represent logical ONE and ZERO. Thus, a single NRZ symbol carries one-bit information. In PAM4 signaling, a symbol contains four levels and exhibits the capability to carry two-bit information. Therefore, every two binary bits are grouped and transmitted together in PAM4 signaling. The four possible groups, 00, 01, 10 and 11, map to the four levels, respectively.

Both NRZ and PAM4 signals belong to digital-amplitude-modulated signals.

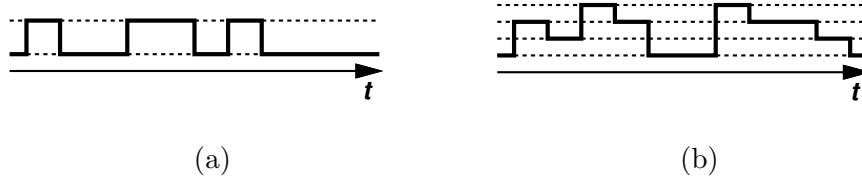


Figure 1.1. Transient waveforms of (a) NRZ signaling, and (b) PAM4 signaling.

Such signal, $x(t)$, can be generally written as

$$x(t) = \sum_{k=0}^{\infty} b_k p(t - kT_b), \quad (1.1)$$

where T_b is the duration of a single symbol, which, in wireline communications, is also called a unit interval (UI); $p(t)$ denotes the pulse function:

$$p(t) = \begin{cases} 1, & \text{for } 0 \leq t \leq T_b, \\ 0, & \text{otherwise.} \end{cases} \quad (1.2)$$

NRZ signal is a pulse train with its amplitude modulated by two levels. If we use $\pm V_0$ to represent logical ONE and ZERO, b_k is expressed as

$$b_k = \begin{cases} +V_0, & \text{for symbol 1,} \\ -V_0, & \text{for symbol 0.} \end{cases} \quad (1.3)$$

In PAM4 signaling, the pulse amplitude is modulated by four levels. To optimize the signal-to-noise ratio (SNR) (explained in Section 1.6), four levels separate uniformly. If the maximum amplitude is also V_0 , b_k is

$$b_k = \begin{cases} +V_0, & \text{for symbol 11,} \\ +\frac{1}{3}V_0, & \text{for symbol 10,} \\ -\frac{1}{3}V_0, & \text{for symbol 01,} \\ -V_0, & \text{for symbol 00,} \end{cases} \quad (1.4)$$

if it is binary coded.

1.2 Spectrum

For an amplitude-modulated signal $x(t)$, it can be proved that its power spectrum density (PSD) is in the form of [5]

$$S_x(f) = \frac{\overline{|b_k P(f)|^2}}{T_b}, \quad (1.5)$$

where $P(f)$ is Fourier transform of $p(t)$. If $x(t)$ data rate is r bit/second and symbol rate is R_b baud, then $T_{b,NRZ} = 1/R_{b,NRZ} = 1/r$ and $T_{b,PAM4} = 1/R_{b,PAM4} = 2/r$. For NRZ and PAM4 signals expressed by Eq.(1.1) - (1.4), the spectra are

$$S_{x,NRZ}(f) = V_0^2 T_b \left[\frac{\sin(\pi f T_b)}{\pi f T_b} \right]^2 = \frac{V_0^2}{r} \left[\frac{\sin(\frac{\pi f}{r})}{\frac{\pi f}{r}} \right]^2, \quad (1.6)$$

and

$$S_{x,PAM4}(f) = \frac{5}{9} V_0^2 T_b \left[\frac{\sin(\pi f T_b)}{\pi f T_b} \right]^2 = \frac{10V_0^2}{9r} \left[\frac{\sin(\frac{2\pi f}{r})}{\frac{2\pi f}{r}} \right]^2. \quad (1.7)$$

Figure 1.2 plots the two spectra.

NRZ spectrum shows a main lobe between frequencies $\pm r$ while PAM4 spectrum main lobe between $\pm 0.5r$. In a $\text{sinc}(\cdot) = \frac{\sin(\cdot)}{(\cdot)}$ function, the main lobe is about 13-dB higher than the first side lobe and contains 90% power of the

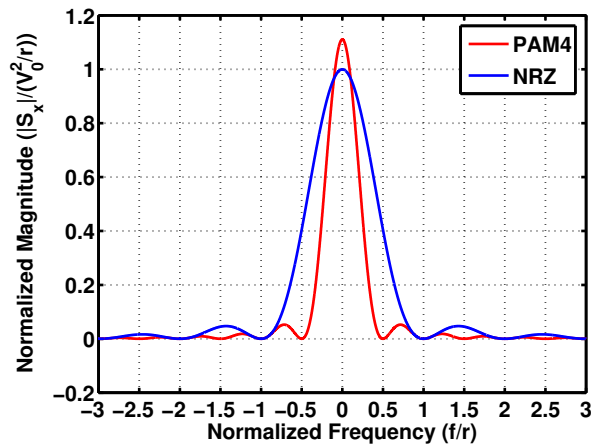


Figure 1.2. Comparison of NRZ and PAM4 spectra.

whole spectrum. For this reason, at the same data rate, we say that PAM4 data has more power concentrated to DC than its NRZ counterpart. Intuitively, at the same data rate, PAM4 signal updates its symbol twice slowly on average compared to NRZ signal.

1.3 Nyquist Frequency

For a transmitter designed for a symbol rate R_b , if we configure the input data such that the output waveform toggles at its highest rate between the highest and the lowest levels, then the output spectrum will show a fundamental tone at frequency $R_b/2$. This frequency is called Nyquist frequency, representing the highest fundamental frequency at the symbol rate R_b .

The name of Nyquist frequency may make us recall the sampling theorem, which says the lowest sampling frequency, f_s , must satisfy that $f_s \geq 2f_M$ to recover the original signal, where f_M is the highest frequency of the original signal and f_s is called Nyquist rate. If we treat wireline transmitted data as a sampled version in which $f_s = R_b$, then the highest frequency of the original signal should satisfy $f_M \leq f_s/2 = R_b/2$, which gives us the same conclusion as above.

In wireline communications, we care about how much information is kept after data traveling through a lossy channel. This concern is quantified by the channel loss at Nyquist frequency. For the same data rate, r , Nyquist frequency of PAM4 signal is at $R_{b,PAM4}/2 = r/4$ while that of NRZ signal at $R_{b,NRZ}/2 = r/2$. The lower Nyquist frequency of PAM4 signal yields a smaller channel loss. For this reason, PAM4 signaling shows its potential in high data-rate wireline communications.

1.4 Effect of Noise on Bit Error Rate

Random data propagating through a lossy channel may experience considerable attenuation. On receiver side, data symbols, ideally, must be sampled by the clock at their midpoints and compared with threshold(s) to detect symbol levels. The noise, $n(t)$, added to the signal degrades eye opening and even causes some logic levels to cross threshold(s) erroneously, leading to an increase in bit-error-rate (BER) of detection.

To derive BER in terms of the amplitude of the additive noise, we assume that the noise amplitude exhibits a Gaussian distribution with zero mean and write the probability density function (PDF) of $n(t)$ as

$$P_n = \frac{1}{\sigma_n \sqrt{2\pi}} \exp \frac{-n^2}{2\sigma_n^2}, \quad (1.8)$$

where σ_n denotes the root mean square (rms) value of the noise.

1.4.1 BER of NRZ Data

In NRZ signal $x(t)$, if ONES and ZEROS occur with equal probabilities, then the PDF of the noiseless signal consists of two pulses at $x = -V_0$ and $x = +V_0$, each having a weight of 1/2 [Fig. 1.3(a)]. With the additive noise $n(t)$, the PDF of the noiseless signal convolves with the PDF of $n(t)$, resulting in a PDF of two Gaussian distributions centered at $\pm V_0$, respectively, with an rms value of σ_n [Fig. 1.3(b)].

The error probability is given by the probability summary of false alarms on logical ONE and ZERO, i.e., the probability when $-V_0 + n(t)$ falls in the region beyond the threshold, 0, and $+V_0 + n(t)$ below 0. The probability of positive

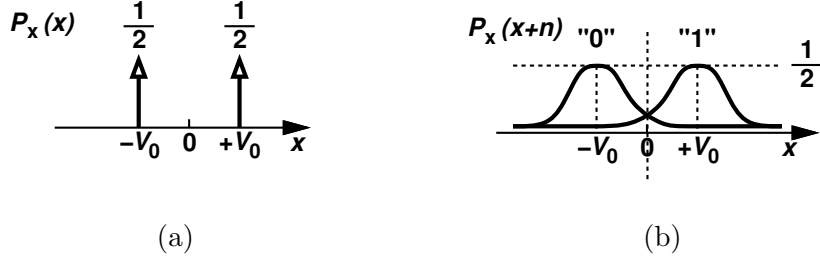


Figure 1.3. PDF of (a) noiseless NRZ signal, and (b) noisy NRZ signal.

$-V_0 + n(t)$ is given by

$$P_{0 \rightarrow 1} = \frac{1}{2} \int_0^{+\infty} \frac{1}{\sigma_n \sqrt{2\pi}} \exp \frac{-(\mu + V_0)^2}{2\sigma_n^2} d\mu. \quad (1.9)$$

Similarly, the probability of negative $+V_0 + n(t)$ is represented by

$$P_{1 \rightarrow 0} = \frac{1}{2} \int_{-\infty}^0 \frac{1}{\sigma_n \sqrt{2\pi}} \exp \frac{-(\mu - V_0)^2}{2\sigma_n^2} d\mu. \quad (1.10)$$

We simplify $P_{0 \rightarrow 1}$ to be

$$\begin{aligned} P_{0 \rightarrow 1} &= \frac{1}{2} \int_{V_0/\sigma_n}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp \frac{-z^2}{2} dz \\ &= \frac{1}{2} Q \left(\frac{V_0}{\sigma_n} \right), \end{aligned} \quad (1.11)$$

where $Q(\cdot)$ is the Q function defined as

$$Q(x) = \int_x^{+\infty} \frac{1}{\sqrt{2\pi}} \exp \frac{-\mu^2}{2} d\mu. \quad (1.12)$$

The error probability is therefore equal to

$$\begin{aligned} P_{tot} &= P_{0 \rightarrow 1} + P_{1 \rightarrow 0} \\ &= Q \left(\frac{V_0}{\sigma_n} \right) \end{aligned} \quad (1.13)$$

For a stationary process, we can use its statistic property, the error probability, to represent the error rate in practice. Thus, BER of NRZ data is equal to

$$\begin{aligned} BER_{NRZ} &= Q \left(\frac{V_0}{\sigma_n} \right) \\ &= Q \left(\frac{V_{pp}}{2\sigma_n} \right), \end{aligned} \quad (1.14)$$

where $V_{pp} = 2V_0$ is the peak-to-peak swing. Notice that the *symbol* error rate of NRZ data is equal to its BER.

1.4.2 BER of PAM4 Data

Similarly to NRZ signal, Fig. 1.4 depicts the PDF of noiseless PAM4 signals under the assumption that four levels occur with equal probabilities. The thresholds are set at $-\frac{2}{3}V_0$, 0 , and $+\frac{2}{3}V_0$. In PAM4 data detection, the code format also affects BER because the false detections between 00 and 01, and 11 and 10 result in one-bit error while those between 00 and 11, and 01 and 10 lead to two-bit error. Figure 1.5 depicts the PDF of noisy PAM4 signals under two coding methods, binary code and Grey code, respectively.

In binary code, the *bit* error probability when symbol 00 is detected erroneously is given by

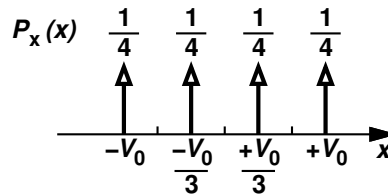


Figure 1.4. PDF of noiseless PAM4 signal.

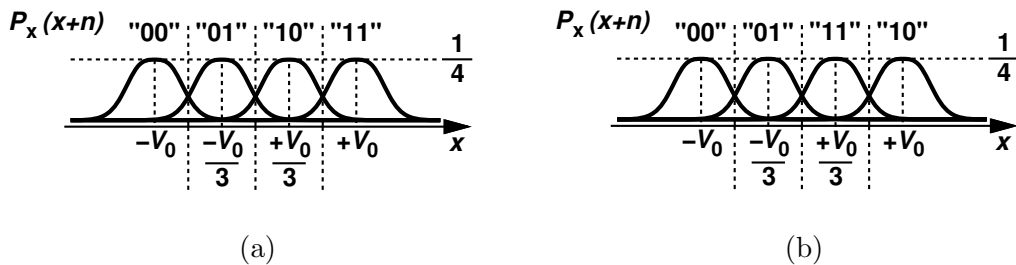


Figure 1.5. PDF of noisy PAM4 signal under (a) binary code, and (b) Grey code.

$$\begin{aligned}
P_{b,00 \rightarrow others} &= P_{00 \rightarrow 01} + P_{00 \rightarrow 10} + 2P_{00 \rightarrow 11} \\
&= P_{00 \rightarrow others} + P_{00 \rightarrow 11} \\
&= \frac{1}{4}Q\left(\frac{V_0}{3\sigma_n}\right) + \frac{1}{4}Q\left(\frac{5V_0}{3\sigma_n}\right).
\end{aligned} \tag{1.15}$$

And when symbol 01 is detected erroneously, the bit error probability is

$$\begin{aligned}
P_{b,01 \rightarrow others} &= P_{01 \rightarrow 00} + 2P_{01 \rightarrow 10} + P_{01 \rightarrow 11} \\
&= P_{00 \rightarrow others} + P_{01 \rightarrow 10} \\
&= 2 \times \frac{1}{4}Q\left(\frac{V_0}{3\sigma_n}\right) + \frac{1}{4}Q\left(\frac{V_0}{3\sigma_n}\right) - \frac{1}{4}Q\left(\frac{V_0}{\sigma_n}\right) \\
&= \frac{1}{2}Q\left(\frac{V_0}{3\sigma_n}\right) + \frac{1}{4}Q\left(\frac{V_0}{3\sigma_n}\right) - \frac{1}{4}Q\left(\frac{V_0}{\sigma_n}\right).
\end{aligned} \tag{1.16}$$

According to the symmetry of the PDF, we have $P_{b,00 \rightarrow others} = P_{b,11 \rightarrow others}$ and $P_{b,01 \rightarrow others} = P_{b,10 \rightarrow others}$. Therefore, the total bit error probability is

$$\begin{aligned}
P_{b,tot} &= P_{b,00 \rightarrow others} + P_{b,01 \rightarrow others} + P_{b,10 \rightarrow others} + P_{b,11 \rightarrow others} \\
&= \frac{3}{2}Q\left(\frac{V_0}{3\sigma_n}\right) + \frac{1}{2}\left[Q\left(\frac{5V_0}{3\sigma_n}\right) + Q\left(\frac{V_0}{3\sigma_n}\right) - Q\left(\frac{V_0}{\sigma_n}\right)\right] \\
&= 2Q\left(\frac{V_0}{3\sigma_n}\right) - \frac{1}{2}\left[Q\left(\frac{V_0}{\sigma_n}\right) - Q\left(\frac{5V_0}{3\sigma_n}\right)\right].
\end{aligned} \tag{1.17}$$

Applying the same analysis on Grey code, we have

$$\begin{aligned}
P_{b,00 \rightarrow others} &= P_{00 \rightarrow 01} + P_{00 \rightarrow 10} + 2P_{00 \rightarrow 11} \\
&= P_{00 \rightarrow others} + P_{00 \rightarrow 11} \\
&= \frac{1}{4}Q\left(\frac{V_0}{3\sigma_n}\right) + \frac{1}{4}Q\left(\frac{V_0}{\sigma_n}\right) - \frac{1}{4}Q\left(\frac{5V_0}{3\sigma_n}\right),
\end{aligned} \tag{1.18}$$

and

$$\begin{aligned}
P_{b,01 \rightarrow \text{others}} &= P_{01 \rightarrow 00} + 2P_{01 \rightarrow 10} + P_{01 \rightarrow 11} \\
&= P_{00 \rightarrow \text{others}} + P_{01 \rightarrow 10} \\
&= 2 \times \frac{1}{4}Q\left(\frac{V_0}{3\sigma_n}\right) + \frac{1}{4}Q\left(\frac{V_0}{3\sigma_n}\right) - \frac{1}{4}Q\left(\frac{V_0}{\sigma_n}\right) \\
&= \frac{1}{2}Q\left(\frac{V_0}{3\sigma_n}\right) + \frac{1}{4}Q\left(\frac{V_0}{\sigma_n}\right).
\end{aligned} \tag{1.19}$$

Thus, we can write the total bit error probability as

$$\begin{aligned}
P_{b,tot} &= P_{b,00 \rightarrow \text{others}} + P_{b,01 \rightarrow \text{others}} + P_{b,10 \rightarrow \text{others}} + P_{b,11 \rightarrow \text{others}} \\
&= \frac{3}{2}Q\left(\frac{V_0}{3\sigma_n}\right) + \frac{1}{2}\left[2Q\left(\frac{V_0}{\sigma_n}\right) - Q\left(\frac{5V_0}{3\sigma_n}\right)\right] \\
&= 2Q\left(\frac{V_0}{3\sigma_n}\right) + \frac{1}{2}\left[2Q\left(\frac{V_0}{\sigma_n}\right) - Q\left(\frac{V_0}{3\sigma_n}\right) - Q\left(\frac{5V_0}{3\sigma_n}\right)\right].
\end{aligned} \tag{1.20}$$

If we ignore the small second term in Eq. (1.17) and (1.20), both of them lead to

$$\begin{aligned}
BER_{PAM4} = P_{b,tot} &\approx 2Q\left(\frac{V_0}{3\sigma_n}\right) \\
&= 2Q\left(\frac{V_{pp}}{6\sigma_n}\right).
\end{aligned} \tag{1.21}$$

Since $Q(\cdot)$ is a monotone decreasing function, Eq. (1.14) and (1.21) indicate that PAM4 signaling exhibits a higher BER than NRZ signaling with the same swing and the same additive noise. The variable inside the $Q(\cdot)$ function shows three-time difference between PAM4 and NRZ signalings due to the fact that PAM4 neighbor levels show only 1/3 separation of NRZ levels. The difference of the coefficient outside $Q(\cdot)$ function comes from that PAM4 data detection shows more cases of wrong detection than NRZ data detection.

Figure 1.6 depicts BER of the two data formats. As we can see from Eq. (1.14) and (1.21), for a certain BER, PAM4 signaling would require a swing more than

three times of that of NRZ signaling. However, the maximum swing is always limited by hardware headrooms. Therefore, standard [4] lowers the required BER of PAM4 signaling, for example, to be about 10^{-6} , and uses forward-error-correction (FEC) coding to bring BER back to around 10^{-15} .

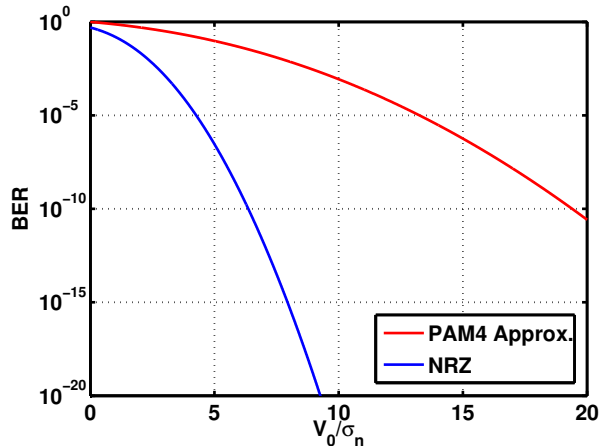


Figure 1.6. BER of NRZ and PAM4 signalings.

1.5 Jitter Due To Additive Noise

If a sequence $D(t)$ is corrupted by additive noise $n(t)$ as shown in Fig. 1.7, the threshold crossing of $D(t) + n(t)$ in the vicinity of $t = t_0$ deviates from the ideal value t_0 by

$$\Delta T_0 = \frac{n(t_0)}{S(t_0)}, \quad (1.22)$$

where $S(t_0)$ denotes the slope of the transition around $t = t_0$. Thus, the sharper the edge, the less deviation.

Let us assume the waveform $D(t)$ is an ideal sequence with fast transition going through a first-order low-pass transfer function with $\omega_{-3dB} = 1/\tau$, and

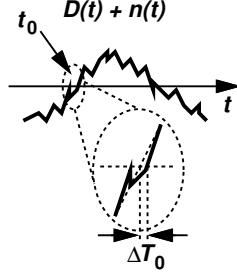


Figure 1.7. Effect of additive noise on jitter.

thus exhibits an exponential transition in the form of

$$D(t) = V_i + (V_f - V_i)[1 - \exp(-t/\tau)], \quad (1.23)$$

where V_i and V_f denote the initial and the final values of the transition, respectively. And the slope of the transition is given by its derivative

$$S(t) = \frac{V_f - V_i}{\tau} \exp(-t/\tau). \quad (1.24)$$

At $t = t_0$, $D(t_0) = (V_f + V_i)/2$ gives $\exp(-t/\tau) = 1/2$. Therefore, we have

$$S(t_0) = \frac{V_f - V_i}{2\tau}, \quad (1.25)$$

and

$$\Delta T_0 = \frac{2\tau n(t_0)}{V_f - V_i}. \quad (1.26)$$

Suppose the bandwidth ω_{-3dB} is η times of the symbol rate, that is $\omega_{-3dB} = 1/\tau = \eta 2\pi R_b = \eta 2\pi/T_b$, we re-write the normalized jitter to be

$$\frac{\Delta T_0}{T_b} = \frac{n(t_0)}{\pi\eta(V_f - V_i)}. \quad (1.27)$$

In a linear system, the transition slope is proportional to the difference between the final and the initial values. If we assume the swings of NRZ and PAM4 signals are both $2V_0$, the transitions in NRZ signal only occur between $\pm V_0$ while those in PAM4 signal between four levels and result in three different slopes.

Except for the transition between $\pm V_0$, the other transition cases in PAM4 signal exhibit smaller slopes and lead to larger jitter than NRZ data. The worst case of the transitions in PAM4 signal gives $V_f - V_i = 2V_0/3$. According to Eq. (1.14) and (1.21), $\text{BER}_{NRZ} = 10^{-12}$ leads to $V_0/n_{rms}(t_0) = 7$ and $\text{BER}_{PAM4} = 10^{-7}$ to $V_0/n_{rms}(t_0) = 16$. With $\eta = 0.7$, we arrive at

$$\left(\frac{\Delta T_{0,rms}}{T_b}\right)_{NRZ} \approx 3.25\%, \quad (1.28)$$

and

$$\left(\frac{\Delta T_{0,rms}}{T_b}\right)_{PAM4,worst} \approx 4.26\%. \quad (1.29)$$

This amount of jitter may not be acceptable in some applications.

1.6 Signal-to-Noise Ratio

As depicted in Fig. 1.8, for NRZ and PAM4 signals under the same swing, PAM4 eyes show 1/3 height of the NRZ eye when the four levels are uniformly apart, indicating that PAM4 signal is more vulnerable to noise, crosstalks and reflections than NRZ signal.

Using signal-to-noise ratio (SNR) to quantify the degradation, PAM4 signal would show a SNR 9.5 dB less than NRZ signal under the same condition of noise, crosstalk and reflections. Ideally, for the same data rate r , if the channel loss at NRZ Nyquist frequency, $r/2$, is L_1 dB, and at PAM4 Nyquist frequency, $r/4$, is L_2 dB, where both L_1 and L_2 are positive, then we would prefer PAM4 data format if $L_1 - L_2 > 9.5$ dB.

SNR of NRZ signal is given by

$$\text{SNR}_{NRZ} = \frac{V_0^2}{2\sigma_n^2}. \quad (1.30)$$

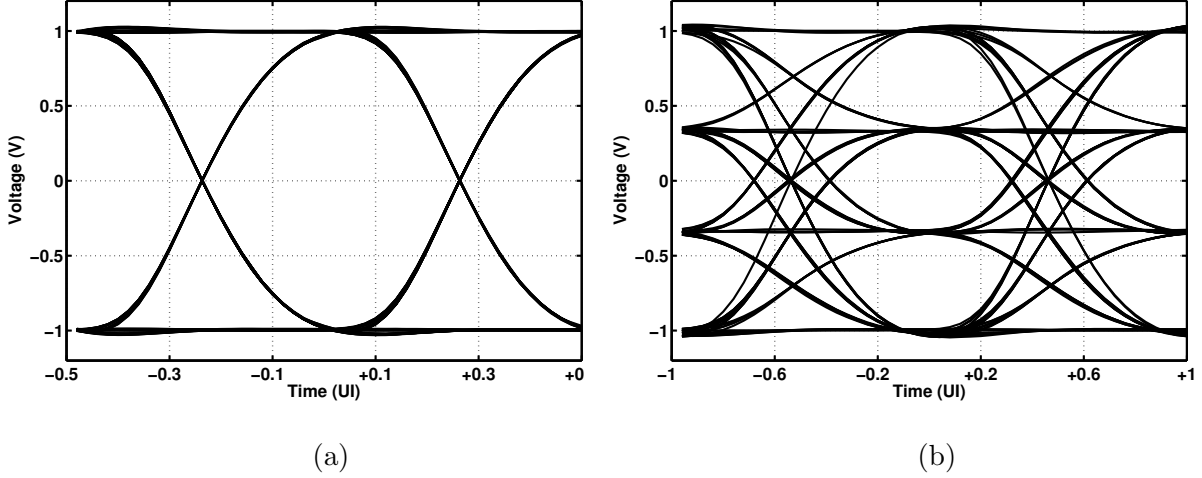


Figure 1.8. Eye diagrams of (a) PAM4 data, and (b) NRZ data under the same swing and at the same data rate. (1 UI is referred to PAM4 data.)

For PAM4 signal, we use the case where the signal toggles between two neighbor levels to represent its SNR in order to quantify the worst immunity to disturbs. Therefore, we get

$$SNR_{PAM4} = \frac{(V_0/3)^2}{2\sigma_n^2} = \frac{V_0^2}{18\sigma_n^2}. \quad (1.31)$$

As a result, BER of NRZ and PAM4 signals in Eq. (1.14) and (1.21) is simplified as a function of SNR:

$$BER_{NRZ} = Q(\sqrt{2SNR_{NRZ}}), \quad (1.32)$$

and

$$BER_{PAM4} \approx 2Q\left(\sqrt{2SNR_{PAM4}}\right). \quad (1.33)$$

If PAM4 levels do not separate uniformly as Fig. 1.9 shows, the minimum eye height would become the bottleneck and limits the immunity to noise, crosstalks and reflections. In this case, SNR would be that when the signal toggles between

the least separated two levels, which is

$$\begin{aligned} SNR_{PAM4} &= \frac{V_{min. \text{ amplitude}}^2}{2\sigma_n^2} \\ &< \frac{(V_0/3)^2}{2\sigma_n^2} = \frac{V_0^2}{18\sigma_n^2}. \end{aligned} \quad (1.34)$$

For this reason, uniformly distributed levels are preferred to maximize PAM4 SNR.

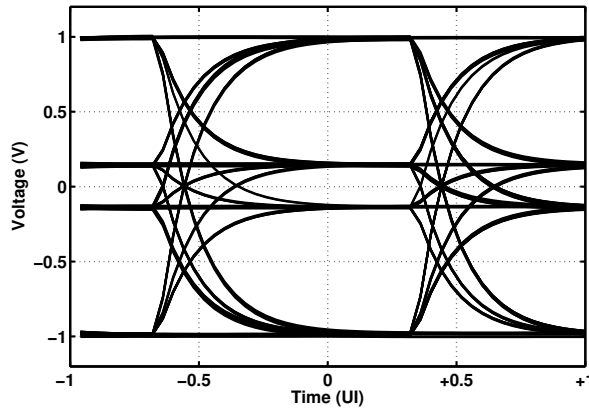


Figure 1.9. Eye diagram with nonlinear PAM4 levels.

1.7 Ratio of Level Mismatch

If nonlinearity happens in PAM4 signal, the three eyes in PAM4 eye diagram would show different heights. The standard [4] uses ratio of level mismatch (RLM) to quantify such distortion. Figure 1.10 depicts one version of the definition, using a test sequence with each symbol lasting for 16 UIs.

If the four levels from low to high are V_1 , V_2 , V_3 and V_4 , respectively, then

$$RLM = \frac{3 \times \min(V_2 - V_1, V_3 - V_2, V_4 - V_3)}{V_4 - V_1}, \quad (1.35)$$

that is, RLM is the ratio of the minimum eye height to 1/3 of the swing. The standard suggests this RLM to be larger than 0.92.

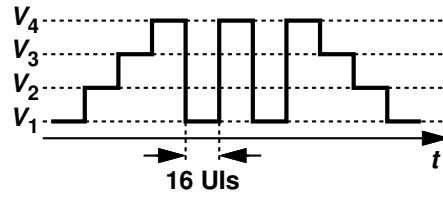


Figure 1.10. A PAM4 linearity test sequence.

The final definition of RLM is still under discussion.

CHAPTER 2

Issues In Wireline Transmitter

A wireline transmitter achieves two basic functions: (1) serializes low-speed paralleled inputs into a high-speed data stream, and (2) delivers the data stream in a certain format and of a certain swing to channel. Therefore, a typical wireline transmitter contains at least two parts, serializers and output drivers. To clock serializers, a wireline transmitter may also integrate functions such as clock generation and distribution.

2.1 Termination

In wireline communications, data travels through transmission lines (T lines) from a transmitter to a receiver. The propagation behavior yields an interesting property: if a step is applied to one end, the instantaneous behavior at that end depends only on a *fundamental* property of the line. Shown in Fig. 2.1, we have $V_i/I_i = Z_0$, where Z_0 represents the “characteristic impedance” of the line and is usually designed to be resistive.

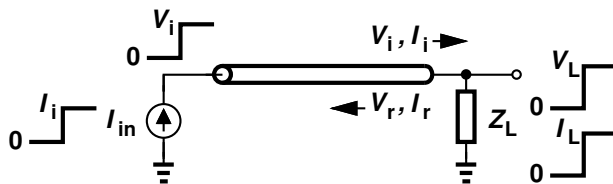


Figure 2.1. Signal generation, propagation and reflection on T line with load.

The generated voltage step V_i propagates on the line, establishing a current of I_i at each point as it travels. If the other end is terminated by an impedance Z_L equal to Z_0 , then the relationship $V_i/I_i = Z_0$ remains valid as the wave reaches the load and the transient ceases thereafter. However, if $Z_L \neq Z_0$, the voltage and current waveforms approaching the load violate Ohm's law, requiring a *reflection* to be generated.

If we call the reflected voltage and current to be V_r and I_r , in order to travel back through the line, they should also satisfy $V_r/I_r = Z_0$. Therefore, the ratio of the reflected waveform to the incident waveform is the same for voltage and current. Calling the ratio Γ , we have $V_r = \Gamma V_i$, $I_r = \Gamma I_i$ and the following relationships:

$$\begin{cases} V_L = V_i + V_r, \\ I_L = I_i - I_r, \\ \frac{V_i}{I_i} = Z_0, \\ \frac{V_L}{I_L} = Z_L. \end{cases} \quad (2.1)$$

It can be proved that

$$\Gamma = \frac{Z_L - Z_0}{Z_L + Z_0}. \quad (2.2)$$

Equation (2.2) also confirms that if $Z_L = Z_0$, no reflection is generated.

When the reflection generated at Z_L end travels back to the source end, it becomes an incident waveform and would result in a second reflection if the source end is not properly terminated. As a result, if without proper terminations, the waveforms are reflected again and again at both ends, and travel back and forth through the line. At the source end, the incident and the reflected waveforms are superposed on the transmitted waveform and may saturate the transmitter. At Z_L end, the detection of the useful incident waveform are corrupted by the reflected and other incident waveforms.

It is true that the termination of $Z_L = Z_0$ at the receiver is enough to stop reflections. However, the parasitic capacitance inevitably introduces impedance mismatch. At tens of gigahertz, such mismatch creates significant reflections that travel back to transmitter side [10]. In addition, in high-speed wireline communications, crosstalk becomes another important disturb. Without a proper termination at transmitter, the crosstalk traveling to transmitter would be reflected to receiver and thus degrade data detection. Therefore, the matched terminations are necessary on both transmitter and receiver sides in high-speed wireline communication systems. Such termination topology [Fig. 2.2] is called double termination.

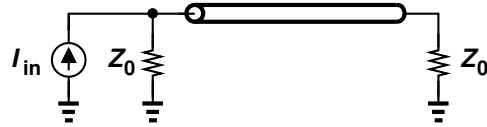


Figure 2.2. Double termination.

2.2 Output Driver

In a transmitter, an output driver is the final stage that delivers current/voltage to the load. An output driver can be modeled by Norton or Thevenin equivalent circuit. To switch between different symbol levels, we can either change the value of the current source in Norton equivalent circuit or the voltage source in Thevenin equivalent circuit. According to this, we have two categories of output drivers, current-mode logic (CML) drivers that steer current to generate logic levels, and voltage-mode drivers which is also called source-series termination (SST) drivers if the termination is matched.

The fundamental power-hungry circuit in a transmitter is the output driver.

For a given voltage swing, this stage must deliver a certain current to the load, e.g., a $100\text{-}\Omega$ differential resistance. In addition, as explained in Section 2.1, on transmitter side, at tens-of-gigahertz speeds, the circuit must also include back-termination resistors on the chip, which are approximately equal to the load impedance. This doubles the necessary supply current for a CML driver or the necessary supply voltage for a SST driver. Moreover, for a CML driver with PAM4 signaling, certain voltage headroom requirements must be met to ensure sufficient linearity as well as a stable output common-mode (CM) level. Thus, the supply voltage well exceeds the single-ended output swing, leading to a low efficiency.

2.2.1 CML Driver

To formulate the driver power consumption, P_{dr} , for a CML PAM4 topology, we consider the structure shown in Fig. 2.3, where half of the most significant bit (MSB) and least significant bit (LSB) stages is shown for simplicity. We can view the circuit as a 2-bit digital-to-analog converter (DAC).

Assuming $R_T = R_L$, and noting that the drain voltage has a CM level equal to $V_{DD} - 3IR_T/2 = V_{DD} - 3IR_L/2$ and a single-ended peak-to-peak swing of $V_{max} = 3I(R_T \parallel R_L) = 3IR_L/2$, we observe that the minimum supply voltage is

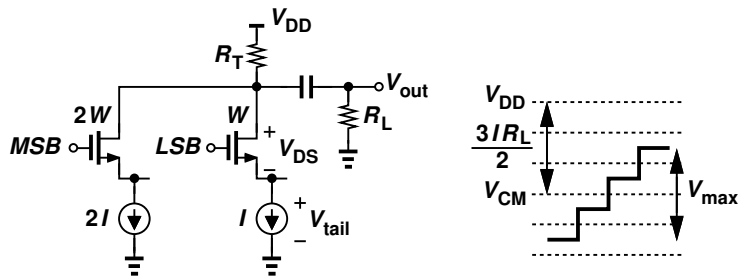


Figure 2.3. CML driver.

given by

$$V_{DD,min} = \frac{3IR_L}{4} + V_{max} + V_{DS} + V_{tail} , \quad (2.3)$$

where V_{DS} and V_{tail} denote the minimum allowable drain-source voltage for the output transistors and the tail currents, respectively. It follows that

$$V_{DD,min} = 1.5V_{max} + V_{DS} + V_{tail} , \quad (2.4)$$

yielding a power consumption of

$$\begin{aligned} P_{dr} &= V_{DD,min}(3I) \\ &= (1.5V_{max} + V_{DS} + V_{tail}) \frac{2V_{max}}{R_L} \\ &= \frac{3V_{max}^2}{R_L} + \frac{2V_{max}(V_{DS} + V_{tail})}{R_L} . \end{aligned} \quad (2.5)$$

Since $V_{DS} + V_{tail}$ is comparable to V_{max} , the second term is nearly equal to the first. For example, if $V_{max} = 350$ mV and $V_{DS} + V_{tail} \approx 500$ mV, and $R_L = 50 \Omega$, we have $P_{dr} \approx 14.35$ mW. The key point here is that the driver power consumption is given by a few fundamental parameters and cannot be reduced significantly. Note that these results also apply to NRZ output stages to some extent, with only V_{DS} being slightly more flexible due to the relaxed linearity requirement in that case.

2.2.2 SST Driver

The foregoing analysis can be repeated for voltage-mode drivers, specifically, those using SST drivers [6, 7, 8]. Depicted in Fig. 2.4 is a single-ended PAM4 SST driver. Such topology incorporates two scaled inverters and series termination resistors R_{T1} and R_{T2} . The choice of $R_{T1} = 1.5R_L$ and $R_{T2} = 3R_L$ yields uniformly distributed PAM4 levels with a maximum single-ended swing of $V_{max} = V_{DD}/2$, and $R_{T1} || R_{T2} = R_L$ ensures proper back termination [6]. In this case,

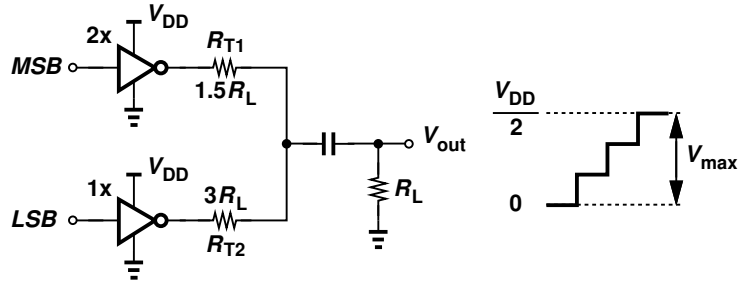


Figure 2.4. SST driver.

the inverter transistors must be so wide as to contribute an output resistance well below their respective series resistors.

Different from CML drivers, this circuit's power consumption is a function of the output voltage. When MSB and LSB are equal, the equivalent circuit [Fig. 2.5(a)] of the differential topology with two such drivers operating differentially gives us a power consumption of $V_{DD}^2/4R_L$. When MSB and LSB are opposite, the equivalent circuit in Fig. 2.5(b) leads to a power consumption of $17V_{DD}^2/36R_L$. If both cases happen with equal probability, the PAM4 SST driver exhibits an average power consumption given by $13V_{DD}^2/36R_L = 13V_{max}^2/9R_L$. For a single-ended swing of 350 mV, we could choose $V_{DD} = 700$ mV and obtain a total power of $13V_{DD}^2/36R_L = 3.54$ mW.

As for NRZ SST drivers, we can treat it as a PAM4 topology with LSB always

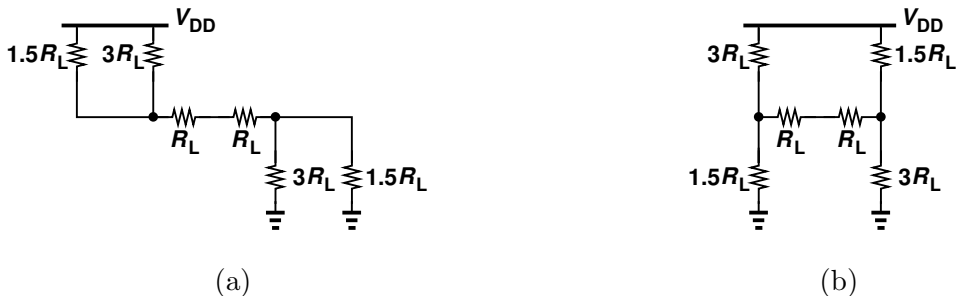


Figure 2.5. Equivalent circuit of PAM4 SST driver when (a) MSB and LSB are equal, and (b) MSB and LSB are opposite.

equal to MSB. Thus, its power consumption is given by $V_{DD}^2/4R_L$.

While draining less power than its CML counterpart, the SST stage faces other issues at high speeds. First, an SST output stage requires rail-to-rail input swings and presents a high input capacitance arising from both NMOS and PMOS transistors in the inverters. Second, the junction capacitance of both NMOS and PMOS transistors in the inverters also leads to a large self-load at the output. As a result, the output of an SST driver exhibits a rapidly degraded eye diagram at high data rate because its rail-to-rail input cannot switch fast enough compared to the small UI. Figure 2.6 plots the comparison of the middle eye opening of a PAM4 CML driver and a PAM4 SST driver outputs in 45-nm technology. In this example, at 80 Gb/s, the SST driver plus its pre-drivers totally draws 14 mW, a number comparable to that of a CML driver.

Third, the actual series resistance, which is R_{T1} or R_{T2} plus the MOSFET conducting resistance, varies not only with process-voltage-temperature (PVT) but also with the output voltage. Although PVT variation is a slowly changing

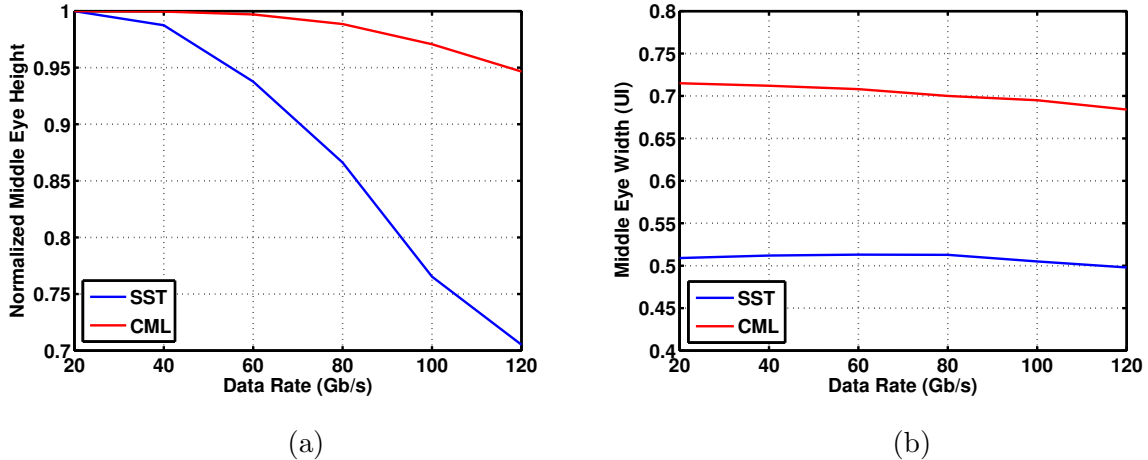


Figure 2.6. Comparison of PAM4 CML and SST drivers' outputs on (a) vertical eye opening, and (b) horizontal eye opening.

part and therefore can be calibrated by a control loop, the variation with the output voltage changes at the data rate, whose effect can be only weakened by larger transistor sizes, which again increases the input capacitance.

Fourth, since the current pulled by an SST driver from V_{DD} depends on the output voltage and thus fluctuates roughly at the data rate, even a package inductance of 100 pH would cause significant ringing.

2.3 Output Swing

In Chapter 1, we see that the relation between signal swing and noise amplitude determines BER. For example, BER of 10^{-12} requires an amplitude $V_0 = 7\sigma_n$ for NRZ signaling. Suppose the resulted 3.25% jitter from the additive noise (Chapter1: Section 1.5) is acceptable in our case, then we ask is it enough for a NRZ transmitter to deliver a swing of $14\sigma_n$?

To answer this question, we model the transceiver front end and the channel in Fig. 2.7. The transmitter front end, the channel and the receiver front end are abstracted to three transfer functions, H_{TX} , H_{ch} and H_{RX} , respectively. V_{TX} is the transmitted signal. $\sigma_{n,TX}$ and $\sigma_{n,RX}$ denote the noise arising from the transmitter and the receiver. The actual channel noise sources spread all over the channel. But we simplify them into two parts, $\sigma_{n,ch1}$ and $\sigma_{n,ch2}$, on transmitter and receiver sides, respectively.

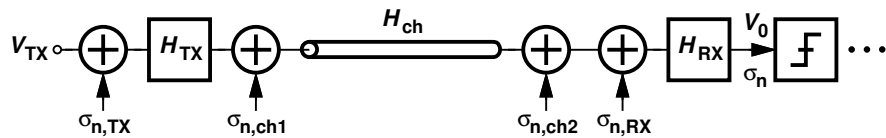


Figure 2.7. Noise model of transceiver front end and channel.

We express the SNR at the input of the receiver slicer as

$$\begin{aligned}
 SNR_{tot} &= \frac{V_0^2}{2\sigma_n^2} \\
 &= \frac{\frac{1}{2}V_{TX}^2 H_{TX}^2 H_{ch}^2 H_{RX}^2}{\sigma_{n,TX}^2 H_{TX}^2 H_{ch}^2 H_{RX}^2 + \sigma_{n,ch1}^2 H_{ch}^2 H_{RX}^2 + (\sigma_{n,ch2}^2 + \sigma_{n,RX}^2) H_{RX}^2}.
 \end{aligned} \tag{2.6}$$

Inverting the denominator and the numerator, we get a more meaningful version as

$$\frac{1}{SNR_{tot}} = \frac{2(\sigma_{n,TX}^2 + \sigma_{n,ch1}^2/H_{TX}^2)}{V_{TX}^2} + \frac{2(\sigma_{n,ch2}^2 + \sigma_{n,RX}^2)}{V_{TX}^2 H_{TX}^2 H_{ch}^2}. \tag{2.7}$$

The first term corresponds to the SNR on transmitter side while the second term represents the added portion of SNR on receiver side. Since the lossy channel heavily attenuates the transmitted signal: $V_{TX}^2 H_{TX}^2 H_{ch}^2 \ll V_{TX}^2$, the second term in Eq. (2.7) is much larger than the first term and therefore dominates the final SNR seen by the slicer. Therefore, the transmitter output swing should satisfy that even after the channel attenuation, the total SNR before the slicer is still high enough to reach the target BER. The offset and the sensitivity of the receiver front end even worsen the situation and thus desire a even larger output swing from the transmitter.

2.4 Skew Between MSB And LSB Paths

Driving a PAM4 output driver requires two-bit inputs, MSB and LSB. The delay mismatch (skew) between the two paths leads to output distortion. The PAM4 eye diagram depicted in Fig. 1.8(b) is the case without the skew. If the four levels are binary coded, the outline of the middle eye is defined by the transitions of $00 \leftrightarrow 10$ and $11 \leftrightarrow 01$. Similarly, the top eye is encircled by the transitions of $11 \leftrightarrow 10$ and $00 \leftrightarrow 11$, and the bottom eye by the transitions of $00 \leftrightarrow 01$, and $00 \leftrightarrow 11$.

If LSB lags MSB by 0.2 UI, the transitions of $00 \leftrightarrow 10$ and $11 \leftrightarrow 01$ keep unchanged because only MSB changes in these cases; the transitions of $00 \leftrightarrow 01$ and $11 \leftrightarrow 10$ happen later by 0.2 UI because only LSB changes. For the cases where both MSB and LSB change, the transition of $01 \leftrightarrow 10$ happens earlier by an amount smaller than 0.2 UI because the output actually tries to become 11 first and then 10. This middle state (1) introduces a glitch at the beginning of the transition, and (2) leads to a larger slope at the start and consequently a faster transition. Vice versa, the transition of $00 \leftrightarrow 11$ happens later.

According to the foregoing analysis, the top and the bottom eyes would happen later than the middle eye. In addition, the top and the bottom eyes would become asymmetric. And the transition region spreads more while transitions are not aligned. Figure 2.8(a) depicts the case with a late LSB.

If clock and data recovery (CDR) in the receiver takes multiple transition cases, the generated clock would carry large amount of jitter. On the other hand, if CDR only takes one transition case, the clock sampling point would be at most

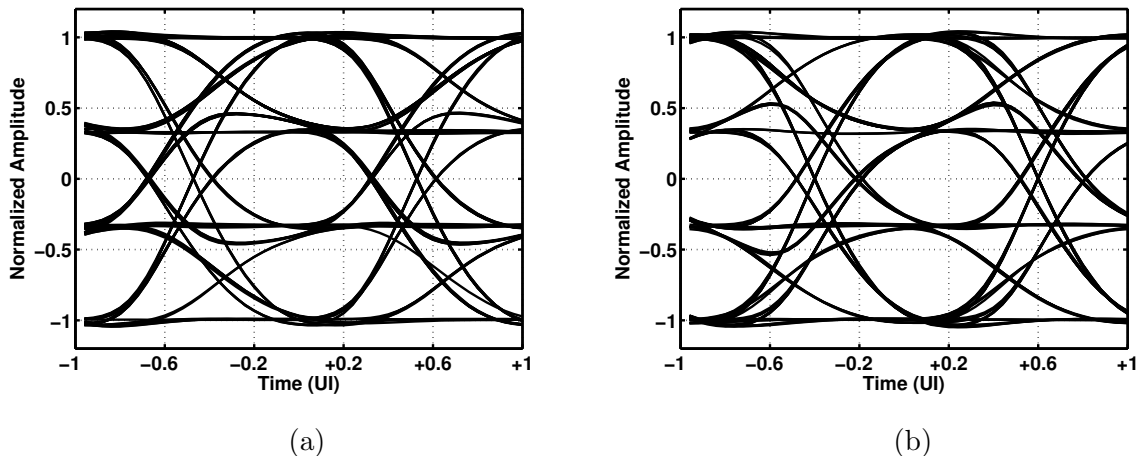


Figure 2.8. PAM4 output eye diagrams with (a) LSB lags MSB by 0.2 UI, and (b) LSB leads MSB by 0.2 UI.

ideal for either the middle eye or the top and the bottom eyes, but not for all three eyes. Besides, hardware mismatches inside PAM4 output drivers and along channels may worsen the case and make the distortion even complicated. Therefore, zero skew between MSB and LSB paths is desired in PAM4 transmitters.

2.5 Timing Of Serializer

Figure 2.9 depicts a typical building block of serializers, which includes two ranks of 2-to-1 multiplexers (MUXes), three latches and a divide-by-two circuit. In practice, the delay exists all over the circuits. Specially, this building block suffers from the divider delay, $t_{CK_1 \rightarrow CK_2}$, and the MUX delay, $t_{CK_2 \rightarrow D_{even}}$, $t_{CK_2 \rightarrow D_{odd}}$ and $t_{CK_1 \rightarrow D_{out}}$.

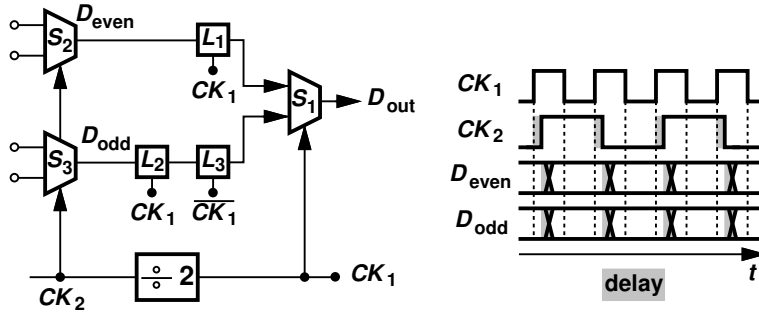


Figure 2.9. Conventional structure and timing of serializer.

In order to let L_1 and L_2 sample D_{even} and D_{odd} successfully, before and after the latching phase when CK_1 becomes low, D_{even} and D_{odd} should settle down well and keep stable for enough time, respectively. We write these conditions as

$$t_{CK_1 \rightarrow CK_2} + t_{CK_2 \rightarrow D_{even/odd}} + t_{setup} + t_{skew} \leq \frac{1}{2}T_{CK_1}, \quad (2.8)$$

and

$$\frac{1}{2}T_{CK_1} + t_{CK_1 \rightarrow CK_2} + t_{CK_2 \rightarrow D_{even/odd}} - t_{skew} \geq t_{hold}, \quad (2.9)$$

where T_{CK_1} denotes the period of CK_1 ; t_{setup} and t_{hold} are the setup and hold time of the latches, respectively; t_{skew} comes from delay mismatches of buffers and routings between CK_1 and CK_2 .

For high-speed communications, the first condition in Eq. (2.8) becomes extremely tight. For example, if D_{out} runs at 40 Gb/s, $T_{CK_1}/2$ is only 25 ps. For 45-nm technology, the divider delay or the MUX delay itself could be already on the level of 25 ps. Thus, their sum is hard to meet Eq.(2.8). As for the second condition [Eq. (2.9)], although it is relatively loose compared to the first one, verifying it across PVT corners is still necessary as $T_{CK_1}/2$ gets small in high-speed communications.

2.6 Jitter On Output Data

Suppose S_1 in Fig. 2.9 is the last stage of the serializer and drives an output driver directly, since every transition of D_{out} is triggered by CK_1 edges, the deviation of CK_1 zero crossings from ideal locations is directly translated to the jitter on D_{out} and finally the transmitter output with a gain roughly equal to one.

To avoid such jitter, a flip-flop is usually applied as a retimer before the output driver as shown in Fig. 2.10. However, the retimer draws significant power to drive the large input capacitance of the output driver at full rate. In addition, the retimer, S_1 and the first divider entail a timing condition twice tighter than Eq. (2.8). Therefore, it merits study on the jitter translation in case of no retimer.

The displacement of CK_1 zero crossings contains a random part due to noise, and a deterministic part. The deterministic part comes from (1) duty cycle distortion (DCD), and (2) clock skew from delay mismatches. The effects of two

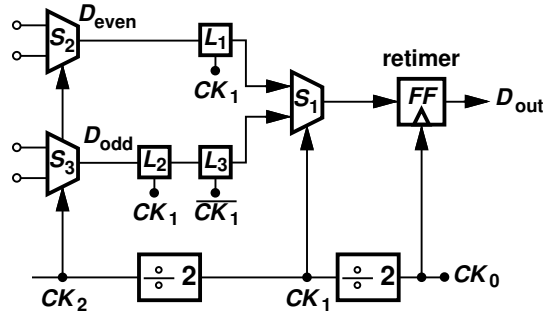


Figure 2.10. Conventional configuration with retimer.

mechanisms depend on the duty cycle of CK_1 .

2.6.1 Duty Cycle = 50%

To study the translation from clock edges to the MUX output edges, we use a 2-to-1 CML MUX in Fig. 2.11 as an example. We care about the jitter on differential output while the clock edges directly affect the edges of single-ended output. Specially, in Fig. 2.11, clock rising edges only affects the single-ended output that goes down; and the falling edges only affects the ones that goes up.

Figure 2.12 depicts the situation with DCD only. ΔT_H is the amount that how much the pulse widths deviate from the ideal value. When CK goes up and \overline{CK} goes down, both edges happen at the correct time. The transition of V_{out} does not deviate from the ideal location. When CK goes down and \overline{CK} goes up,

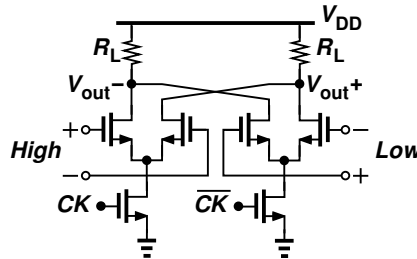


Figure 2.11. A 2-to-1 CML MUX.

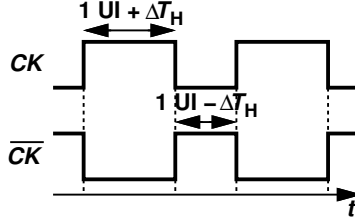


Figure 2.12. DCD of 50%-duty-cycle clocks.

since both edges happen later than the correct time by ΔT_H , V_{out+} , V_{out-} and V_{out} are all delayed by ΔT_H . Therefore, the deviated portion of the pulse width, ΔT_H , results in a peak-to-peak jitter of ΔT_H at the MUX output, that is,

$$J_{pp} = \Delta T_H. \quad (2.10)$$

The situation only with skew is shown in Fig. 2.13. When CK goes up and \overline{CK} goes down, ideally, in the region of transition,

$$\begin{cases} V_{out+} = kt \\ V_{out-} = -kt, \end{cases} \quad (2.11)$$

where $k(> 0)$ denotes the transition slope of the single-ended output; and the reference $t = 0$ is at the middle of the output transition. Thus, $V_{out} = V_{out+} - V_{out-} = 2kt$. The ideal zero crossing happens at $t = 0$.

If, like the dashed waveforms in Fig. 2.13, \overline{CK} lags CK by ΔT_{sk} , the output transitions become

$$\begin{cases} V_{out+} = k(t - \Delta T_{sk}), \\ V_{out-} = -kt. \end{cases} \quad (2.12)$$

Thus, in the region where V_{out+} and V_{out-} both change, we have

$$V_{out} = 2k\left(t - \frac{1}{2}\Delta T_{sk}\right). \quad (2.13)$$

Therefore, the zero crossing moves to $t_0 = \Delta T_{sk}/2$. Similarly, for the case when CK goes down and \overline{CK} goes up, and the cases where the polarities of the two inputs are switched, we get the zero crossings all at $\Delta T_{sk}/2$. The result is equivalent to delaying V_{out} by $\Delta T_{sk}/2$. Therefore, as long as the skew of 50%-duty-cycle clocks is smaller than the output transition time, it does not translate to the jitter at the MUX output.

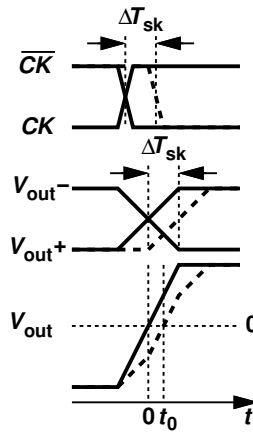


Figure 2.13. Skew of 50%-duty-cycle clocks.

2.6.2 Duty Cycle = 25%

In some cases, we may also design the multiplexing ratio to be 4-to-1 and use 25%-duty-cycle clocks, ϕ_1 - ϕ_4 , to drive a direct 4-to-1 MUX.

For the case only with DCD [Fig. 2.14], the high times of ϕ_1 - ϕ_4 incur errors equal to ΔT_{H1} - ΔT_{H4} , respectively, where $\Delta T_{H1} + \Delta T_{H2} + \Delta T_{H3} + \Delta T_{H4} = 0$. We observe that the falling edge of ϕ_1 and the rising edge of ϕ_2 at $t = t_1$ are displaced by ΔT_{H1} , those of ϕ_2 and ϕ_3 at $t = t_2$ by $\Delta T_{H1} + \Delta T_{H2}$, etc. Thus, the corresponding transitions on the differential output move by ΔT_{H1} and $\Delta T_{H1} + \Delta T_{H2}$, respectively. Therefore, the peak-to-peak jitter at the MUX output can

be expressed as

$$J_{pp} = \max(\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4) - \min(\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4), \quad (2.14)$$

where $\epsilon_1 = \Delta T_{H1}$, $\epsilon_2 = \Delta T_{H1} + \Delta T_{H2}$, etc.

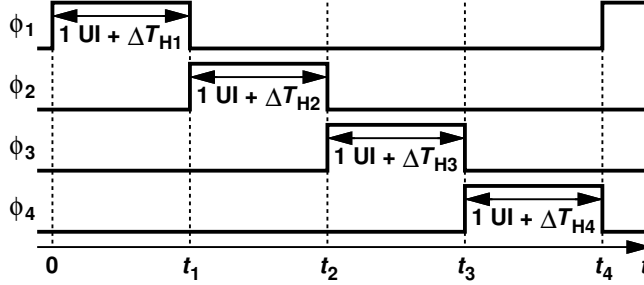


Figure 2.14. DCD of 25%-duty-cycle clocks.

The effect of skew is illustrated in Fig. 2.15, where we assume the falling edge of ϕ_1 incurs an error of ΔT_{sk1} , and the rising edge of ϕ_2 , an error of ΔT_{sk2} . In this case, ideally, we have

$$\begin{cases} V_{out+} = +kt \\ V_{out-} = -kt. \end{cases} \quad (2.15)$$

Thus, the ideal zero crossing is at $t = 0$. Due to the skew, the transitions become

$$\begin{cases} V_{out+} = +k(t - \Delta T_{sk2}) \\ V_{out-} = -k(t - \Delta T_{sk1}). \end{cases} \quad (2.16)$$

As a result, in the region where V_{out+} and V_{out-} both changes,

$$V_{out} = 2k\left(t - \frac{\Delta T_{sk1} + \Delta T_{sk2}}{2}\right). \quad (2.17)$$

Therefore, the differential output of the MUX suffers from a zero-crossing displacement equal to $(\Delta T_{sk1} + \Delta T_{sk2})/2$. Extending this result to all four phases, we have

$$J_{pp} = \max(\delta_1, \delta_2, \delta_3, \delta_4) - \min(\delta_1, \delta_2, \delta_3, \delta_4), \quad (2.18)$$

where $\delta_1 = (\Delta T_{sk1} + \Delta T_{sk2})/2$, $\delta_2 = (\Delta T_{sk2} + \Delta T_{sk3})/2$, etc. These results are for differential outputs; the single-ended output jitter can be shown to be *larger*.

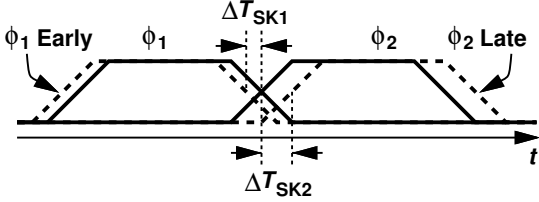


Figure 2.15. Skew of 25%-duty-cycle clocks.

CHAPTER 3

Equalization In Transmitter

In wireline communications, signals are inevitably degraded by low-pass channels. In order to help receivers make correct data detection, we must (1) compensate the degradation from low-pass channels, and (2) maintain the bandwidth wide enough along data paths to avoid furthermore degradation. The equalization techniques widely used nowadays aim to compensate and reduce the effect of channels so as to provide sufficient margin for receiver slicers to make right decisions. On the other hand, inductive peaking plays an important role in the family of broadband techniques.

3.1 Pre-Emphasis

As Eq. (1.1) and (1.2) show, since the data can be decomposed to be a summary of pulses with modulated amplitudes and different delays, therefore, the study begins with the channel effect on a single pulse. Shown in Fig. 3.1 are transmitted and received pulses of a logical ONE. Due to the low-pass channel, the received pulse (1) rises slowly and reaches a lower level, making the slicer hard to detect ONE, and (2) engages a long tail that lasts for several UIs and is superposed on the following symbols, disturbing their detection.

On transmitter side, what can we do with the degraded received pulse? First, to improve the lowered high level, can we send a pulse with a higher amplitude?

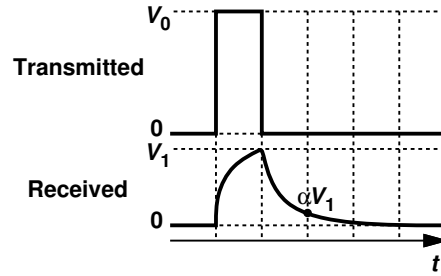


Figure 3.1. Transmitted and received pulses.

Unfortunately, the maximum swing that a transmitter can deliver is fundamentally limited by hardware headrooms and the supply. Second, can we cancel the long tail based on the information we have on transmitter side? If serializing function is correct, the transmitter front end knows all the symbols it delivers. For the example in Fig. 3.1, since it is known that the long tail of ONE after the channel may disturb the detection of the following symbols, the transmitter delivers another delayed, inverted and scaled ONE such that the interference from the early ONE is reduced and even cancelled at the moments of detecting the following symbols [Fig. 3.2]. The key point here is that it is not the long tail being completely cancelled but its interference at the moments of sampling the following symbols being reduced or removed.

To describe the equalization clear, people call the current symbol as the main cursor, the past symbols as the post-cursors and the future symbols as the pre-cursors. As shown in Fig. 3.3, if the pulse becomes wider than 1 UI due to channel dispersion such that the best sampling point lies after the transition beginning by more than 1 UI, at the sampling moment, since the following symbol already starts its transition, as a result, the following symbol may disturb the detection of the current symbol. This is the reason why some equalizations take pre-cursors into account.

Such equalization technique bears multiple names, pre-emphasis, pre-amplify

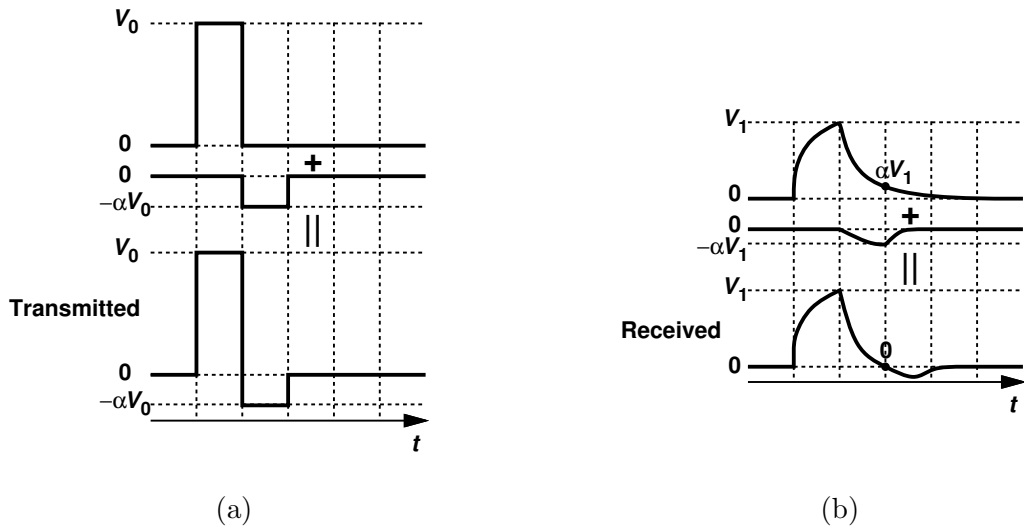


Figure 3.2. (a) Transmitted waveform, and (b) received waveform with equalization.

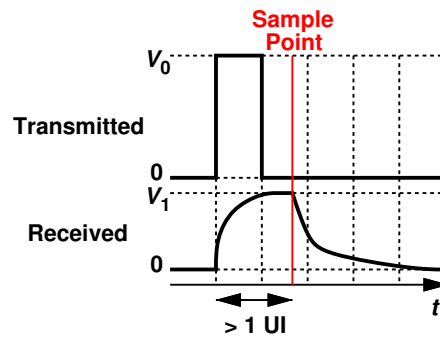


Figure 3.3. Reason for pre-cursors.

and feedforward equalization (FFE). But the truth behind the names is that it is actually the low-frequency components that are attenuated (de-emphasis).

Figure 3.4 depicts an ideal two-tap pre-emphasis (the main cursor and the 1st post cursor). The transfer function is

$$\frac{D_{out}}{D_{in}} = 1 - \alpha Z^{-1}. \quad (3.1)$$

With $Z = e^{-j\omega}$, the amplitude frequency response is

$$\left| \frac{D_{out}}{D_{in}} \right| = \sqrt{1 + \alpha^2 - 2\alpha \cos \omega}. \quad (3.2)$$

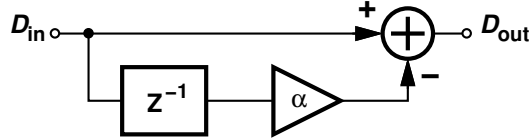


Figure 3.4. A two-tap pre-emphasis.

According to the above equation, the gain magnitude at DC is $1 - \alpha$ while at $\omega = \pi$ is $1 + \alpha$, where $\omega = \pi$ corresponds to Nyquist frequency according to sampling theorem. To intuitively understand this, we suppose the maximum allowed output amplitude is V_0 , without the pre-emphasis, the nominal voltage of ONE is $+V_0$ and ZERO is $-V_0$. In one extreme case, if D_{in} keeps to be ONE at $+V_0$, then the actual output, D_{out} , is a DC voltage of $(1 - \alpha)V_0$. In the other extreme case, if D_{in} keeps toggling between ONE and ZERO, D_{out} toggles between $(1 + \alpha)V_0$ and $-(1 + \alpha)V_0$ with an amplitude of $(1 + \alpha)V_0$. Thus, we gain a boosting of $(1 + \alpha)/(1 - \alpha)$ at Nyquist frequency that equalizes the low-pass channel.

However, notice that with the pre-emphasis in Fig. 3.4, the maximum output amplitude, $(1 + \alpha)V_0$, has exceeded the maximum allowed value, V_0 . In order to keep the boosting, the actual output has to be scaled down by $1 + \alpha$. Figure. 3.5 plots the two frequency responses. As we can see, it is, actually, the DC component that has been attenuated by $(1 - \alpha)/(1 + \alpha)$ so as to create a relative boosting at high frequencies. Intuitively, since the low-pass channel attenuates the fast toggling part more than the less toggling part, thus, the less-toggling part is attenuated intentionally on transmitter side such that after the channel it is on the similar level to the fast toggling part.

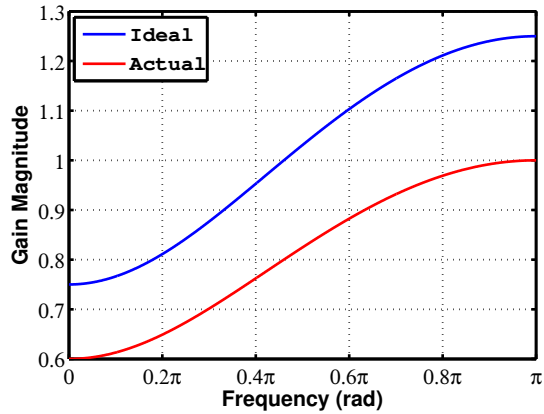


Figure 3.5. Magnitude of de-emphasis transfer functions.

Ideally, the pre-emphasis in transmitters should compensate all of the channel loss. However, since the high boosting ratio is on the penalty of significant attenuation on DC component, we cannot sacrifice too much for boosting and put transmitted signals vulnerable to noise, reflections and crosstalks. The typical boosting ratio of the transmitter pre-emphasis is about 5-6 dB.

3.2 Inductive Peaking

In high-speed wireline communication systems, broadband designs become particularly difficult in transmitter front end. To deliver the required current, the transmitter front end inevitably exploits large transistors, presenting a large input capacitance to the preceding stage. In addition, the large self-load arising from multiple branches in multiplexers also limits the bandwidth. In these circumstances, inductive peaking proves to be useful.

To understand the mechanism of inductive peaking, we begin with simple cases in Fig. 3.6. If a step in I_{in} from 0 to I_0 happens at $t = 0$, in Fig. 3.6(a), the output voltage, V_{out} , jumps instantaneously to $I_0 R$. If a capacitor is put in

parallel with R as shown in Fig. 3.6(b), V_{out} cannot jump immediately but follows an exponential shape:

$$V_{out} = I_0 R (1 - \exp(-\frac{t}{RC})). \quad (3.3)$$

And it takes the time of $5RC$ for V_{out} to reach 99% of $I_0 R$. In this scenario, we see an interaction between R and C because of KCL and KVL. On one hand, although the voltage across R changes immediately with I_R , I_R cannot take all of I_{in} due to I_C . On the other hand, although charging the capacitor takes time, I_C cannot take all of I_{in} , either, to maximize the charging rate.

To break interaction, we insert a switch in series with R in Fig. 3.6(c). After the current step happens, the switch turns off to force all of I_{in} to charge C until V_{out} reaches $I_0 R$ at $t = RC$. After $t = RC$, the switch turns on to switch all of I_{in} through R and makes V_{out} keep at $I_0 R$. In this way, V_{out} takes the time of RC to reach the final value, indicating an remarkable improvement from $5RC$.

Now, the question is how to control the switch to turn on exactly at $t = RC$? Notice that at the beginning, the switch blocks I_{in} through R (keeps I_R to be zero), and after the time of RC , allows the current through R . This reminds us of the operation of an inductor, which tends to resist the current change. Therefore, we replace the switch with an inductor in Fig. 3.6(d), which leads us to inductive peaking.

3.2.1 Inductive Shunt Peaking

In Fig. 3.7, we apply inductive peaking in a real circuit where the other side of the differential structure is not shown for simplicity. Such connection is called “shunt peaking” because the resistor/inductor combination appears in parallel with the output.

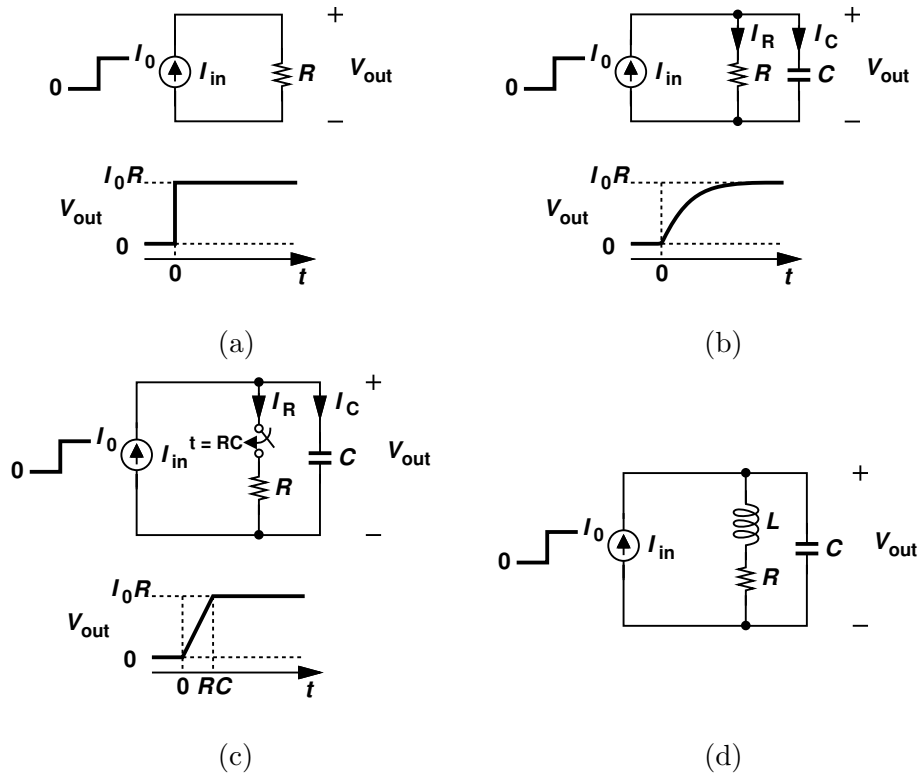


Figure 3.6. Evolution of inductive peaking: (a) only resistor as load, (b) add a capacitor, (c) add a switch in series with resistor, and (d) replace switch with an inductor.

As described before, the switch is desired to turn on exactly at $t = RC$. Replaced by an inductor, the actual operation depends on the inductance. On one hand, if the inductance is too small, the blocking on the resistor current is weak, yielding a negligible improvement on the transition. On the other hand, if the inductance is too large, the capacitor is charged for a too long time so that V_{out} , the capacitor voltage, exceeds the final value, resulting in an overshoot and ringing. Interestingly, although we call it “peaking”, we actually want a flat frequency response, which leads to a fast transient transition but without much overshoot or ringing.

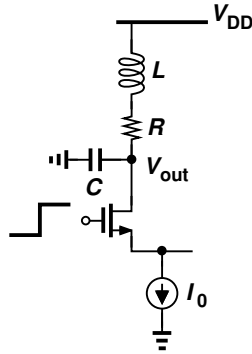


Figure 3.7. Configuration of shunt peaking.

A proper value of inductance can be chosen according to the equation [9]:

$$L = mR^2C, \quad (3.4)$$

where m is a design parameter typically in the range of 0.25 to 0.41 for optimal peaking. $m = 0.41$ yields a maximally flat frequency response of amplitude (MFA) while $m = 0.33$ a maximally flat envelope delay response (MFED). Generally, if the input is a sine wave, we prefer MFA to broaden the bandwidth most; if the input is a square wave, we prefer MFED to keep the square shape. Figure 3.8 compares the magnitude of frequency responses and transient waveforms. The MFA increases the bandwidth by 72% and the MFED 57%. As the step response in Fig. 3.8(b) shows, the MFED keeps the flat shape of the step waveform while MFA introduces a little overshoot although its magnitude of frequency response is maximally flat.

3.2.2 Inductive Series Peaking

In some circumstances, the shunt peaking by monolithic inductors becomes hard in layout floor plans because the supply needs to be routed to both inductors and the core circuit of transistors, which two are usually far away. In some occasions such as cascaded stages, the capacitor can be split into two parts, the output

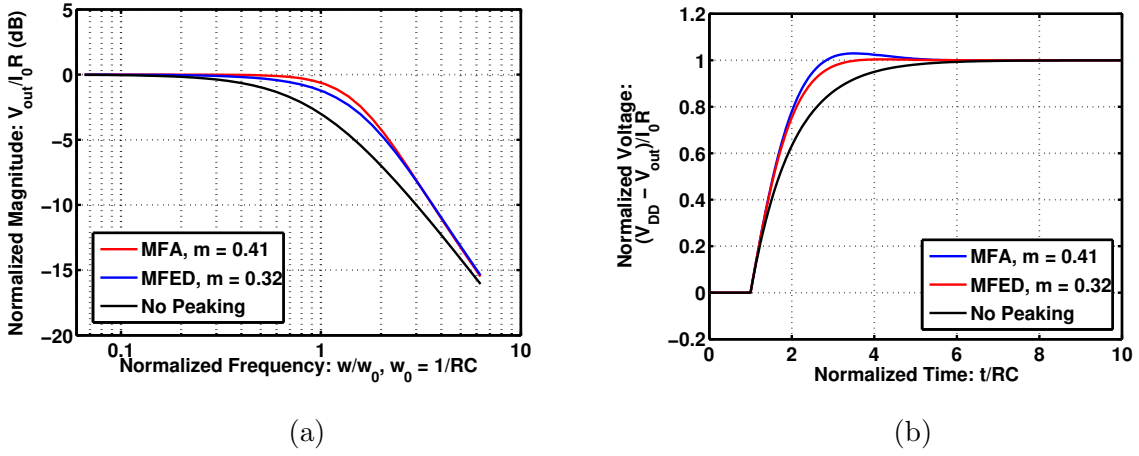


Figure 3.8. Shunt peaking: (a) frequency responses, and (b) transient waveforms.

capacitance of the first stage and the input capacitance of the second stage. This provides the possibility of inductive series peaking. Shown in Fig. 3.9(a) is a circuit with inductive series peaking.

In Fig. 3.9(b), the equivalent circuit is divided into two parts, I and II. If we treat the inductor as a switch for the first-order approximation, right after the current step happens, V_1 shows an exponential transition of part I with a time constant $\tau = RC_1$; after the switch turns on at $t = 5RC_1$, nearly all current flows to part II, charging V_{out} linearly with a slope I_0/C_2 . Thus, the total transition time is about $5RC_1 + RC_2$, smaller than $5R(C_1 + C_2)$ without peaking.

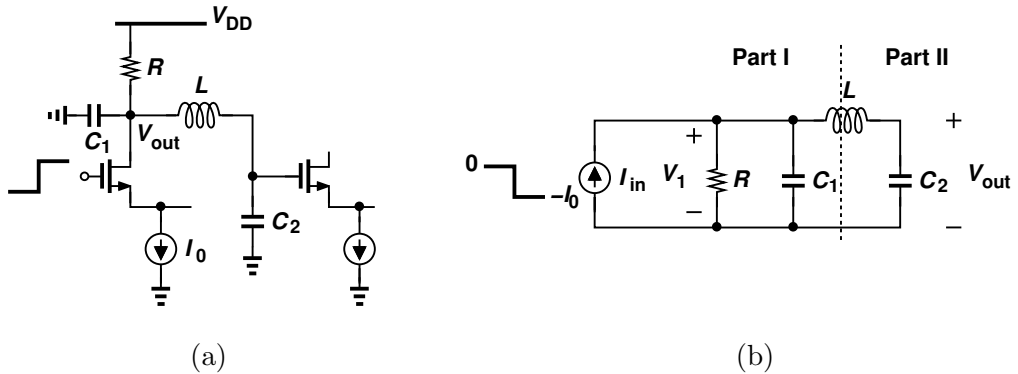


Figure 3.9. (a) Configuration of series peaking, and (b) equivalent circuit.

The former analysis points out an important information is that the part with the resistor shows an exponential transition, which is slower than that of the part with only capacitor. Therefore, the optimal structure of series peaking is (1) put the inductor between two capacitors, and (2) put the resistor on the side of smaller capacitance. Figure 3.10 depicts the two cases. For the case in Fig. 3.10(b), the transition time is about $RC_1 + 5RC_2$. This way ensures that the larger coefficient, i.e., 5, is always together with the smaller capacitance.

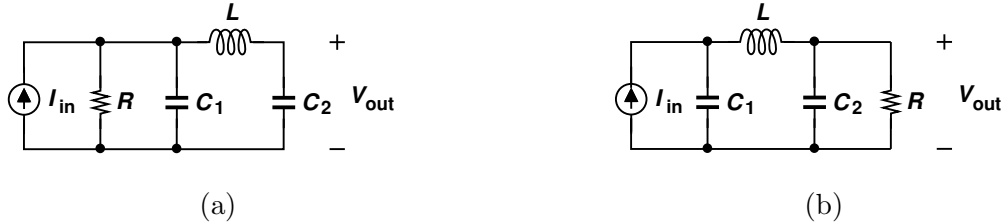


Figure 3.10. Configuration of two cases: (a) $C_1 < C_2$, and (b) $C_1 > C_2$.

Similar to shunt peaking, the inductor value in series peaking also needs proper design [9]:

$$L = mR^2(C_1 + C_2). \quad (3.5)$$

The optimal value depends on the ratio of C_2/C_1 . Figure 3.11 plots frequency responses and transient waveforms of MFA and MFED for $C_1 < C_2$. For the case where $C_1 > C_2$, the values of optimal capacitor ratio correspond to C_1/C_2 . The series peaking extends the bandwidth by 100% in MFA and 75% in MFED.

3.2.3 T-Coil Peaking

As the shunt and the series peakings increase the bandwidth by allowing less current through the resistor when charging the capacitance, “T-coils” provide an alternative means with greater extension by mutual coupling. Intuitively, as shown in Fig. 3.12, if the transistors results in a positive current I_1 through L_1 ,

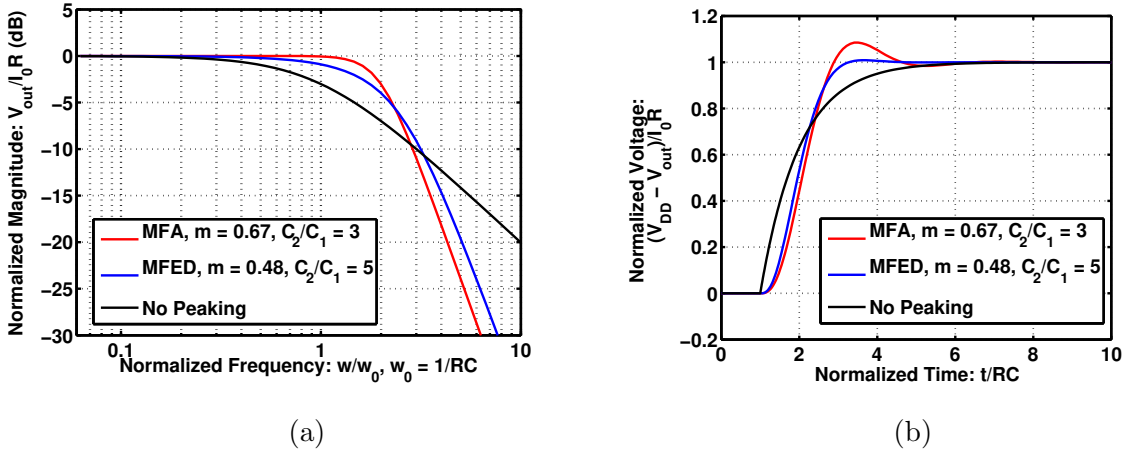


Figure 3.11. Series peaking: (a) frequency responses, and (b) transient waveforms.

the mutual coupling will generate a positive I_2 through L_2 . Thus, the current that discharges C_L is $I_C = I_1 + I_2$, leading to a higher discharging rate than that with I_1 only.

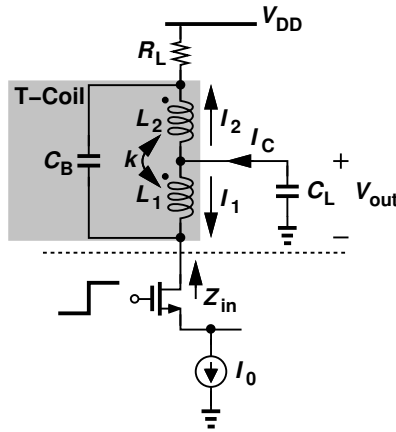


Figure 3.12. Configuration of T-coil peaking.

Another important attribute of T-coil peaking is that the input impedance, Z_{in} , can remain resistive and equal to R_L at *all* frequencies and for *any* value of

C_L if parasitic losses are ignored and the following conditions are satisfied [10]:

$$L_1 = L_2 = \frac{R_L^2 C_L}{2(1+k)}, \quad (3.6)$$

$$\frac{C_B}{C_L} = \frac{1}{4} \frac{1-k}{1+k}, \quad (3.7)$$

$$k = \frac{4\zeta^2 - 1}{4\zeta^2 + 1}, \quad (3.8)$$

where ζ is the damping factor of the transfer function V_{out}/I_{in} (proven to be on second order).

Figure 3.13 compares frequency responses and transient waveforms. The improvement of the bandwidth is 182% of MFA and 172% of MFED. The normalized magnitude and phase of Z_{in} are plotted in Fig. 3.14. Both MFA and MFED exhibit a resistive input impedance $Z_{in} = R_L$ across frequencies.

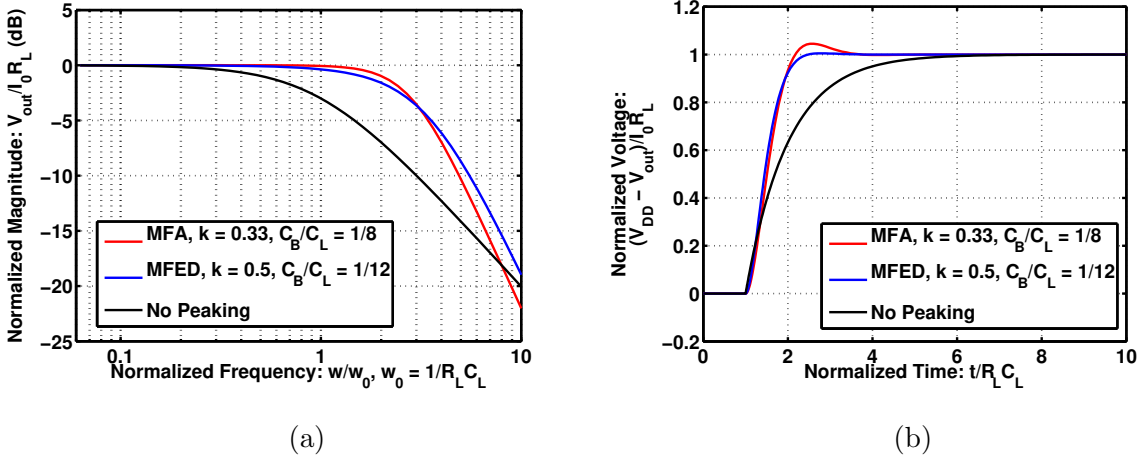


Figure 3.13. Tcoil peaking: (a) frequency responses, and (b) transient waveforms.

The two properties, bandwidth extension and constant resistive input impedance, make T-coil peaking widely used in I/O pads with ESD. The ESD usually introduces large parasitic capacitance, which disrupts the matched termination and limits the bandwidth. Figure 3.15 shows the configurations with T-coil peaking

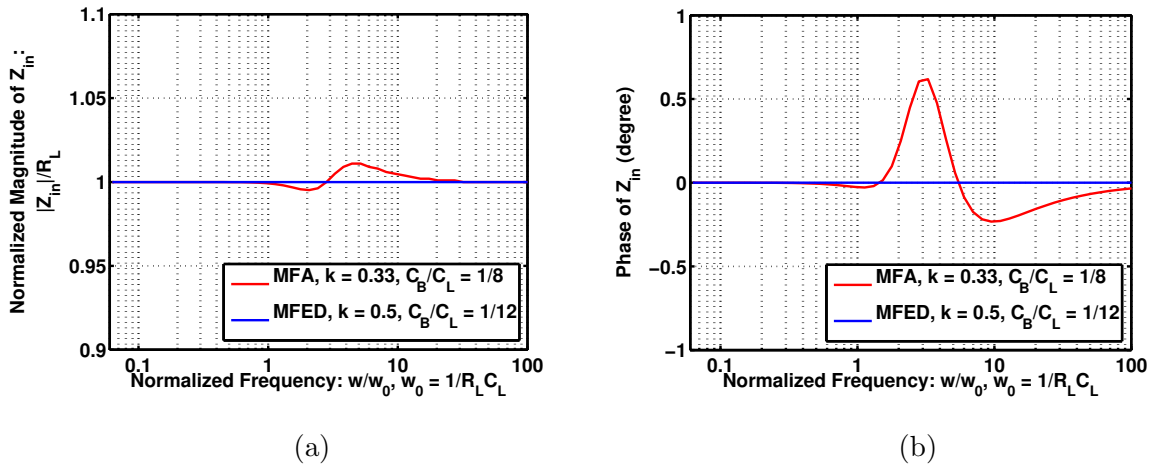


Figure 3.14. (a) Normalized magnitude and (b) phase of input impedance with T-coil peaking.

applied to input and output pads, respectively [11]. Notice that for output pads in Fig. 3.15(b), two ports of the T-coil are swapped. But the good features still keep due to the reciprocity of the peaking network.

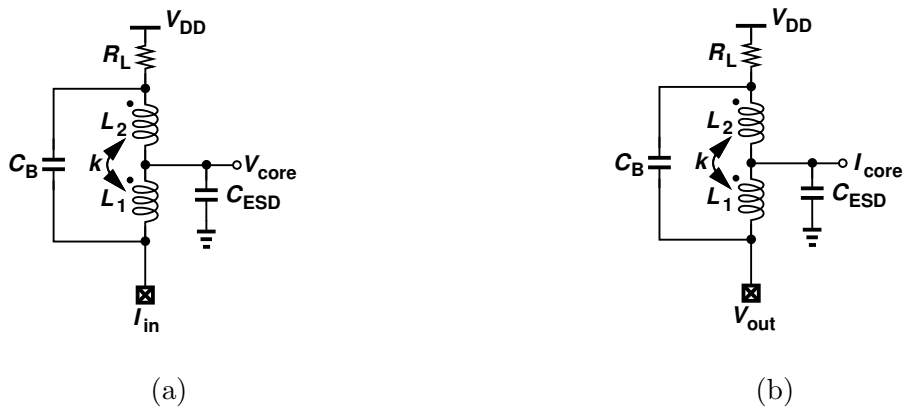


Figure 3.15. T-coil peaking (a) for input pad, and (b) for output pad.

CHAPTER 4

A 40-Gb/s NRZ Transmitter

4.1 Design Considerations

As explained in Chapter 2, in transmitter design, the output driver tends to consume a high power as it must deliver large currents with relatively large voltage swings. This issue is exacerbated at tens of gigabits per second for two reasons: (1) the need for on-chip back-termination resistors doubles the power, and (2) the use of FFE requires additional strength in the driver.

Two other difficulties arise in conventional transmitters due to the use of a full-rate retimer and hence the need for a full-rate frequency divider [Fig. 4.1(a)]. First, the divider must operate at 40 GHz while driving at least two multiplexers, potentially drawing a high power. Second, the divider delay is subtracted from the timing margin available to the retimer, severely limiting the speed [10, 12].

Another challenge in the transmitter front end of Fig. 4.1(a) relates to the driver's large input capacitance, C_{dr} . To achieve sufficient bandwidth ($\approx 0.7 \times 40 \text{ Gb/s} = 28 \text{ GHz}$) at this interface, we can either introduce a predriver or design the retimer with high currents and low impedances, both power-hungry solutions. For example, the two-stage driver in [12] draws 26.4 mW. It is possible to remove the full-rate retimer and divider [Fig. 4.1(b)] so as to avoid the speed and delay constraints imposed by the latter. In this case, however, the clock duty cycle error and the data path mismatches within each MUX directly translate to jitter

at the output.

The transmitter front end can be further simplified if the MUX and driver stages are merged [Fig. 4.1(c)] [14]. Here, the power consumed by the MUXes is not “wasted.” The large input capacitance of the MUX, C_{MUX} , is now driven at 20 Gb/s, and the latches within the MUXes operate at 20 GHz, both still challenging problems. In the next step, we contemplate the use of multi-phase

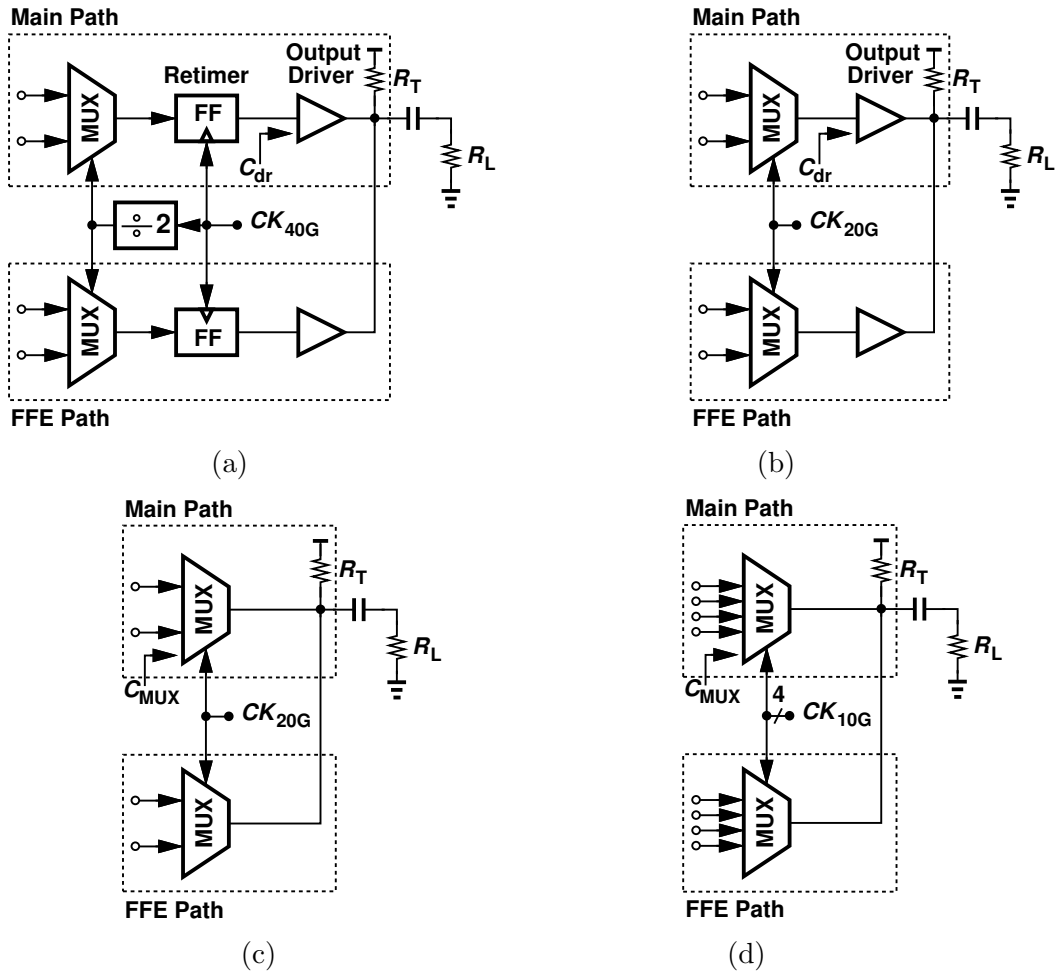


Figure 4.1. (a) Full-rate front end with retimer and divider, (b) half-rate front end, (c) half-rate front end with combined 2-to-1 MUX and driver, and (d) quarter-rate front end with combined 4-to-1 MUX and driver.

multiplexing [15] and combine the idea with Fig. 4.1(c), arriving at the front end shown in Fig. 4.1(d). In this case, C_{MUX} is as large as C_{dr} in Fig. 4.1(a), and is driven at 10 Gb/s, but the main path contains four of these input capacitances. Moreover, multi-phase clocking still requires four latches for each 4-to-1 MUX [12]. Our proposed transmitter architecture addresses both of these issues.

4.2 Transmitter Architecture

Figure 4.2 shows the transmitter architecture. It consists of a main serializer path, an FFE path with a programmable strength, and a phase-locked loop (PLL) for clock generation. The main path comprises a 128-to-8 CMOS MUX, which produces data at a rate of 5 Gb/s, an 8-to-4 “current-integrating” MUX (IMUX), and a 4-to-1 CML MUX. The FFE employs four programmable 4-to-1 CML MUX slices. The PLL receives a reference frequency of 312.5 MHz and delivers 25%-duty-cycle clock phases, ϕ_1 - ϕ_4 , at 10 GHz, 50%-duty-cycle phases, CK_1 - CK_4 , at 5 GHz, etc.

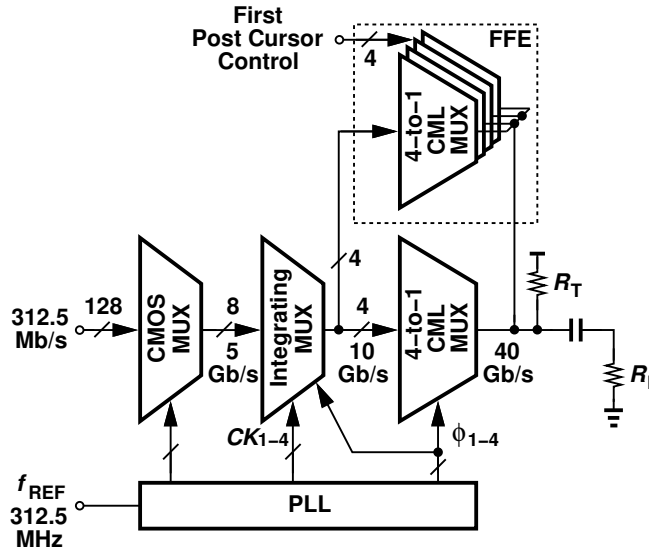


Figure 4.2. Proposed NRZ transmitter architecture.

The proposed transmitter achieves more than a two-fold improvement in the power efficiency as a result of three new concepts: (1) the integrating MUX drives the large input capacitance of the CML MUX with very low power consumption ($410 \mu\text{W} \times 4$ for the 8-to-4 selector), (2) the use of quadrature clock phases with 25% and 50% duty cycles completely eliminates high-speed latches in the data path, and (3) the integrating selector incorporates a timing scheme that readily accommodates the first FFE post cursor.

4.3 Integrating MUX

The direct 4-to-1 MUX/driver in Fig. 4.2 presents two issues, namely, a large input capacitance, $C_{MUX} \approx 96 \text{ fF}$, and proper timing in the preceding stage to guarantee that each input is available when one of ϕ_1 - ϕ_4 is asserted. The integrating MUX efficiently deals with both issues.

In order to drive C_{MUX} at a bit rate of $r_b = 10 \text{ Gb/s}$, we can opt for a CML stage [Fig. 4.3(a)]. Here, we must choose $1/(2\pi R_L C_{MUX}) \approx 0.7r_b$ for minimal ISI, and $R_L = V_0/I_{SS}$ to obtain a single-ended peak-to-peak swing of V_0 . That is, the CML stage consumes $1.4\pi r_b C_{MUX} V_0 V_{DD}$. For example, if $C_{MUX} \approx 100 \text{ fF}$, $V_0 \approx 400 \text{ mV}$, and $V_{DD} = 1 \text{ V}$, the four CML stages driving the 4-to-1 MUX consume a total of 7 mW.

Alternatively, the MUX can be driven by an integrating stage [Fig. 4.3(b)], where first the output is reset to V_{DD} and then the tail current turns on to impress the data level on C_{MUX} . In this case, the power consumption is given by $r_b C_{MUX} V_0 V_{DD}$, a factor of 4.4 lower than that of the CML topology. Additionally, the differential pair transistors in the integrating stage present less input capacitance.

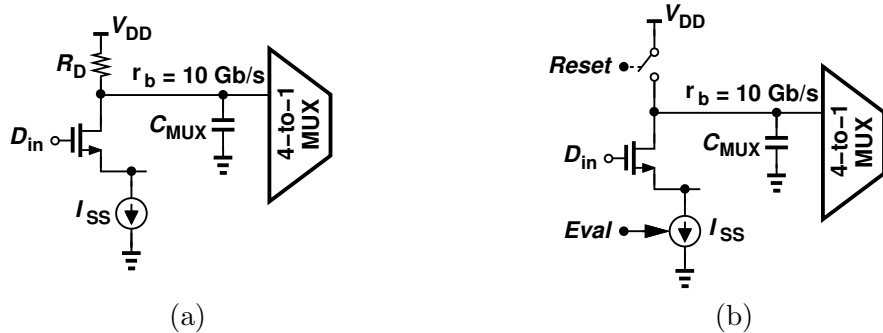


Figure 4.3. Driving the 4-to-1 MUX by (a) a CML stage, or (b) an integrating stage.

It is desirable to incorporate multiplexing within the integrating stage of Fig. 4.3(b). As shown in Fig. 4.4, two differential pairs receive D_{in1} and D_{in2} , but only one is enabled according to the select command, CK_1 . Thus, when ϕ_2 goes high, D_{in1} or D_{in2} travels to the output. As in a standard selector, CK_1 has a 50% duty cycle and the same rate as the inputs (5 GHz), but ϕ_1 and ϕ_2 have a 25% duty cycle and run at 10 GHz, creating much more flexibility in the overall architecture (explained below).

The integrating MUX operates as follows. First, X and Y are reset to V_{DD}

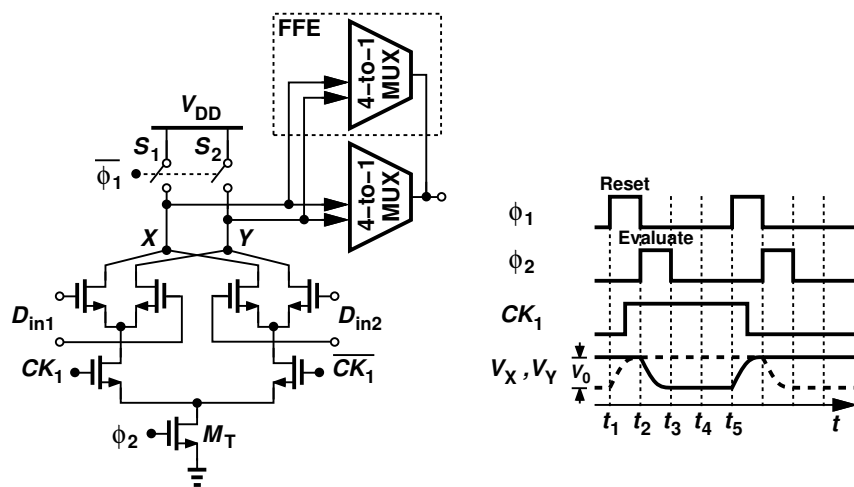


Figure 4.4. Proposed 2-to-1 integrating MUX.

while the tail current source, M_T , is off. Next, CK_1 arrives to select D_{in1} or D_{in2} , and then ϕ_2 rises to perform evaluation. In this mode, V_X or V_Y falls for about 25 ps, providing the desired swing, V_0 . When ϕ_2 goes low, the tail current ceases and the output is held for approximately 50 ps.

The use of both 25% and 50% duty cycles enhances two aspects of the design. First, the main 4-to-1 MUX senses V_X and V_Y from t_3 to t_4 whereas the FFE branch receives these values from t_4 to t_5 . Since this time offset is equal to 1 UI at 40 Gb/s, FFE is implemented with no latches, thus saving power. Second, the topology in Fig. 4.4 provides a hold period, during which V_X and V_Y are constant, so that the subsequent stages can sense the signals reliably. Without the 25%-duty-cycle phases, on the other hand, V_X or V_Y would continue to fall after t_3 , creating unequal swings for the main and FFE paths and hence substantial ISI.

The integrating MUX of Fig. 4.4 merits two more remarks. First, the stacking of transistors still lends itself to a 1-V supply because all of the inputs have rail-to-rail swings. Second, since the value of V_0 is PVT-dependent, the circuit is designed so as to produce a sufficient swing for the 4-to-1 MUXes if M_T is weak and also reset X and Y to V_{DD} in 25 ps if M_T is strong and S_1 and S_2 are weak.

The proposed transmitter employs four integrating multiplexers to serialize data from 8×5 Gb/s to 4×10 Gb/s. These outputs directly drive the direct 4-to-1 multiplexers in the main and FFE paths.

4.4 Main and FFE Multiplexers/Drivers

The waveforms in Fig. 4.4 indicate that the integrating MUX provides a stable output for two time slots, each 25 ps long. We allocate the first slot to the 4-to-1 MUX in the main path and the second to that in the FFE path.

Figure 4.5 shows the main and FFE 4-to-1 MUX/driver circuits in simplified form. Four differential pairs controlled by ϕ_1 - ϕ_4 select one of the inputs for 25 ps, delivering the 40-Gb/s data to the 50- Ω on-chip back-termination resistors and the 50- Ω loads. The differential output voltage swing (without FFE action) is at least 440 mV across PVT corners.

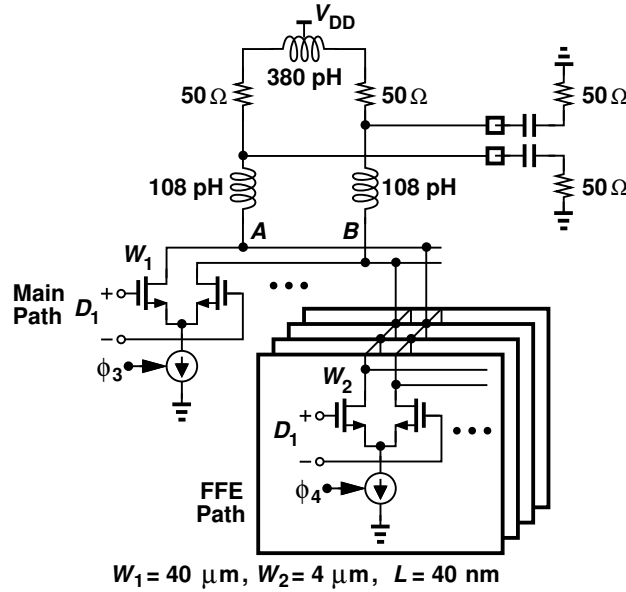


Figure 4.5. Main and FFE data paths.

The FFE path consists of four programmable slices that provide a relative tap coefficient ranging from 0 to 0.4. Each slice contains four differential pairs controlled by ϕ_1 - ϕ_4 and scaled down by a factor of 10 with respect to those in the main path.

The interface between the IMUX and the final drivers is illustrated in Fig. 4.6. Here, three operations occur in succession: first, ϕ_2 is high from t_2 to t_3 for the IMUX to generate proper levels at X and Y ; next, ϕ_3 is high, allowing the 4-to-1 MUX in the main path to sense V_X and V_Y ; last, ϕ_4 is high, enabling the FFE MUX.

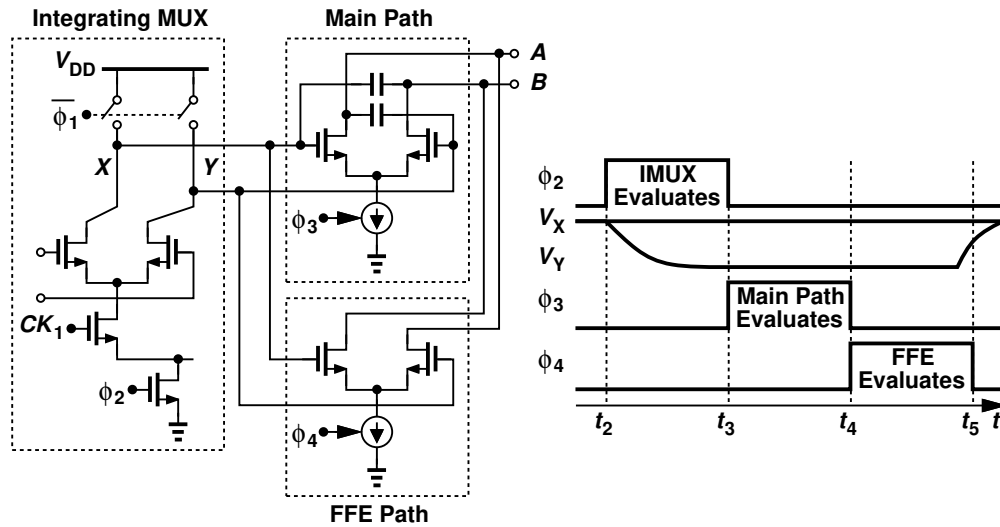


Figure 4.6. Interface between the integrating MUX and the main and FFE drivers/MUXes.

4.5 Experimental Results

The 40-Gb/s NRZ transmitter has been fabricated in TSMC's 40-nm CMOS technology and tested with a 1-V supply. Figure 4.7 shows a photograph of the die, whose active area measures $330 \mu\text{m} \times 175 \mu\text{m}$.

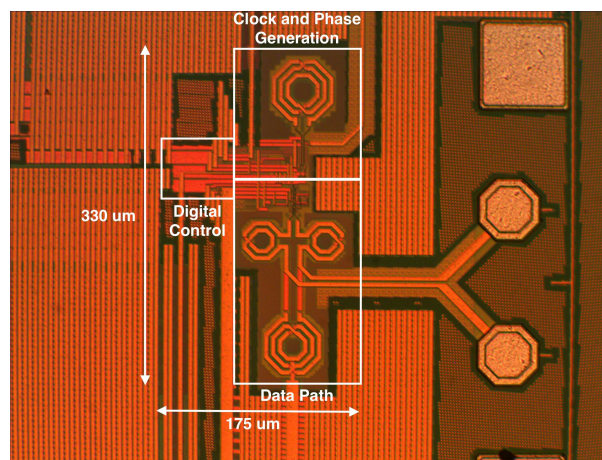


Figure 4.7. TX die photograph.

Figure 5.34(a) plots the measured output spectrum of the PLL at 20 GHz and Fig. 5.34(b) shows the measured phase noise after this clock is divided by 2. The phase noise is -110 dBc/Hz at 10 GHz. Integrated from 10 kHz to 100 MHz, the jitter is equal to $332 f_{s_{\text{rms}}}$. The reference spurs are at -45 dBc.

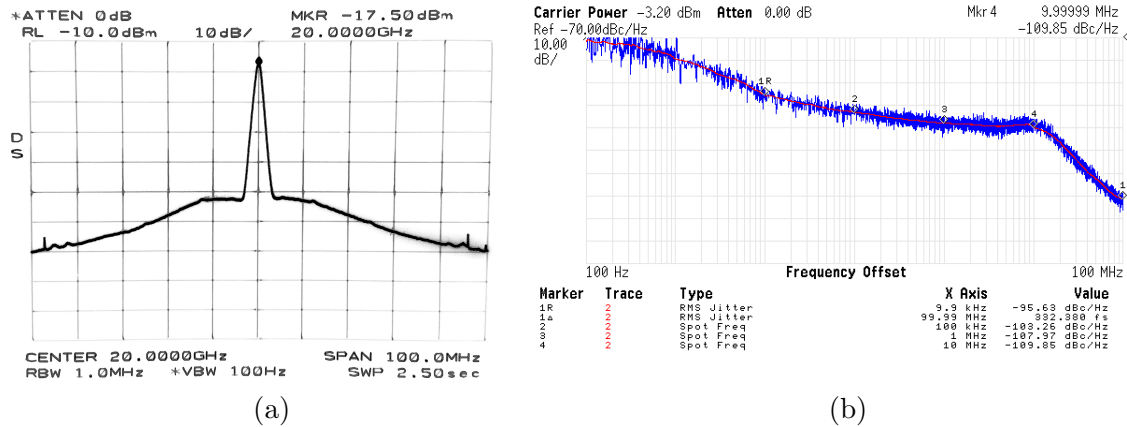


Figure 4.8. (a) Measured spectrum of 20-GHz clock, and (b) phase noise profile of 10-GHz clock.

Figure 4.9(a) shows the transmitter output eye diagram with no FFE action. The differential voltage swing is $460 \text{ mV}_{\text{pp}}$. Figure 4.9(b) shows the output with the FFE tap strength of 0.4, yielding a 7.4-dB boost. The output bit stream has also been captured and checked to ensure correct serialization of the 128 312.5-Mb/s inputs to the 40-Gb/s output.

In order to examine the effect of mismatches (Chapter 2: Section 2.6), we apply the input data so as to create a 20-GHz periodic 0101 sequence at the TX output. The duty cycle and delay mismatches produce spurs at 10-GHz offset. Figure 4.10 shows the single-ended measured spectrum, indicating a spur level of -34 dBc. Translating this value to rms jitter in the single-ended output, we arrive at 225 fs, in reasonable agreement with Monte Carlo simulations.

The transmitter consumes 32 mW from a 1-V supply: 9.0 mW in the main

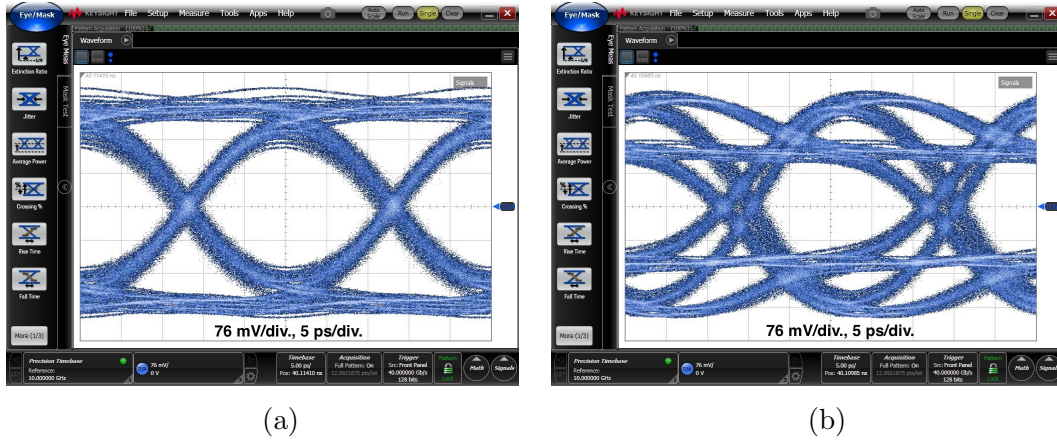


Figure 4.9. Measured eye diagrams with (a) FFE off, and (b) four FFE slices on.

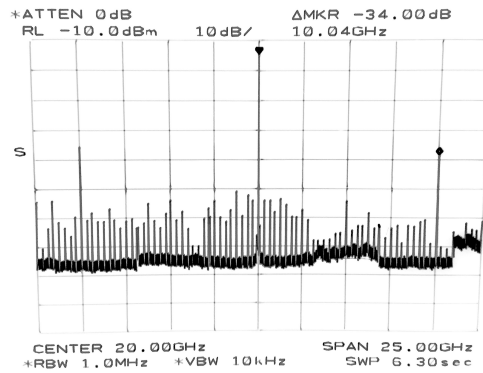


Figure 4.10. Measured spectrum of single-ended output delivering 0101 sequence. and FFE 4-to-1 MUX/drivers, 1.6 mW in the four integrating MUXes, 3.4 mW in the VCO and 18.0 mW in the 128-to-8 serialization, PLL core, divider chain and clock distribution. Table 4.1 summarizes the measured performance of our transmitter and compares it to the prior art. We have achieved a factor of 2.28 improvement in the power efficiency.

Table 4.1. PERFORMANCE SUMMARY.

Reference	Kim ISSCC'15	Hafez JSSC'15	Navid JSSC'15	Huang CICC'15	This Work	
Technology (nm)	14	65	28	65	45	
Data Rate (Gb/s)	16 - 40	31.68 - 48.4	40	40	40	
FFE	4-tap	no	2-tap	2-tap	2-tap	
PN (dBc/Hz)	-	-127.5	-128	-	-104	
f_{offset} (MHz)	-	10	100	-	10	
RMS Jitter (fs)	-	251	162	-	332	
Integ. Range (MHz)	-	0.0001 - 10	10 - 10000	-	0.01 - 100	
Power (mW)	Data Path	-	41.4	130	-	11
	Whole TX*	518**	88	-	80	32
Power Eff. (pJ/bit)	Data Path	-	0.86	3.25	-	0.28
	Whole TX*	12.95**	1.82	-	2	0.4

* Data path and clock path (PLL, phase generation and clock distribution).

** Excluding power of PLL.

CHAPTER 5

An 80-Gb/s PAM4 Transmitter

5.1 Background

A number of PAM4 transmitters operating at tens of gigabits per second have been reported [17]-[22]. Among these, the 56-Gb/s designs in [8] and [18] achieve a power of 101 mW and 200 mW, respectively. The 64-Gb/s transmitter in [19] draws 145 mW. These values exclude the PLL. It is therefore prudent to identify the power-hungry functions in transmitters before deciding on the architecture and its building blocks.

The foregoing analysis of the output driver in Chapter 2 indicates that the power consumption and power efficiency of an output driver is fundamentally limited by the required output swing, back-termination and hardware headrooms. The example of a target 700-mV output swing leads to about 15-mW power consumption of either CML output driver or SST driver plus its pre-driving stages.

To put matters in perspective, we ask, if the driver power can be maintained roughly around 15 mW, where does the remainder of the 100 - 200 mW go in actual designs, e.g., in [8, 18, 19]? We expect that the overall serializer that multiplexes the data from low speeds to the final data rate also draws considerable power. The issue is exacerbated in a PAM4 transmitter owing to the need for two separate MUX chains for the MSB and LSB paths (Section 5.2). For example,

serialization from 312.5 Mb/s to 40 Gb/s (up to the inputs of the output driver) would require 3×254 latches if 3-latch MUX cells are used in a binary tree. Even though the number of latches drops by a factor of 2 from one rank to the next, the increase in speed at least doubles the power per latch. Consequently, the serializer can consume tens of milliwatts in 45-nm technology (Section 5.3).

The generation and distribution of the clock and its divided versions can also draw a high power. Among the prior PAM4 transmitters, [19] includes the distribution in the overall power numbers but not the PLL and phase generation. The design in [18] reports a PLL power of 20 mW at 14 GHz, excluding phase generation and distribution. Thus, the PLL also merits investigation if the overall transmitter power must be minimized.

5.2 Transmitter Architecture

Figure 5.1 shows the proposed transmitter architecture, which consists of MSB and LSB data paths, an output driver/DAC, and a clock generation module. Each serializer consists of a CMOS MUX, a charge-steering MUX, and a direct 4-to-1 MUX. The co-design of the data paths and the PLL allows the former to employ new circuit topologies that substantially reduce the power. Specifically, the feedback dividers provide quadrature phases, ϕ_1 - ϕ_4 , 45° phases, SEL_1 - SEL_4 , etc., making it possible to avoid latches in the entire serializer (Section 5.3).

We should remark that, owing to our “direct-and-conquer” approach, the highest-frequency clock distribution in the transmitter of Fig. 5.1 occurs at 10 GHz rather than 40 GHz. This benefit accrues because the transmitter has been architected such that the direct 4-to-1 MUX operates with a 10-GHz clock and also this MUX is not followed by a retimer. Of course, the distribution of the four

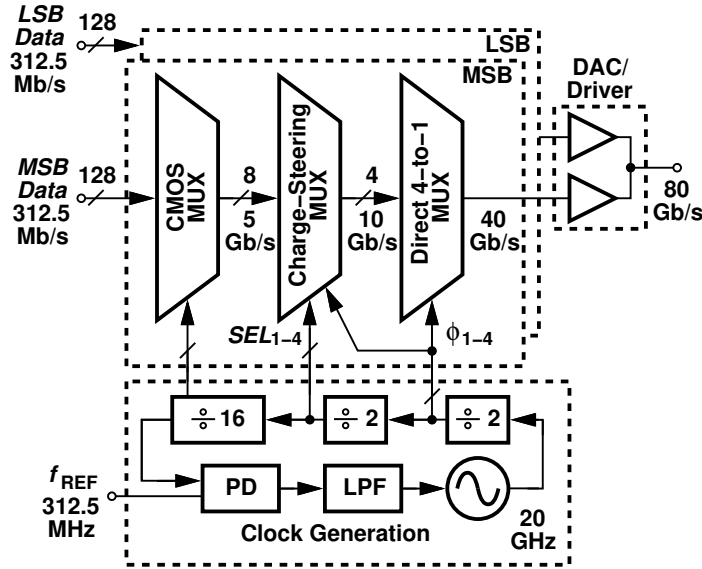


Figure 5.1. Proposed transmitter architecture.

phases, ϕ_1 - ϕ_4 , does require careful layout, and the clock and MUX mismatches must be managed properly.

It is important to recognize the necessity for separating the MSB and the LSB paths. If the data is serialized through a single path, then, it must be deserialized to an MSB and an LSB before it reaches the DAC. That is, the data would need to be multiplexed up to 80 Gb/s in NRZ form and subsequently demultiplexed to two 40-Gb/s streams. Such a transmitter must support an NRZ speed of 80 Gb/s internally and would be much more difficult to design.

The interface between the MSB and LSB serializers and the driver/DAC in Fig. 5.1 entails a critical issue. Since the DAC MSB cell presents twice as much input capacitance as the LSB cell does, the two serializers preceding the DAC must have proportionally scaled drive strengths so to avoid a systematic skew between the MSB and the LSB waveforms. Such a skew manifests itself as jitter and distortion at the final output (Chapter 2: Section 2.4). Thus, the drive strength of the direct 4-to-1 MUX stage in the MSB serializer is scaled up by a

factor of 2, but the stages before this MUX remain mostly unscaled.

5.3 Serializer Design

As mentioned in Section 5.2, the transmitter must employ two serializer paths for the MSB and the LSB, potentially consuming a high power. In this work, we propose a number of techniques to ameliorate this issue: (1) the use of three logic styles in Fig. 5.1 allows the optimum speed-power trade-off, (2) a new “latchless” MUX design, (3) charge steering [23] as a paradigm that affords a higher speed than CMOS logic and a lower power consumption than CML, and (4) a direct latchless 4-to-1 MUX that considerably reduces the number of high-speed stages. We describe these concepts below.

5.3.1 CMOS MUX

Rail-to-rail CMOS logic provides robust operation with a power of the form fCV_{DD}^2 , where f denotes the frequency at which C charges from 0 to V_{DD} . In the context of transmitter design, we must decide on the maximum reliable speed that this style can support. The architecture in Fig. 5.1 comfortably utilizes rail-to-rail stages to serialize the data from 312.5 Mb/s to 5 Gb/s.

The 128-to-8 binary-tree CMOS MUX requires 120 2-to-1 MUX cells. As shown in Fig. 5.2(a), a typical cell comprises three latches and one selector, with L_1 and L_2 holding the inputs so as to block glitches from preceding stages, and L_3 serving to avoid input change when the clock has selected that input. However, if the timing of D_{in1} and D_{in2} is known and well-controlled, L_1 and L_2 can be omitted [Fig. 5.2(b)] [12]. In this case, the assumption is that D_{in1} and D_{in2} change on one edge of the clock and settle before the next edge of the clock.

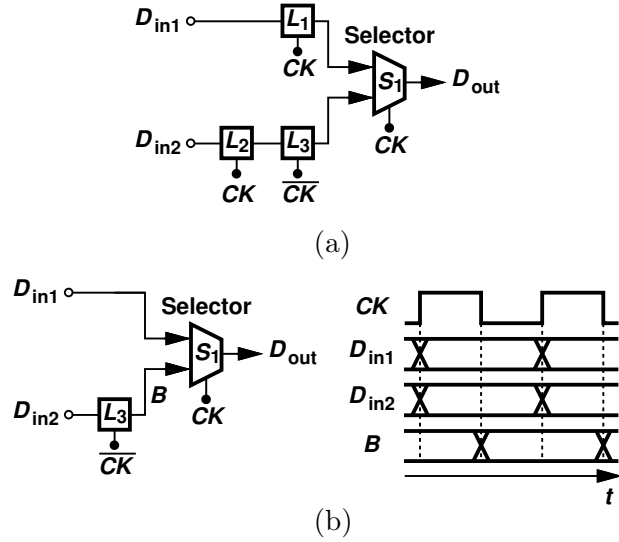


Figure 5.2. (a) Conventional three-latch MUX cell, and (b) simplified MUX cell.

Also, L_3 ensures that the selector inputs do not make simultaneous transitions.

If the multiplexing clock is available in quadrature phases, CK_I and CK_Q , the serializer design can be improved. For example, [18] utilizes such phases to establish a longer hold time for the MUX input. We introduce a new serialization approach that exploits CK_I and CK_Q to eliminate all latches in the data path.¹ Illustrated in Fig. 5.3, the idea is to create the necessary delay between each selector's inputs by proper choice of the clock edges in consecutive stages. Let us consider how D_{even} and D_{odd} avoid simultaneous transitions, noting that selectors S_2 and S_3 are driven by $CK_{2,I}$ and $CK_{2,Q}$, respectively. We make two observations: (1) the edges of these two clocks have an offset equal to $T_{CK2}/4$, and hence D_{odd} changes $T_{CK2}/4$ seconds after D_{even} does, and (2) since the edge separation between CK_1 and CK_2 (≈ 200 ps) is long enough for D_{even} or D_{odd} to settle, no glitch appears at the input of S_1 . Thus, the three-cell structure consisting of S_1 , S_2 , and S_3 can be repeated in the preceding ranks so long as the

¹The use of quadrature clocks does not translate to a power penalty because every selector would need a clock in any other architecture as well.

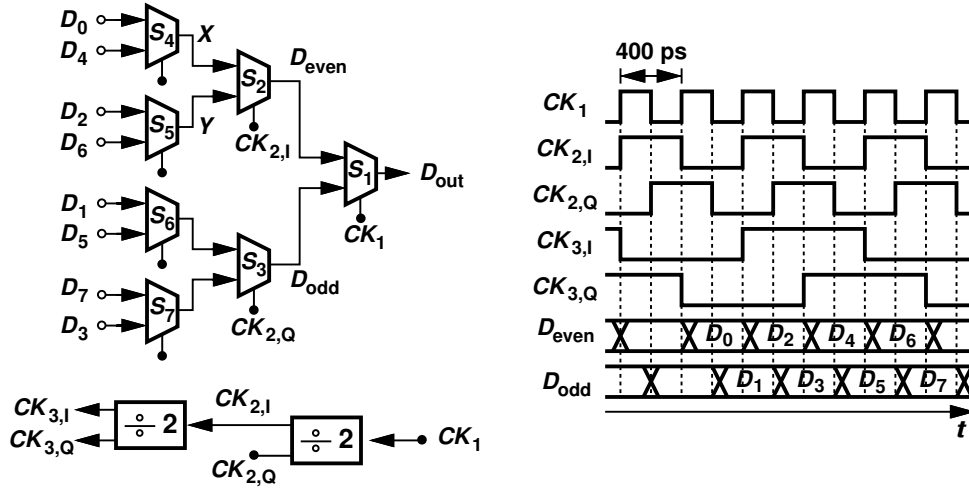


Figure 5.3. Proposed timing scheme to remove latches by applying I and Q clocks.

clock phases are chosen accordingly.

The 120 selectors necessary for multiplexing 312.5 Mb/s to 5 Gb/s can incur a high power consumption in their clock path. We therefore wish to minimize the dimensions of the clocked transistors and the length of the clock wires. On the other hand, the drive strength of the last selector must suffice for the operation of the subsequent (charge-steering) MUX, calling for wide transistors. In addition, S_1 must deliver the final output at 5 Gb/s with small enough delay so as to leave enough timing margin for the charge-steering MUX.

Based on the above considerations, the selector unit is realized as shown in Fig. 5.4(a). This topology occupies a small area - allowing short interconnects for the entire CMOS serializer - and achieves sufficient speed. For S_1 in Fig. 5.3, the transistor dimensions are chosen as $W_N = 1 \mu\text{m}$, $W_P = 1.5 \mu\text{m}$, and $L = 40 \text{ nm}$, leading to a power consumption of $22 \mu\text{W}$ for this unit (in both the data and the clock paths). The eye diagram shown in Fig. 5.4(b) represents this output. Since the stages preceding this selector operate at progressively lower frequencies, the unit design is scaled down by a factor of 2 from one rank to the rank preceding

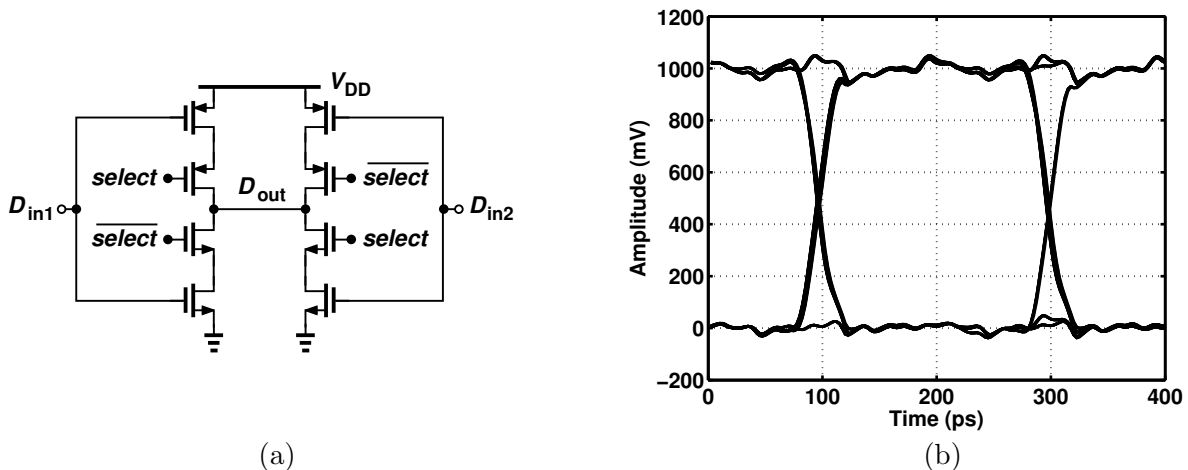


Figure 5.4. (a) CMOS selector used in this work, and (b) simulated output eye diagram of the last stage of CMOS MUX.

it, until a minimum allowable width of 120 nm is reached. Note that the latchless topology does not exhibit glitches because it benefits from ample timing margin between the I and Q edges. Also, the clocking action applied to the selector does not allow device or timing mismatches to accumulate through the serializers. The entire 128-to-8 serializer draws 365 μW in the data path.² The single-ended output is converted to complementary form by means of an inverter after the final CMOS MUX stage.

5.3.2 Charge-Steering MUX

For operation above 5 Gb/s, charge steering proves more viable than CMOS logic. By virtue of their moderate voltage swings ($\approx 300 \text{ mV}_{\text{pp}}$ single-ended), charge-steering circuits achieve a higher speed [23]. We propose a number of techniques that improve the performance of charge-steering stages in the context

²A 3-latch approach would require a power consumption of about 11 mW for the 128-to-8 serializer including the clock path.

of the 8-to-4 MUX in Fig. 5.1.

We begin with the simple charge-steering selector shown in Fig. 5.5. When CK is low, S_1 - S_3 are on, C_T is discharged to ground and X and Y are precharged to V_{DD} . When CK rises, the output begins to track V_{in1} or V_{in2} depending on the logical value of SEL . Capacitor C_T continues to draw charge from X or Y until its voltage reaches approximately one threshold voltage below the input common-mode level, at which point V_X or V_Y approaches its minimum value.³

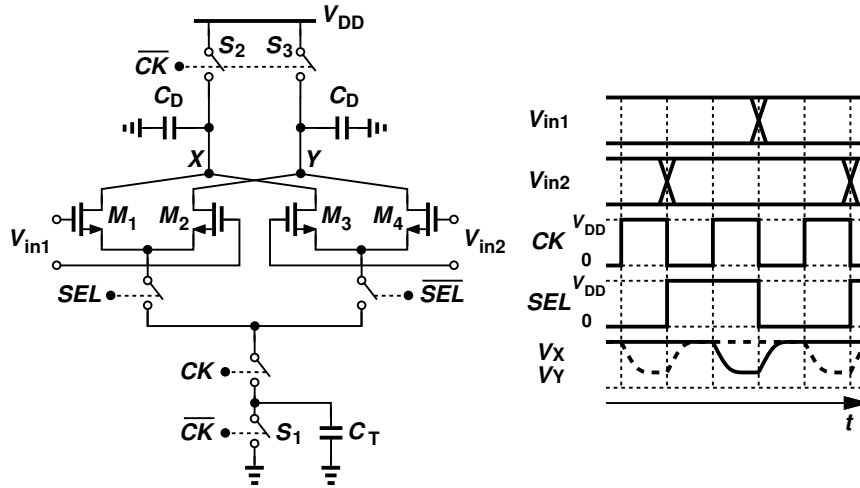


Figure 5.5. Simple charge-steering 2-to-1 MUX.

The reset action at the output nodes removes ISI but occupies about half of the clock cycle, during which the next stage must not sense X and Y . Note that none of the transistors need operate in saturation because the rail-to-rail input and clock swings guarantee complete steering of the charge. In this topology, CK runs at twice the SEL frequency, which itself is equal to the input data rate.

If used in the transmitter architecture of Fig. 5.1, the above charge-steering selector faces a critical issue: the levels produced at X and Y deteriorate due

³The charge-steering MUX does not allow the output low level to reach zero regardless of the clock period, a point of contrast to current-integrating circuits.

to the kickback noise of the next stage, namely, the direct 4-to-1 MUX (Section 5.3.3).⁴ Fortunately, V_{in1} and V_{in2} in Fig. 5.5 are produced by the CMOS serializer and have rail-to-rail swings. Exploiting these swings, we add a small helper of a PMOS selector to the circuit as depicted in Fig. 5.6(a). Here, for a given input state, one of M_5 - M_8 conducts, providing a resistive path from X or Y to V_{DD} [Fig. 5.6(b)], and hence restoring the high level even in the presence of kickback noise from the next stage. Figure 5.7 plots the selector’s simulated output waveforms with and without the PMOS differential pairs, indicating an improvement of about 100 mV in the high level. The PMOS devices primarily restore the output common-mode level, providing a greater voltage headroom for the direct 4-to-1 MUX tail devices.

Another difficulty in the charge-steering selector design is that, at 10 Gb/s, nodes X and Y in Fig. 5.5 do not precharge to V_{DD} completely, thereby suffering from ISI and degraded levels. This is alleviated by introducing switch S_F in Fig. 5.6(a), which ensures $V_X \approx V_Y$ during precharge.

Since the above selector’s output is unavailable in the precharge mode, the 8-to-4 charge-steering MUX and the direct 4-to-1 MUX in Fig. 5.1 must be co-designed to ensure compatibility between their timings. We propose the use of quadrature clock phases with 25% duty cycle for both. To this end, we modify the selector’s clocks as shown in Fig. 5.8. Here, the clock phase RST performs precharge and reset for 25 ps and the EVL phase evaluates the input also for 25 ps. The command SEL selects one input after each precharge interval. Thus, the output is available from t_3 to t_4 .

The 8-to-4 MUX requires four two-input selectors whose timings must agree with those of the direct 4-to-1 MUX. This is accomplished as illustrated in

⁴This issue is also present if a current-integrating MUX is used.

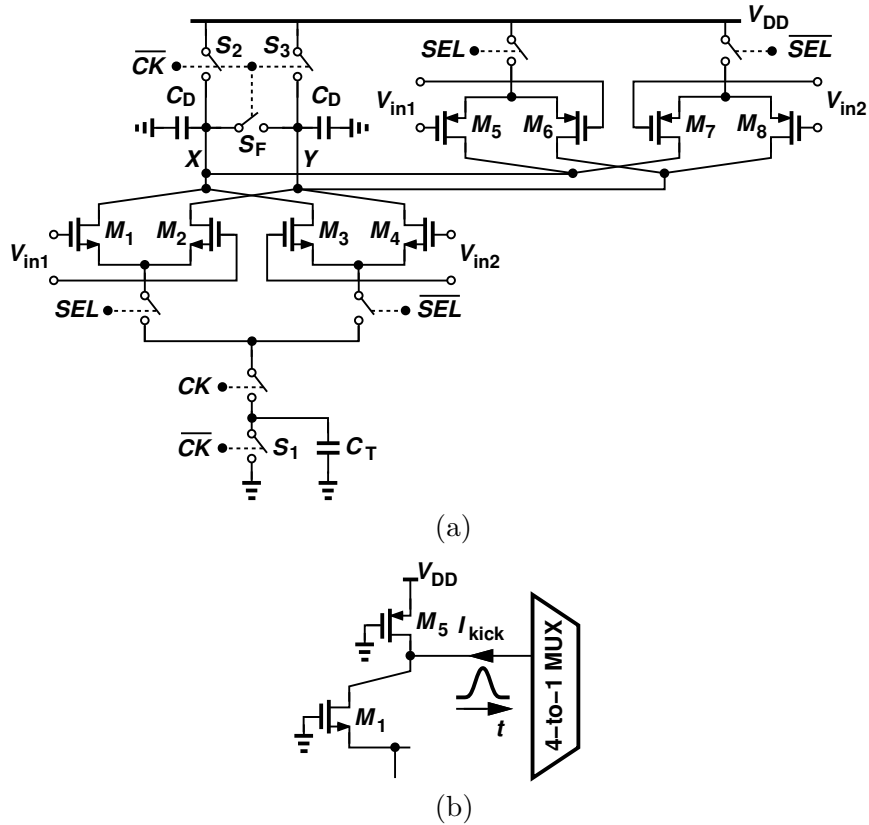


Figure 5.6. (a) Proposed charge-steering MUX, and (b) role of PMOS pull-up device in suppressing the effect of kickback noise.

Fig. 5.9(a), where ϕ_1 - ϕ_4 denote the four phases of the 10-GHz clock with 25% duty cycle and SEL_1 - SEL_4 are the 45° phases of the 5-GHz clock with 50% duty cycle. The first selector on the left operates with ϕ_1 and ϕ_2 in the same manner as in Fig. 5.8, i.e., $RST = \phi_1$, $EVL = \phi_2$. For the next selector, ϕ_2 and ϕ_3 act as RST and EVL , respectively, and SEL_2 , which is 25 ps behind SEL_1 , drives the SEL input. The remaining two selectors run on other rotated phases, and the four outputs D_a - D_d appear in succession.

The idealized situation depicted in Fig. 5.9(a) assumes a zero delay between the rising edge of ϕ_2 and the rising edge of SEL_1 and similarly for other phases. In reality, however, SEL_1 is obtained by frequency division and incurs a delay

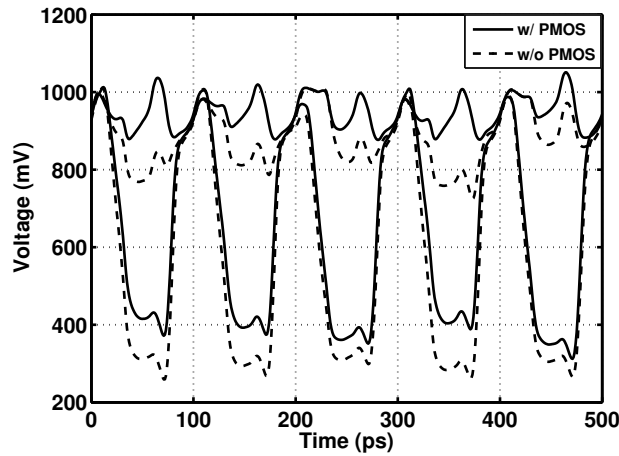


Figure 5.7. Simulated output waveforms of charge-steering MUX with and without PMOS differential pairs.

of about 20 ps. Thus, the charge-steering action is delayed by this amount, shortening the time available for evaluation to about zero. To resolve this issue, we recognize that the select command in Fig. 5.8 can be asserted even before *EVL* arrives. We therefore apply SEL_4 , rather than SEL_1 , to the first selector and rotate the rest accordingly. Figure 5.9(b) shows the resulting assignment of SEL_1 - SEL_4 .

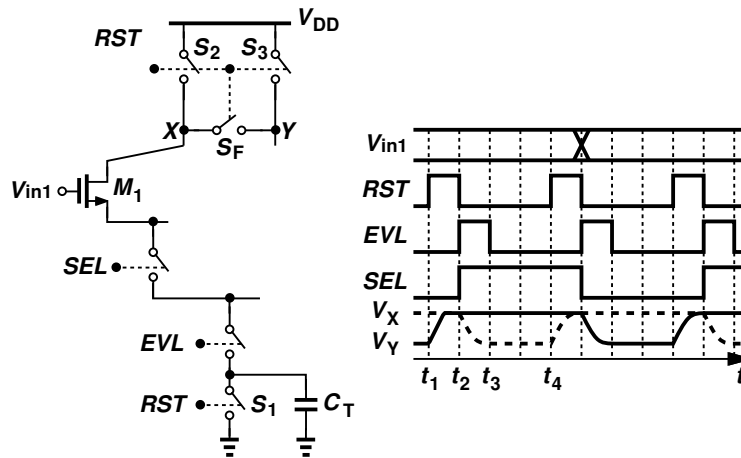


Figure 5.8. Timing diagram of charge-steering MUX with 25% duty-cycle clocks.

While the present prototype does not include feedforward equalization (FFE), our scheme makes it possible to add FFE with minimal power penalty. We briefly explain the idea here based on the charge-steering MUX of Fig. 5.6(a) and refer to a similar FFE implementation based on an integrating MUX in Chapter 4. To create a post cursor tap, we first decompose the following direct 4-to-1 MUX and the output driver into, for example, four slices, three of which are driven by the main cursor and the fourth by the post cursor. Since the MUX of Fig. 5.6(a)

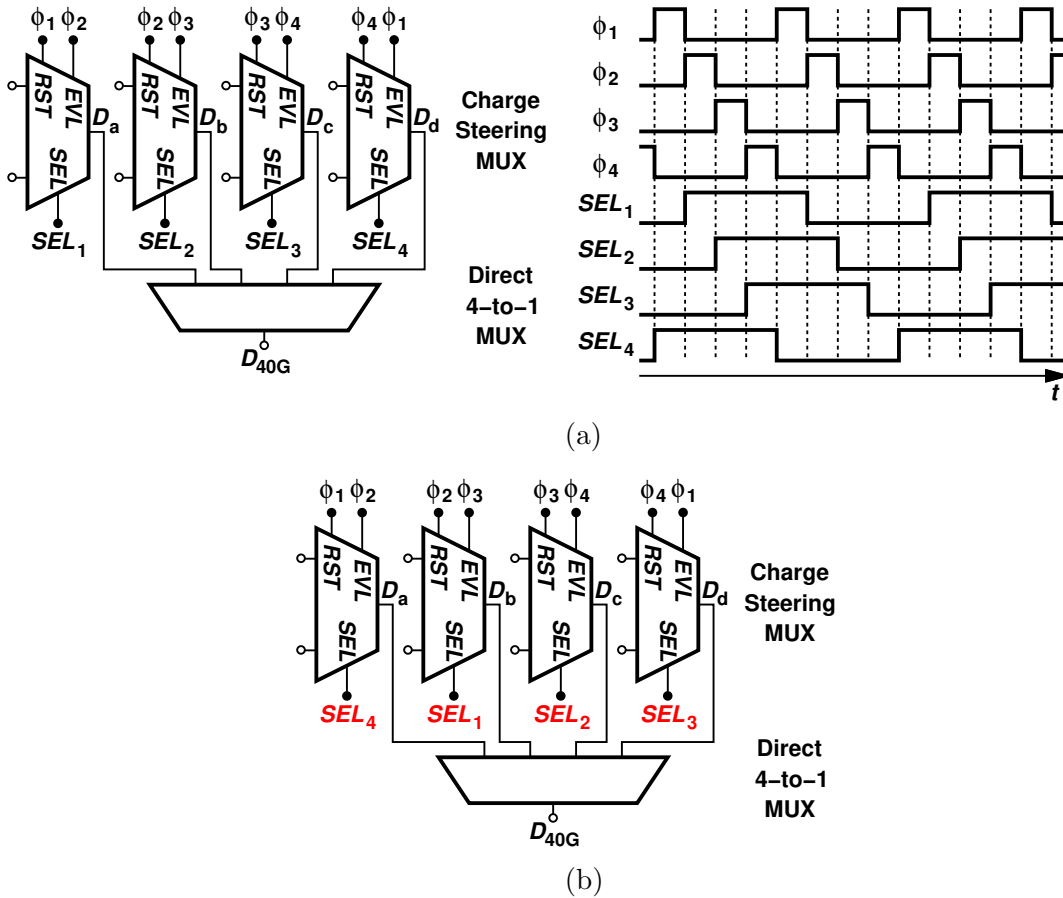


Figure 5.9. (a) Timing diagram of charge-steering MUX with 25% duty-cycle clocks, (b) four charge-steering MUXes with idealized waveforms, and (c) rotation of SEL_1 - SEL_4 in four charge-steering MUXes to accommodate the clock delay.

holds the output for 2 UI [Fig. 5.8], the second UI (from t_3 to t_4) can be used to drive the post cursor without adding any latches. This overall strategy can be applied to both the MSB and the LSB paths.

5.3.3 Direct 4-to-1 MUX

Serialization of data from 10 Gb/s to 40 Gb/s in 45-nm CMOS technology inevitably calls for CML implementations. Let us consider the 4-to-1 binary-tree topology shown in Fig. 5.10(a), where each selector is preceded by one latch to avoid simultaneous input transitions. This arrangement must employ six tail currents (a total of 12 for MSB and LSB paths) and also deal with the loss of timing budget due to the divider delay. Moreover, in each of the MSB and LSB paths, at least four clocked transistors plus the divider are driven at 20 GHz and at least eight at 10 GHz. We can ask whether the latchless serialization described in Section 5.3.1 is applicable here as well. Such an approach would save a total of six high-speed latches but would necessitate quadrature phases of the 10-GHz clock with a 50% duty cycle. The charge-steering MUX, on the other hand, requires 25%-duty-cycle phases at 10 GHz. We must therefore develop a CML MUX that can operate with the latter.

We opt for a direct 4-to-1 CML structure that can utilize these phases. Figure 5.10(b) depicts the result. The four differential pairs are enabled in succession such that each senses an input that is evaluated and held by the preceding charge-steering selector. Inductive peaking deals with the heavy capacitive load (≈ 82 fF for the MSB path and ≈ 40 fF for the LSB path) presented by the large input transistors of the next stage (the output driver/DAC) and the self-load from the four differential pairs.

Direct 4-to-1 MUX topologies have been reported [12], but our approach

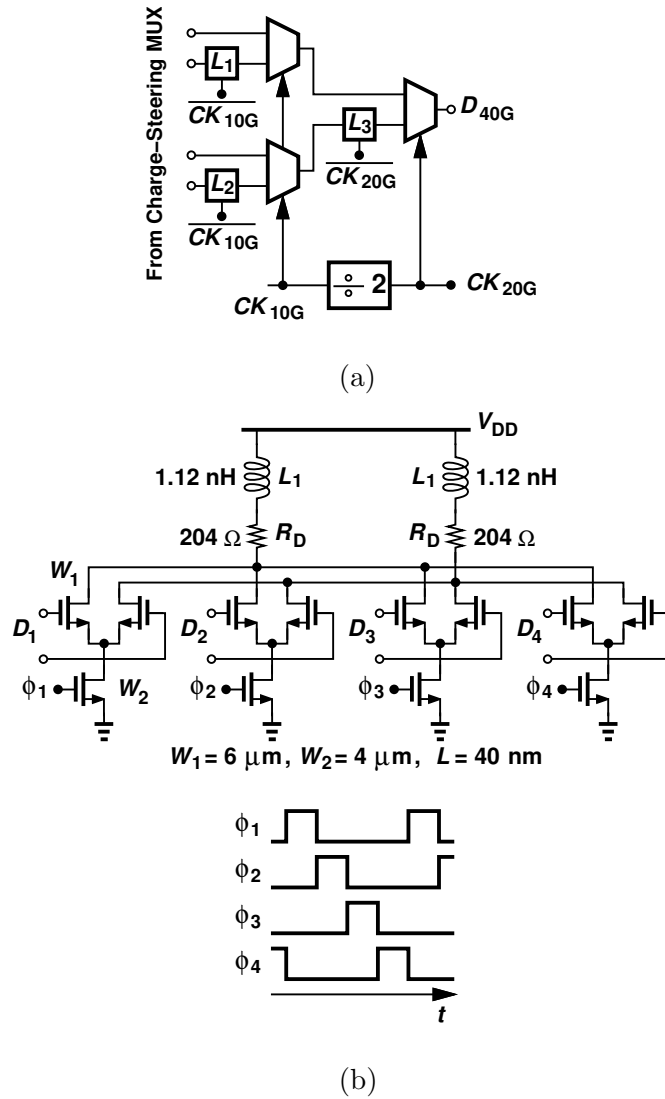


Figure 5.10. (a) Binary-tree 4-to-1 MUX, and (b) direct 4-to-1 CML MUX.

merits some remarks. First, at a clock frequency of 10 GHz, the use of single clocked transistors driven by ϕ_1 - ϕ_4 proves more efficient than generating overlapping quadrature phases and using stacked transistors to perform a NAND gate [25]. Second, with rail-to-rail swings for ϕ_1 - ϕ_4 , the clocked transistors need only be 8 μm wide for the MSB path and 4 μm wide for the LSB path to draw a sufficient current, but the MUX output swing exhibits some dependence upon

PVT. Nevertheless, so long as the output swing is large enough to ensure complete switching in the following driver, this dependence is benign. The values shown in Fig. 5.10(b) correspond to the LSB path; the design is linearly scaled up by a factor of 2 for the MSB path.

In Section 5.4, we address the task of generating the clock phases and observe that their duty cycle can be slightly less or greater than 25% depending on the circuit topology. We must therefore quantify the effect of this systematic departure upon the MUX performance. Plotted in Fig. 5.11 are the width and height of the transmitter’s output eye as a function of the duty cycle. Here, the middle eye of PAM4 is examined. We note that (1) the width in fact prefers a duty cycle of about 23%,⁵ and (2) the height is less sensitive, prefers about 28%, and can tolerate from about 22% to 33%. Figures 5.12(a) and (b) depict simulated examples, indicating that erring toward smaller values is more tolerable because the eye in the former exhibits a greater opening. The simulations leading to Fig. 5.12 include the direct 4-to-1 MUX and the output driver (with inductive peaking) with a clock transition time of 15 ps. These simulations can be repeated with a channel model and other imperfections to determine the optimum duty cycle.

As mentioned in Section 5.3.2, the MUX of Fig. 5.10(b) draws transient kickback currents from its inputs. The kickback arises when one tail device turns on and its current must initially flow from the C_{GS} of the corresponding differential pair transistors. For the MSB path, the resulting gate current has a peak of 260 μA and lasts about 20 ps. The PMOS differential pairs in the charge-steering selector alleviate the issue as shown in Fig. 5.13. For the MSB path, the tail capacitance in Fig. 5.6(a) is doubled to ensure sufficient voltage swings at X and

⁵Since the turn-off and turn-on delays of the direct 4-to-1 MUX tails are not equal, the neighboring branches briefly overlap in time for a duty cycle of 25%.

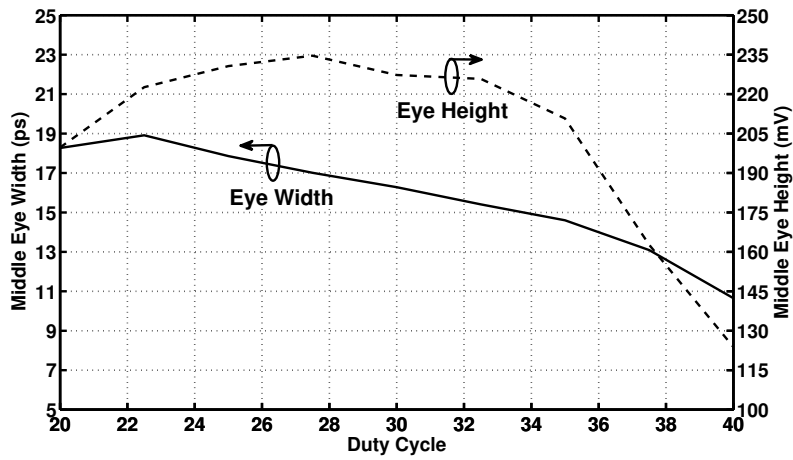


Figure 5.11. Dependence of height and width of PAM4 middle eye upon duty cycle.

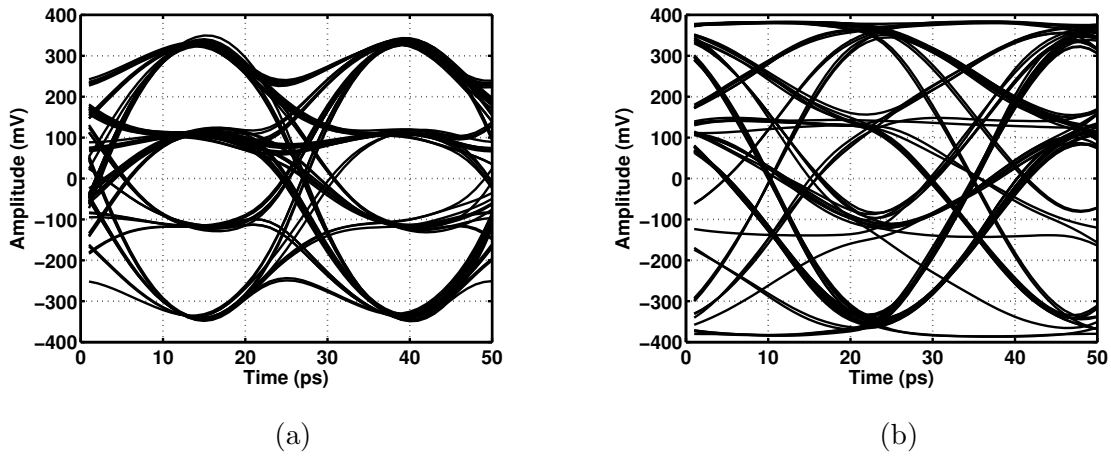


Figure 5.12. (a) Output eye for 20% duty cycle, and (c) output eye for 37.5% duty cycle.

Y , and the precharge switches are widened by a factor of 2 to guarantee proper reset.

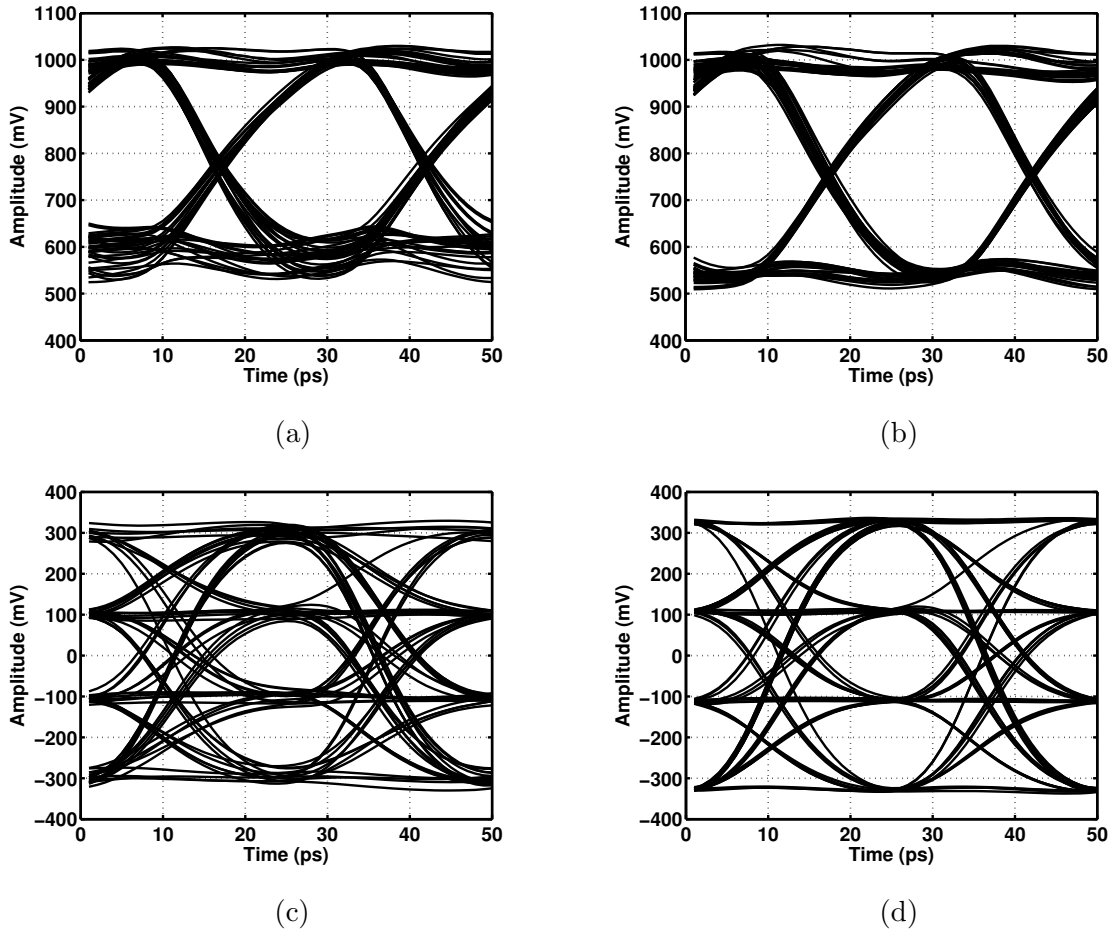


Figure 5.13. Direct 4-to-1 MUX single-ended output eye-diagrams (a) without and (b) with PMOS differential pairs, and transmitter differential output eye-diagrams (c) without and (d) with PMOS differential pairs.

5.4 Output Driver/DAC

The 40-Gb/s MSB and LSB data streams are combined in the output driver to produce the final 80-Gb/s PAM4 signal. Based on the analysis in Chapter 2, for the data rate of 80 Gb/s and 45-nm technology, we prefer a CML topology.

Figure 5.14 shows the realization, where three nominally identical differential pairs act as a 2-bit DAC. The 300-pH inductors broaden the bandwidth in the

presence of the output pad capacitance (≈ 50 fF), which includes the pad and two small ESD diodes.⁶ The overall circuit consumes 13 mW from a 1-V supply.

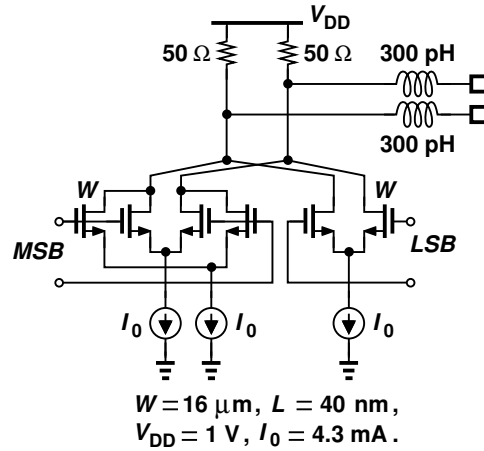


Figure 5.14. Topology of the PAM4 CML output driver/DAC.

The use of short-channel devices raises concern regarding the nonlinearity of the DAC: since the output resistance varies with the digital input, the output eye can be distorted. The effect is exacerbated by the fact that the input high level is close to V_{DD} , forcing the transistors into the triode region for some output PAM4 levels.

In Chapter 1, we explained that uniform eye heights are desired in a PAM4 eye diagram. In Chapter 2, we see that the output swing trades off with the headroom of the current source in a CML output driver. Therefore, we should study the nonlinearity of the CML output driver, a two-bit current-steering DAC essentially.

We start the analysis with a general case of $N + 1$ different output levels. Shown in Fig. 5.15 is a single-ended current-steering DAC with input code k ,

⁶Series peaking in this case simplifies the layout as the inductors become part of the routing to the pads.

$0 \leq k \leq N$. Each current source is modeled by an ideal current source I_0 in parallel with its Norton equivalent impedance r_o . The termination, R_T , is equal to channel characteristic impedance, R_L .

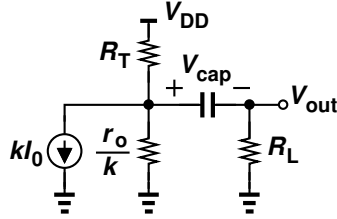


Figure 5.15. Equivalent circuit of single-ended current-steering DAC.

With the input code of k , the output level is given by

$$V_{out} = \frac{V_{DD} - V_{cap}(1 + R_T/(r_o/k)) - kI_0R_T}{2 + R_T/(r_o/k)}, \quad (5.1)$$

where V_{cap} is the voltage across the AC coupling capacitor:

$$V_{cap} = \frac{1}{2} \left(\frac{V_{DD}}{R_T} - NI_0 \right) \left(R_T \parallel \frac{r_o}{N} \right). \quad (5.2)$$

Figure 5.16(a) depicts the output of a seven-bit DAC with $V_{DD} = 1$ V, $R_T = R_L = 50 \Omega$, $r_o = 12.7$ k Ω , $NI_0 = 12$ mA and $N = 127$, and the ideal linear output. Comparing the two, we plot integral nonlinearity (INL) of the single-ended current-steering DAC in Fig. 5.16(b). The region around the middle output level exhibits most severe nonlinearity of about 7 LSB.

Similarly, we also write the output of a differential current-steering DAC [Fig. 5.17] as

$$\begin{aligned} V_{out} &= V_{out+} - V_{out-} \\ &= \frac{(I_0r_o - V_{DD})r_oR_T(N - 2k)}{2r_o^2 + 1.5Nr_oR_T + k(N - k)R_T^2}. \end{aligned} \quad (5.3)$$

Figure 5.18 shows the comparison between the nonlinear output and the ideal

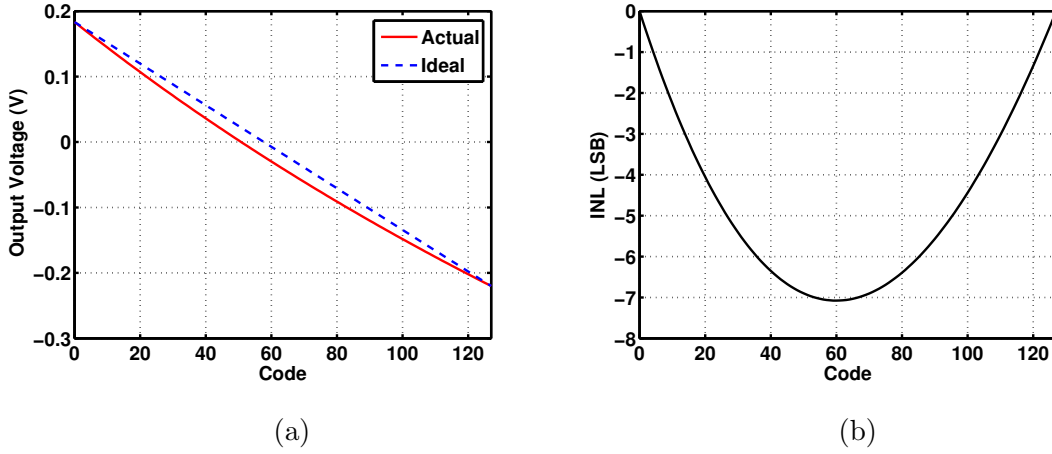


Figure 5.16. (a) Output of 7-bit single-ended current steering DAC, and (b) INL.

output, and INL under the same condition as the single-ended DAC. Different from that of the single-ended topology, INL of the differential current-steering DAC exhibits zero INL at the middle of the code. In addition, since INL is forced to zero at the middle, the maximum INL is less than 0.6 LSB, more than ten times smaller than that of the single-ended topology.

For a two-bit current-steering DAC that delivers PAM4 output, Fig. 5.19 plots its INL of the single-ended and the differential topologies with N set to be 3, and the same NI_0 and r_o/N as the seven-bit case, i.e., $I_0 = 4$ mA and $r_o = 300$ Ω . It can be proved that the INL of the two-bit DAC is

$$INL = \frac{R_T^2}{2r_o^2 + 1.5r_oR_T + 2R_T^2} \text{ (LSB)}, \quad (5.4)$$

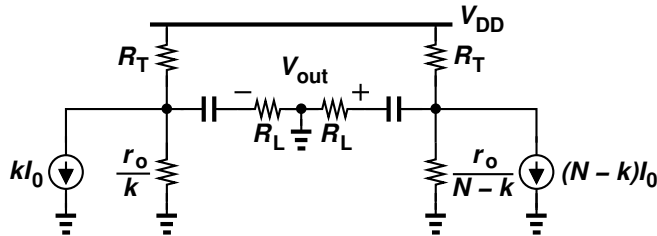


Figure 5.17. Equivalent circuit of differential current-steering DAC.

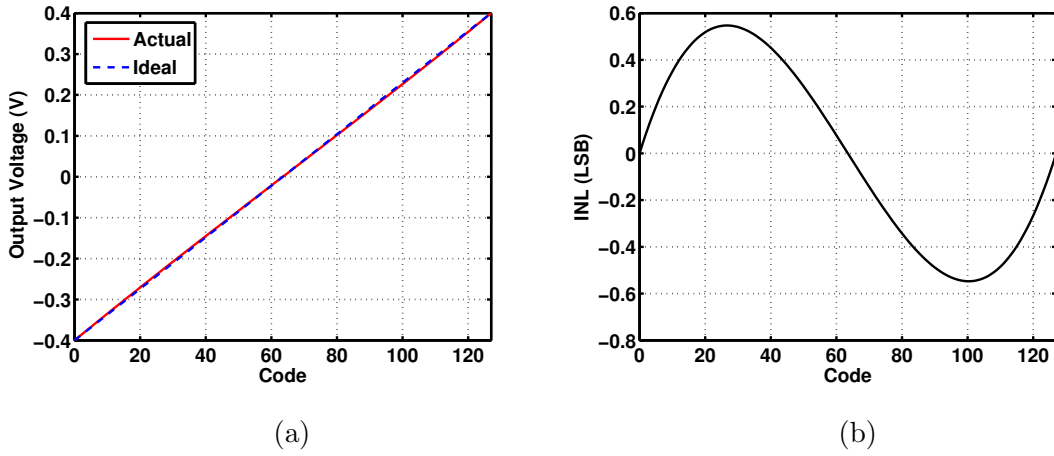


Figure 5.18. (a) Output of 7-bit differential current steering DAC, and (b) INL.

which yields 0.1 LSB in this case.

For the singled-ended driver, the INL yields a RLM of 86% and for the differential driver, 98%, necessitating the use of the differential output driver in a PAM4 transmitter, and proving enough linearity of the two-bit differential topology.

Although the differential PAM4 CML output driver exhibits enough high RLM, the standard [4] also requires the fluctuation amplitude of the output CM

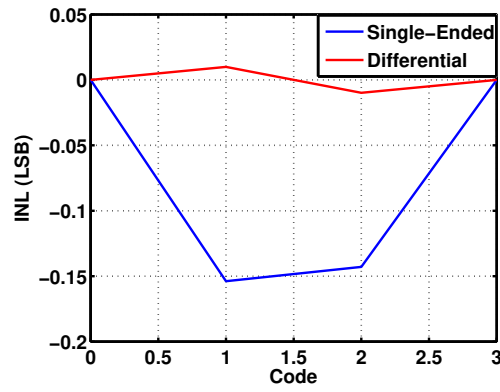


Figure 5.19. Comparison of INL between single-ended and differential PAM4 CML output driver.

level to be less than about 30 mV, which means we must still leave enough headroom for the current source to ensure the large enough Norton equivalent impedance and small enough CM fluctuation.

5.5 Clock Generation

As explained in Section 5.3, the transmitter in Fig. 5.1 extensively exploits quadrature and 45° clock phases with 25% or 50% duty cycles to perform serialization without the use of latches. The generation and distribution of these phases thus play a central role in the overall performance and power consumption.

The most critical clock phases are those running at 10 GHz with a duty cycle of 25% because their mismatches directly translate to jitter at the output of the 4-to-1 MUX. To create these phases, we can (1) directly generate 10-GHz overlapping quadrature clocks by means of two coupled LC oscillators and use AND gates to convert the duty cycle to 25%, (2) generate a 20-GHz differential clock, apply it to a standard $\div 2$ circuit and AND the results, or (3) generate a 20-GHz differential clock and apply it to a $\div 2$ circuit that inherently produces outputs with a 25% duty cycle. From Fig. 5.11, we target an optimal duty cycle of around $25\% \pm 3\%$. The first approach is less attractive as quadrature LC VCOs suffer from a high phase noise and require at least two symmetric inductors, complicating the floor plan. The second method demands that CMOS static AND gates operate at 10 GHz, a difficult and power-hungry task. The third solution is potentially the most efficient since it avoids the logic altogether.

We begin with the divider topology illustrated in Fig. 5.20(a) [26], whose outputs have a duty cycle of approximately 25%. While achieving a high speed, this structure faces two drawbacks: (1) the logical low levels at the output are de-

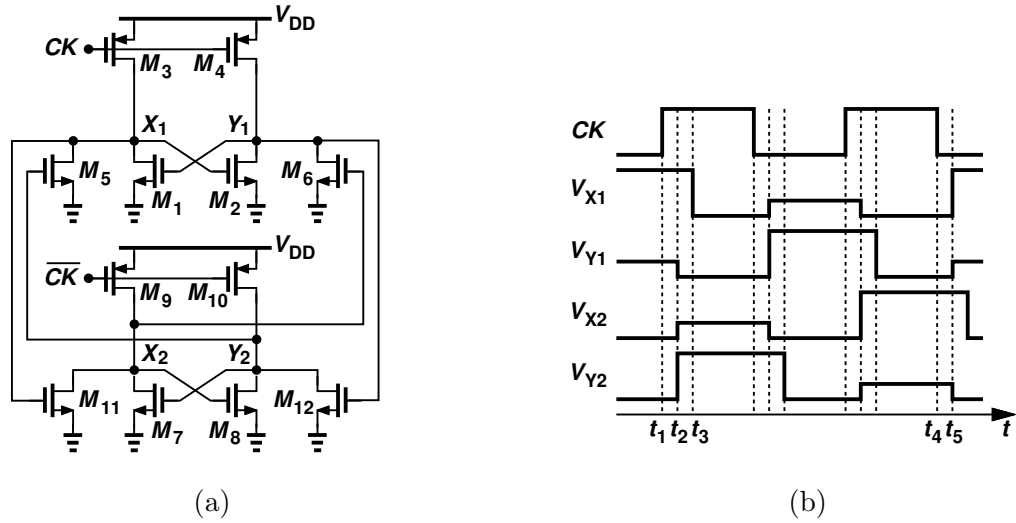


Figure 5.20. (a) Divider topology to generate 25%-duty-cycle clocks directly [26], and (b) divider's waveforms.

graded for about one quarter of the time, and (2) the duty cycle is in fact greater than 25% by one gate delay. To understand the cause of these issues, we examine the circuit's operation with the aid of the waveforms shown in Fig. 5.20(b). Suppose CK is low, V_{X1} is high, and the other three outputs are low. At $t = t_1$, CK rises and \overline{CK} falls, turning on M_{10} and pulling V_{Y2} to V_{DD} at $t = t_2$ (while M_{12} is off). Since V_{X1} is still high, M_{11} is on, but M_9 has also turned on. Thus, the low level in V_{X2} degrades and a static current flows. Now, the rising edge at Y_2 drives M_5 and brings V_{X1} down at $t = t_3$. That is, the high-to-low transition at V_{X1} occurs two gate delays after the rising edge on CK . The operation proceeds in a similar manner until $t = t_4$, when CK falls, causing V_{X1} to rise at $t = t_5$. In summary, V_{X1} incurs one gate delay on its falling edge and two on its rising edge, exhibiting a duty cycle of 25% plus one gate delay, a significant error at 10 GHz.

In order to eliminate the static current, a cross-coupled pair can be inserted in series with the drains of the clocked transistors [27], but, owing to the greater

gate delay, the duty cycle error increases further. As an alternative approach, let us consider the static latch topology shown in Fig. 5.21(a), where M_a and M_b are controlled by the inputs. If, for example, CK falls when D_{in} is high, M_5 does not fight M_3 anymore. Nevertheless, the duty cycle still remains well above the desired value. To address this issue, we recognize in Fig. 5.20(b) that any rising edge on CK can be allowed to pull V_{X1} to zero. In other words, CK can directly lower V_{X1} rather than through V_{Y2} . This observation leads us to add two clocked devices, M_c and M_d , as shown in Fig. 5.21(b) such that they can respectively force V_{X1} or V_{Y1} to zero when CK goes high. Proper ratioing of $W_{5,6}$ and $W_{c,d}$ yields the desired duty cycle.

The series combination of PMOS devices in Fig. 5.21(b) degrades the divider's speed significantly. We then change all of the transistors to their opposite type,

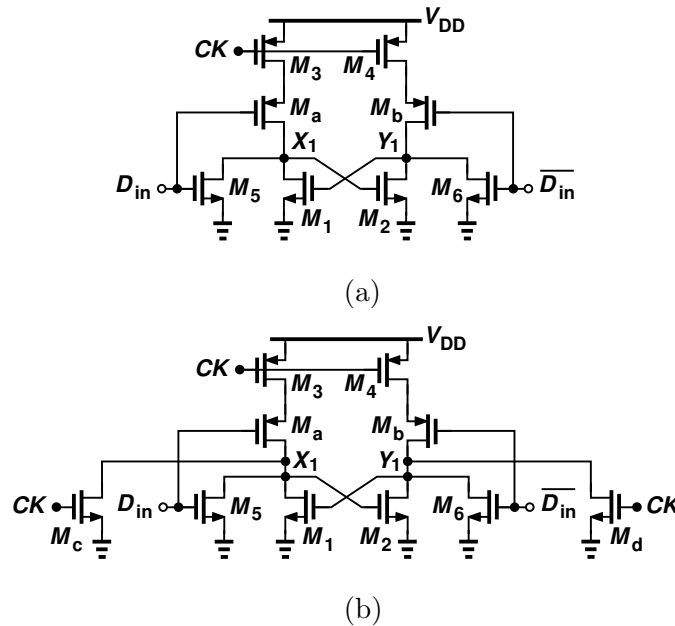
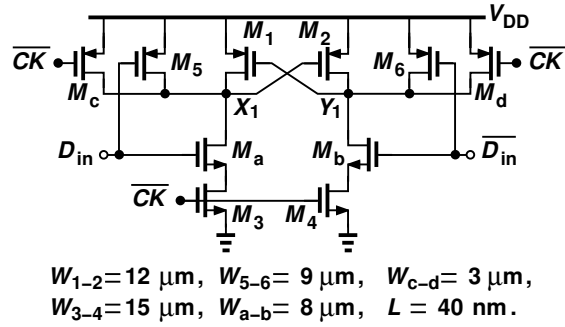
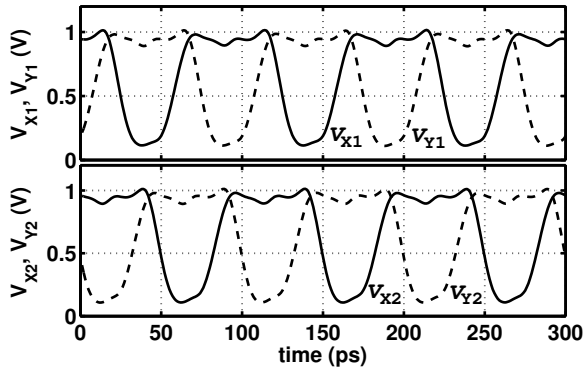


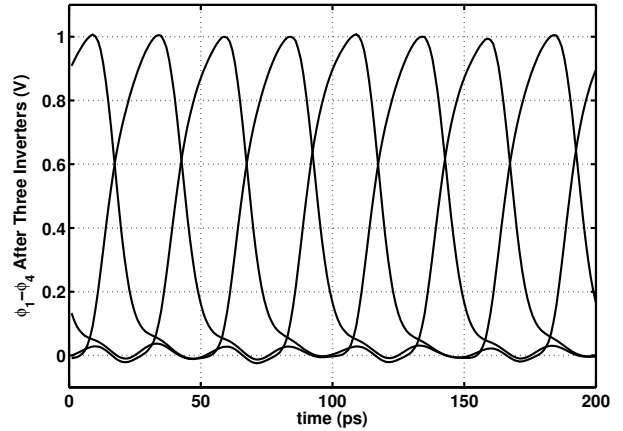
Figure 5.21. (a) Latch topology to remove static current of M_a and M_b , and (b) M_c and M_d driven by CK to reduce transition delay of falling edge on V_{X1} and V_{Y1} .



(a)



(b)



(c)

Figure 5.22. (a) Proposed latch topology with stacked NMOS devices, (b) simulated waveforms of the divider outputs, and (c) the outputs after three inverters.

arriving at the proposed latch design depicted in Fig. 5.22(a)⁷ and the simulated waveforms in Fig. 5.22(b) and (c).

According to simulations, the topology of Fig. 5.21(b) reaches a maximum speed of 23 GHz and that in Fig. 5.22(a), 29 GHz. The divider is followed by three inverters to deliver the four phases to the charge-steering MUX and the direct 4-to-1 MUX in Fig. 5.1. The divider core consumes 3.7 mW at an output

⁷The ratios chosen here lead to a duty cycle range of 24% to 32% across SS, SF, FS, FF, and TT corners.

frequency of 10 GHz, and the buffer inverters, 8.2 mW. Figure 5.23 depicts the relation between the output duty cycle after three inverters and the transistor size ratio. The curve exhibits a linear tuning property due to the “delay interpolation” between $M_{5,6}$ and $M_{c,d}$.

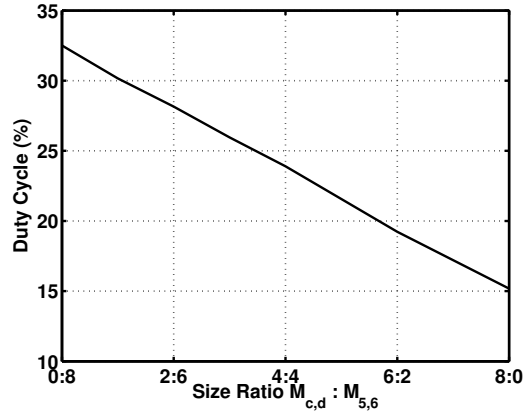


Figure 5.23. Relation of duty cycle upon transistor ratio.

As mentioned in Section 5.2, with no retimer after the 4-to-1 MUX, the mismatches between the clock phases produce jitter. Monte Carlo simulations of the divider, its buffers, the four charge-steering 2-to-1 MUXes, the direct 4-to-1 MUX and the output driver/DAC indicate an average jitter of $205 \text{ fs}_{\text{rms}}$ and one-sigma of $75 \text{ fs}_{\text{rms}}$ due to mismatches. We also observe in Section 5.8 that the measured transmitter output jitter in the 40-Gb/s NRZ mode is only $479 \text{ fs}_{\text{rms}}$ and the measured DCD is $100 \text{ fs}_{\text{rms}}$, concluding that the matching is acceptable.

The second divide-by-2 stage in Fig. 5.1 runs at an input frequency of 10 GHz but, with only 25%-duty-cycle phases available from the preceding divider, it must operate with a clock high level that lasts less than 25 ps. Moreover, the circuit must provide eight output phases, SEL_j and $\overline{SEL_j}$ for $j = 1, \dots, 4$. For this purpose, we introduce another new divider topology that exploits all four 10-GHz phases. Shown in Fig. 5.24(a), the circuit incorporates four latches that

are consecutively driven by ϕ_1 - ϕ_4 , thereby shifting two ONEs and two ZEROs by 25 ps every time ϕ_j pulsates. Figure. 5.24(b) depicts the C²MOS latch used here, with the cross-coupled inverters guaranteeing differential operation. The overall circuit draws 1.9 mW at an input frequency of 10 GHz.

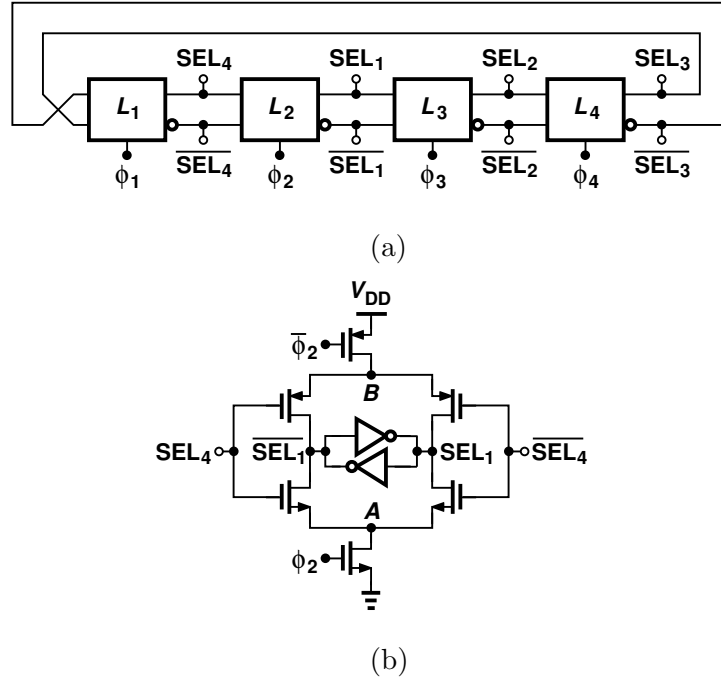


Figure 5.24. (a) Divide-by-2 stage to generate eight-phase clocks, and (b) C²MOS latch used in the divider.

5.6 PLL Design

In most high-speed wireline transmitters, the phase-locked loop and the clock distribution network draw considerable power. In this work, the PLL generates a 20-GHz output that is subsequently divided to produce the phases and frequencies necessary for serialization. With $UI = 25$ ps, we target an overall PLL jitter of $300 f_{s_{\text{rms}}}$ for negligible degradation of the transmitted data.

The PLL jitter arises from the reference spurs, the VCO phase noise, and the multiplied reference phase noise. The closed-loop bandwidth, f_{BW} , must therefore be optimized in terms of these three imperfections.

To quantify the (deterministic) jitter due to the reference spurs, we write $V_0 \cos(\omega_c t + \beta \sin \omega_m t) \approx V_0 \cos \omega_c t - \beta V_0 \sin \omega_c t \sin \omega_m t$ and note that the normalized spur level, $\beta/2$, is also half of the peak of the jitter and must be multiplied by $\sqrt{2}$ to yield the rms value. Thus, a spur level of K dBc translated to roughly $\beta/\sqrt{2} = \sqrt{2} \times 10^{K/20}$ radians of rms jitter. For example, if the spurs are at -50 dBc, the jitter is around $36 \text{ fs}_{\text{rms}}$, and hence negligible. We also note that a crystal oscillator phase noise, S_{REF} , of about -150 dBc/Hz at 312.5 MHz rises by $20 \log 64 = 36 \text{ dB}$ within the loop bandwidth as it reaches the output. Thus, f_{BW} must be chosen so as to minimize the sum of $64 S_{REF} f_{BW}$ and the shaped VCO phase noise. This PLL design chooses $f_{BW} = 20 \text{ MHz}$.

In order to achieve a wide bandwidth with acceptable spur levels, we modify the RF synthesizer architecture introduced in [29] for operation with $f_{REF} = 312.5 \text{ MHz}$ and $f_{VCO} = 20 \text{ GHz}$. Shown in Fig. 5.25(a), the loop consists of an XOR phase detector (PD), a master-slave sampling filter (MSSF), a VCO, and a divider chain. As described in [29], the master-slave sampling action yields a small ripple on the control line and hence low spurs at the output. The settling behavior shown in Fig. 5.25(b) exhibits a ripple of $15 \text{ mV}_{\text{pp}}$ after settling.

In this work, we exploit an LC VCO with complementary cross-coupled transistors [Fig. 5.26(a)] for a nearly rail-to-rail output swing. In a manner similar to [23] and [28], the VCO drives the first stage of the divider chain directly without buffers, omitting the large amount of power burnt on buffers. Owing to a PLL closed-loop bandwidth of 20 MHz , the phase noise requirement for the LC VCO is greatly relaxed, allowing the oscillator power to be as low as 3.5 mW . The VCO

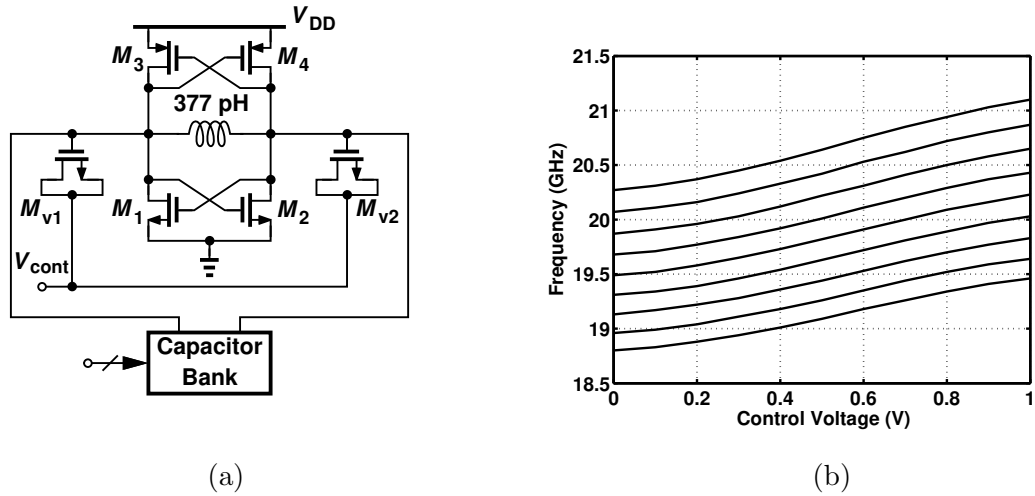


Figure 5.26. (a) VCO implementation, and (b) simulated frequency tuning.

5.7 Floor Plan

The transmitter floor plan must deal with two general issues: (1) the layout style for the MSB and LSB serializer so as to achieve a compact design and hence minimal interconnect capacitances, and (2) the placement of the spiral inductors so as to reduce their mutual coupling and yet maintain short interconnects.

As explained in Section 5.3, the CMOS serializer in Fig. 5.1 exploits reverse scaling by a factor of 2 from one rank to that preceding it. That is, the transistor widths are halved while the number of 2-to-1 selectors is doubled. This design approach also leads to a modular layout: if one rank employs $2M$ selectors, the rank following it can have the same number except that every two adjacent selectors are placed in parallel to perform $2\times$ scaling. As illustrated in Fig. 5.27(a), the first rank consists of 64 unit selectors (with smallest transistor dimensions) that multiplex 128 inputs to 64. The next rank can also have 64 unit selectors, but with each two grouped to create $2\times$ scaling for greater driving strength. Consequently, all of the ranks in the CMOS serializer can use the same layout

structure and benefit from pitch matching.

The actual floor plan is made more compact by rearranging each rank to form an array. Depicted in Fig. 5.27(b), the first rank places every eight selectors in one row, and the second rank merges each two of these selectors. The overall floor plan now has a shorter height, presenting less capacitance to the clock phases arriving from the frequency dividers.

The transmitter incorporates inductors in the MSB and LSB 4-to-1 multiplexers [Fig. 5.10(b)], in the output driver [Fig. 5.14], and in the VCO. The coupling

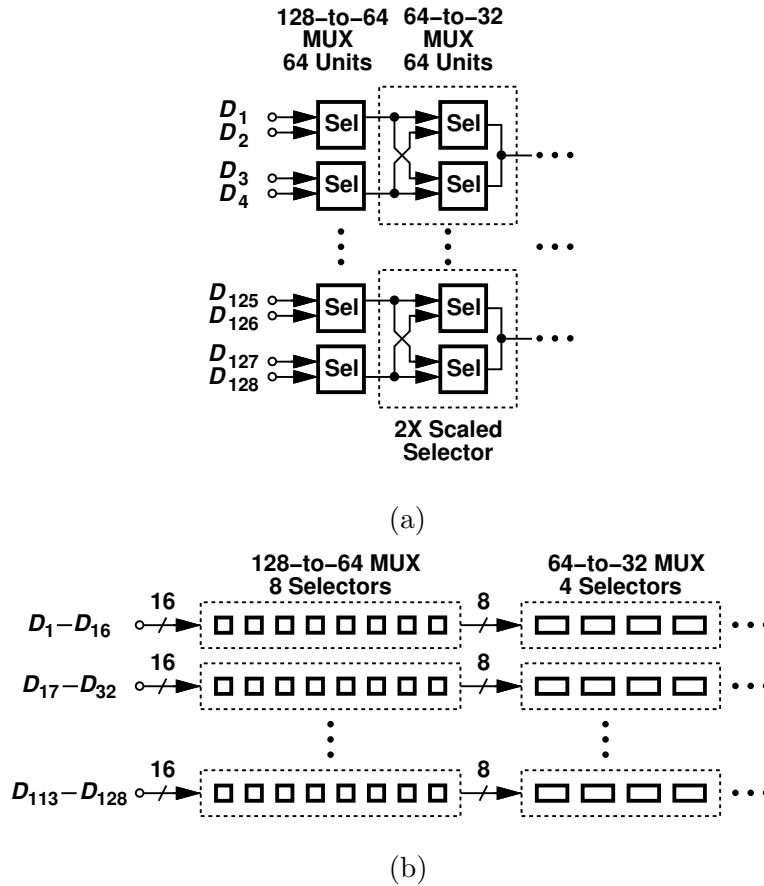


Figure 5.27. (a) Modular-based scaling between the first and the second ranks, and (b) modular placement in CMOS MUX array.

of random data from the inductors in the first two circuits to that in the VCO can produce considerable jitter. The floor plan favors the VCO performance and is shown in Fig. 5.28. Realized as compact, stacked structures, the MSB and LSB MUX inductors are placed on top, with inevitably long interconnects, a minor issue as its Q is less critical. The output series peaking inductors are also positioned at about $150\ \mu\text{m}$ from the core. According to HFSS simulations, the coupling factor between L_{VCO} and L_{S1} is about 0.14%. As explained in Section 5.6, the measured clock phase noise agrees well with the PLL simulations excluding this coupling, suggesting that the VCO is negligibly corrupted by the random data.

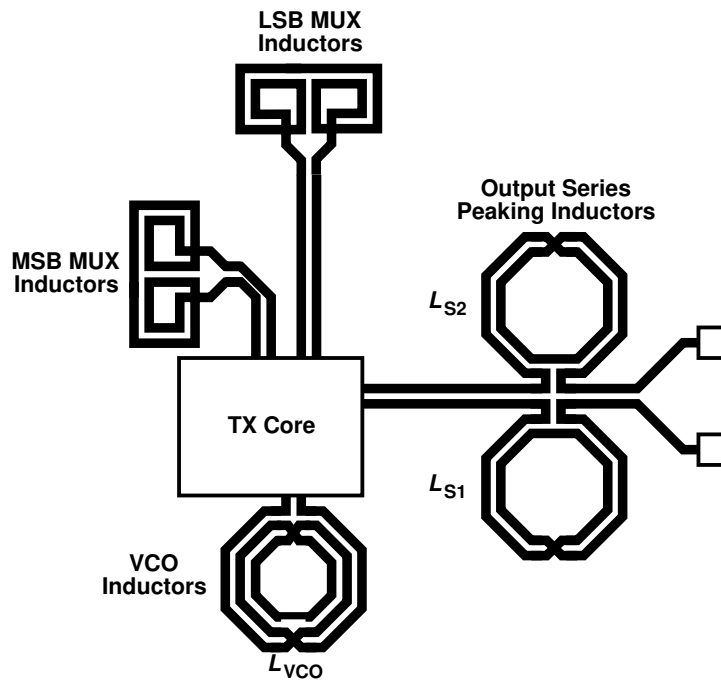


Figure 5.28. Placement of inductors in layout floor plan.

5.8 Experimental Results

The PAM4 transmitter has been fabricated in TSMC's 45-nm digital CMOS technology. Figure 5.29 shows a photograph of the die, whose active area is about $330\ \mu\text{m} \times 320\ \mu\text{m}$. The die has been directly mounted on a printed-circuit board and tested on a high-speed probe station. All of the measurements have been performed with a 1-V supply.

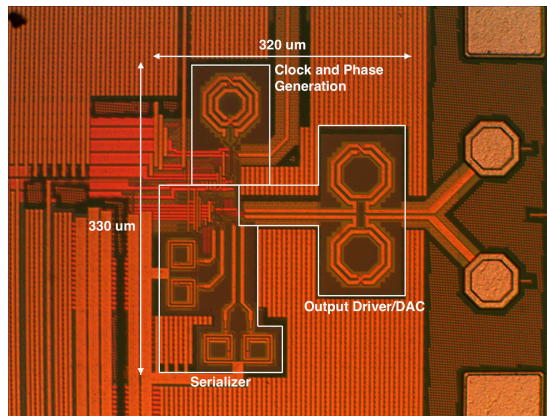


Figure 5.29. Die photograph.

The overall transmitter consumes 44 mW. Table 5.1 shows the breakdown of the power consumption at 80 Gb/s. To separate the power of the clock distribution from the PLL, we simulate the divider chain in two cases: (1) while it drives the data path, and (2) while it does not. The difference between the power values, 4.1 mW, is that necessary for clock distribution.

Figure 5.30 shows the measured TX output in the NRZ mode at 40 Gb/s. Figure 5.31 shows the output in the PAM4 mode at 40 Gb/s and 80 Gb/s. The differential voltage swing is $630\ \text{mV}_{\text{pp}}$. The use of a 1-V supply for the entire system limits the output swing to about 630 mV. If the output driver supply is raised to 1.2 V and the tail currents in Fig. 5.14 to 24 mA, the swing can reach 1.2 V. The data pattern is PRBS7. The vertical eye opening is 170 mV, the

Table 5.1. POWER BREAKDOWN.

Blocks		Power (mW)
Data Path (MSB + LSB)	Output Driver/DAC	13.72
	CML MUX	5.66
	Charge-steering MUX	1.61
	CMOS MUX	0.73
Clock Path	Divider Chain and Buffers	18.25
	XOR + MSSF + Nonoverlap Gen.	0.62
	VCO	3.46
Total		44.05

horizontal eye opening is 0.56 UI for the middle eye and 0.43 UI for the top and bottom eyes. Shown in Fig. 5.32, the output bit pattern has been captured and checked against the input data to verify correct serialization.

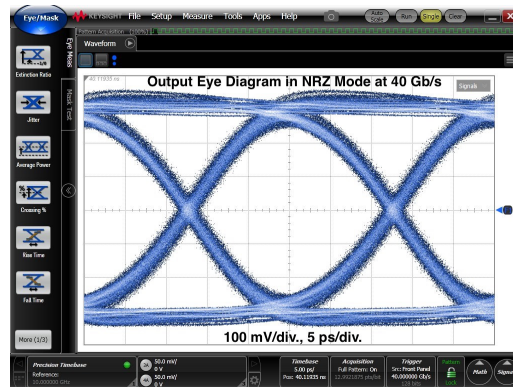
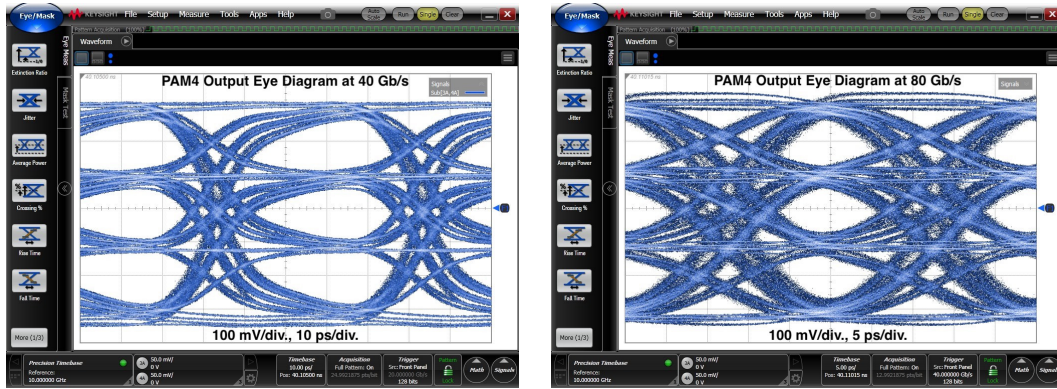


Figure 5.30. Output eye diagram in NRZ mode at 40 Gb/s.

The linearity of the PAM4 waveform is quantified by RLM (Chapter 1). To measure the RLM, the input data pattern is chosen so that the output PAM4 waveform contains different symbols with each lasting for 16 UIs as shown in Fig. 5.33. Our measured RLM is around 99%, exceeding the 92% specification [4].

The 20-GHz clock generated by the PLL has also been characterized. The



(a)

(b)

Figure 5.31. PAM4 output eye diagrams at (a) 40 Gb/s, and (b) 80 Gb/s.

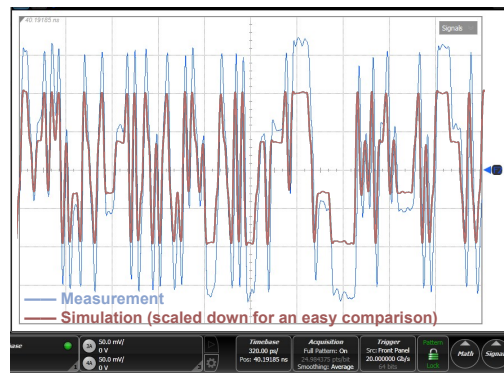


Figure 5.32. Comparison between simulated and measured waveform.

measured spectrum and phase noise profile are shown in Fig. 5.34. The reference spurs are at -45 dBc. Figure 5.34(b) plots the measured phase noise of the 10-GHz clock. Due to our equipment limitation, the maximum offset is 1 GHz, but we note from Fig. 5.34(c) that the integrated jitter reaches a plateau of 200 fs beyond approximately 200 MHz. In fact, noting that the phase noise is around -140 dBc/Hz for offsets greater than 200 MHz, we observe that the range from 1 GHz to 5 GHz (the Nyquist rate) contributes $[\sqrt{4 \text{ GHz} \times 10^{-14}}/2\pi] \times 100 \text{ ps} \approx 100 \text{ fs}$, which, combined with the 205-fs value found in Fig. 5.34(b), amounts to 228 fs. That is, the phase noise beyond 1 GHz is negligible. This is also verified by

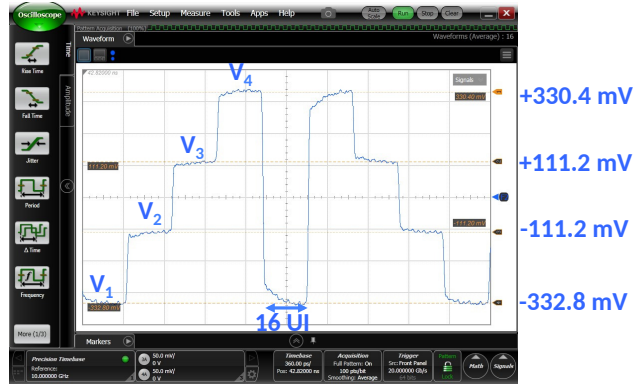
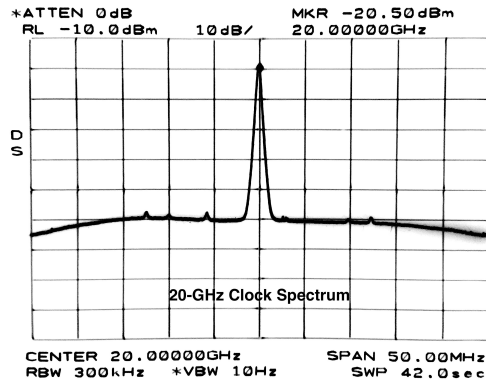


Figure 5.33. RLM test sequence.

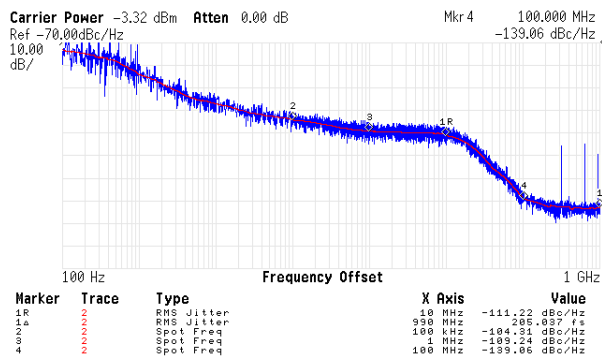
simulation of the data path, including the output driver, and observing a flat phase noise up to 10 GHz.

To examine the effect of mismatches in ϕ_1 - ϕ_4 , we apply the input data so as to create a 20-GHz periodic 0101 NRZ sequence at the TX output. Shown in Fig. 5.35, a spur level of -41 dBc at 10-GHz offset in the single-ended output indicates a deterministic jitter of 100 fs_{rms} jitter due to mismatches among ϕ_1 - ϕ_4 and within the 4-to-1 MUX. The relation between the spur and the jitter is obtained in Section 5.6.

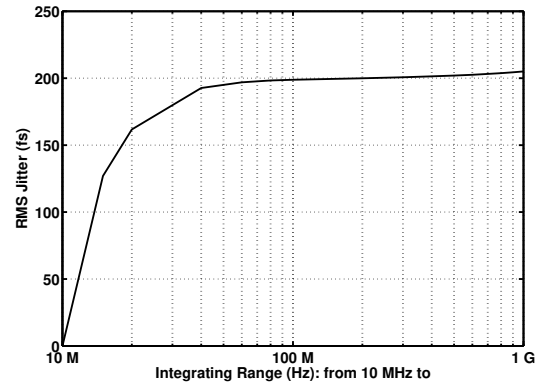
Table 5.2 compares our measured performance with that of the prior art. We note that, if the PLL power consumption is excluded, our design achieves a nearly six-fold improvement in power efficiency. Even if we prorate the power consumption of our output DAC from 13.7 mW to about 32 mW to account for the larger output swing of 1.2 V_{pp,d} in [19], our power efficiency is still higher by approximately a factor of 4 (excluding the PLL). Even though our prototype does not include FFE, the discussion in Section 5.3.2 shows that adding FFE would entail negligible power penalty.



(a)



(b)



(c)

Figure 5.34. (a) Spectrum of 20-GHz clock, phase noise profile of 20-GHz clock divided by two externally, and (c) relation of jitter and integrating range of 20-GHz clock divided by two externally..

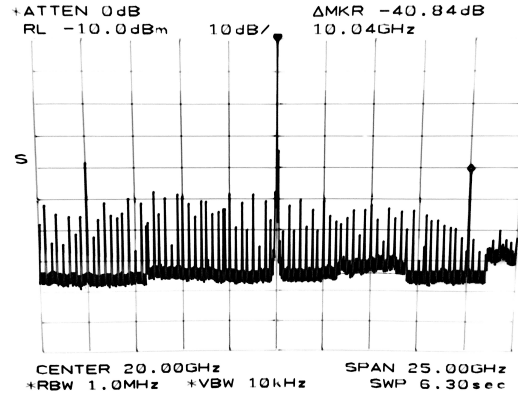


Figure 5.35. Measured spectrum of single-ended output delivering 20-GHz 0101 NRZ sequence.

Table 5.2. PERFORMANCE SUMMARY.

	Peng ISSCC'17	Steffan ISSCC'17	Dickson ISSCC'17	This Work
Technology (nm)	40	28	14	45
Data Rate (Gb/s)	56	64	56	80
Output Driver Type	CML	CML	SST	CML
Driver Supply (V)	1.5	1.2	0.95	1
Max. Output $V_{pp,d}$ (mV)	600	1200	900	630
RLM	N/A	0.94	N/A	0.99
RMS Jitter (fs)	688	290	318	205
Integ. Range (MHz)	0.0001 - 1000	0.5 - 8000	N/A	10 - 1000
Power (mW)	Exc.*	200	145***	101
	Inc.**	220	-	-
Power Eff. (pJ/bit)	Exc.**	3.57	2.26***	1.8
	Inc.**	3.93	-	-
Active Area (mm²)	0.8*	N/A	0.035*	0.1

* Excluding PLL power but including clock distribution.

** Including PLL power and clock distribution.

*** Without I&Q clock generation.

CHAPTER 6

Conclusion

This research studies architectures and circuit techniques that reduce power consumption of wireline transmitters operating at tens of gigabit per second. Theoretical analysis and practical design issues of PAM4 signaling have been described in details.

A NRZ transmitter with 2-tap FFE has been described in Chapter 4. It has shown that the MUX output without being followed by retimers carries small enough jitter such that the power-hungry retimers in transmitter front end have been safely removed. The functions of the final MUX and the output driver have been also merged. A current-integrating MUX has been introduced to drive 96-fF load with only 0.4 mW. The timing scheme along with the current-integrating MUX makes 2-tap FFE free of latches. These techniques afford NRZ operation at 40 Gb/s with 7.4-dB boosting under 32 mW, achieving a 2.28 improvement on power efficiency compared to state of the art works.

A PAM4 transmitter running at 80 Gb/s with nearly six-fold improvement on power efficiency than the prior art has been described in Chapter 5. It has been recognized that the combination of different logic styles makes it feasible to balance speed-power tradeoff. The MUX cell has been simplified to be latchless structure, totally saving 720 latches. A charge-steering MUX with improved immunity to kickback noise has been introduced to handle 10-Gb/s multiplexing. The direct multi-phase MUX has avoided high-speed latches before it. It has been

proved that the proposed frequency divider with a 25% output duty cycle draws only 1/4 power of the conventional AND-gate method. This PAM4 transmitter delivers 80-Gb/s data with 630-mV_{pp} swing under 1-V supply, and consumes 44 mW including a type-I PLL using MSSF. The output PAM4 levels exhibits a RLM of 99%.

For future work, the timing scheme of the PAM4 transmitter is ready to accommodate 2-tap FFE function without any latch. The correction of mismatches in duty cycles and propagation delays of 25%-duty-cycle clocks is worth of study in order to promote their use in quarter-rate “retimerless” transmitter front end at even higher data rate. The concept of the multi-phase MUX can be extended to even larger multiplexing ratio, in which broadband techniques and circuits need to be studied furthermore.

REFERENCES

- [1] C. Menolfi *et al.*, “A 25Gb/s PAM4 Transmitter in 90nm CMOS SOI,” *IEEE ISSCC Dig. Tech. Papers*, pp. 72-73, Feb. 2005.
- [2] 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, Apr. 2005.
- [3] B. Garlepp *et al.*, “A 1-10 Gbps PAM2, PAM4, PAM2 partial response receiver analog front end with dynamic sampler swapping capability for backplane serial communications,” *Symposium on VLSI Circuits Dig. of Tech. Papers*, pp. 376 -379, June 2005.
- [4] <http://www.ieee802.org/3/bs/>.
- [5] H. Taub and D. L. Schilling, *Principles of Communication Systems*, Second Edition, McGraw-Hill, 1986.
- [6] J. Kim *et al.*, “A 16-to-40Gb/s Quarter-Rate NRZ/PAM4 Dual-Mode Transmitter in 14nm CMOS,” *ISSCC Dig. Tech. papers*, pp. 60-61, Feb. 2015.
- [7] M. Bassi, F. Radice, M. Bruccoleri, S. Erba, and A. Mazzanti, “A High-Swing 45 Gb/s Hybrid Voltage and Current-Mode PAM-4 Transmitter in 28 nm CMOS FDSOI,” *IEEE J. Solid-State Circuits*, vol. 51, no. 11, pp. 2702-2715, Nov. 2016.
- [8] T. O. Dickson *et al.*, “A 1.8pJ/b 56Gb/s PAM-4 Transmitter with Fractionally Spaced FFE in 14nm CMOS,” *IEEE ISSCC Dig. Tech. Papers*, pp. 118-119, Feb. 2017.
- [9] P. Staric and E. Margan, *Wideband Amplifiers*, Springer, 2015.
- [10] B. Razavi, *Design of Integrated Circuits for Optical Communications*, McGraw-Hill, 2003.
- [11] 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.
- [12] A. A. Hafez, M.-S. Chen and Ch.-K. K. Yang, “A 32-48 Gb/s Serializing Transmitter Using Multiphase Serialization in 65 nm CMOS Technology,” *IEEE J. Solid-State Circuits*, vol. 50, pp. 763-775, Mar. 2015.

- [13] R. Navid *et al.*, “A 40 Gb/s Serial Link Transceiver in 28 nm CMOS Technology,” *IEEE J. Solid-State Circuits*, vol. 50, pp. 814-827, Apr. 2015.
- [14] K. Huang *et al.*, “A 190mW 40Gbps SerDes Transmitter and Receiver Chipset in 65nm CMOS Technology,” *Proc. CICC*, Sep. 2015.
- [15] Ch.-K. K. Yang, R. Farjad-Rad and M. A. Horowitz, “A 0.5- μ m CMOS 4.0-Gbit/s Serial Link Transceiver with Data Recovery Using Oversampling,” *IEEE J. Solid-State Circuits*, vol. 33, pp. 713-722, May 1998.
- [16] Y. Chang, A. Manian, L. Kong, and B. Razavi, “A 32-mW 40-Gb/s CMOS NRZ Transmitter,” accepted by *Proc. CICC*, Apr. 2018.
- [17] Y. Frans *et al.*, “A 56-Gb/s PAM4 Wireline Transceiver Using a 32-Way Time-Interleaved SAR ADC in 16-nm FinFET,” *IEEE J. Solid-State Circuits*, vol. 52, no. 4, pp. 1101-1110, Apr. 2017.
- [18] P.-J. Peng, J.-F. Li, L.-Y. Chen, and J. Lee, “A 56Gb/s PAM-4/NRZ Transceiver in 40nm CMOS,” *IEEE ISSCC Dig. Tech. Papers*, pp. 110-111, Feb. 2017.
- [19] G. Steffan *et al.*, “A 64Gb/s PAM-4 Transmitter with 4-Tap FFE and 2.26pJ/b Energy Efficiency in 28nm FDSOI,” *IEEE ISSCC Dig. Tech. Papers*, pp. 116-117, Feb. 2017.
- [20] K. Gopalakrishnan *et al.*, “A 40/50/100Gb/s PAM-4 Ethernet Transceiver in 28nm CMOS,” *IEEE ISSCC Dig. Tech. Papers*, pp. 62-63, Feb. 2016.
- [21] A. Nazemi *et al.*, “A 36Gb/s PAM4 Transmitter Using an 8b 18GS/s DAC in 28nm CMOS,” *IEEE ISSCC Dig. Tech. Papers*, pp. 58-59, Feb. 2015.
- [22] J. Lee, P.-C. Chiang, and C.-C. Weng, “56Gb/s PAM4 and NRZ SerDes Transceivers in 40nm CMOS,” *Symposium on VLSI Circuits Dig. of Tech. Papers*, pp. 118 -119, June 2015.
- [23] J. W. Jung and B. Razavi, “A 25-Gb/s 5-mW CMOS CDR/Deserialzer,” *IEEE J. Solid-State Circuits*, vol. 48, no. 3, pp. 684-697, Mar. 2013.
- [24] Y. Lu, K. Jung, Y. Hidaka, and E. Alon, “Design and Analysis of Energy-Efficient Reconfigurable Pre-Emphasis Voltage-Mode Transmitters,” *IEEE J. Solid-State Circuits*, vol. 48, no. 8, pp. 1898-1909, Aug. 2013.
- [25] C.-K. K. Yang, R. Farjad-Rad, and M. A. Horowitz, “A 0.5- m CMOS 4.0-Gbit/s Serial Link Transceiver with Data Recovery Using Oversampling,” *IEEE J. Solid-State Circuits*, vol. 33, no. 5, pp. 713-722, May 1998.

- [26] B. Razavi, K. F. Lee, and R. H. Yan, "Design of High-Speed, Low-Power Frequency Dividers and Phase-Locked Loops in Deep Submicron CMOS," *IEEE J. Solid-State Circuits*, vol. 30, no. 2, pp. 101-109, Feb. 1995.
- [27] I. Fabiano, M. Sosio, A. Liscidini, and R. Castello, "SAW-less analog front-end receivers for TDD and FDD," *IEEE ISSCC Dig. Tech. Papers*, pp. 82-83, Feb. 2013.
- [28] A. Manian and B. Razavi, "A 40-Gb/s 14-mW CMOS Wireline Receiver," *IEEE J. Solid-State Circuits*, vol. 52, no. 9, pp. 2407-2421, Sep. 2017.
- [29] L. Kong and B. Razavi, "A 2.4 GHz 4 mW Integer-N Inductorless RF Synthesizer," *IEEE J. Solid-State Circuits*, vol. 51, no. 3, pp. 626-635, Mar. 2016.
- [30] P. Andreani and A. Fard, "More on the $1/f^2$ Phase Noise Performance of CMOS Differential-Pair LC-Tank Oscillators," *IEEE J. Solid-State Circuits*, vol. 41, no. 12, pp. 2703-2712, Dec. 2006.
- [31] Y. Chang, A. Manian, L. Kong, and B. Razavi, "An 80-Gb/s 44-mW Wireline PAM4 Transmitter," accepted by *IEEE J. Solid-State Circuits*, 2018.
- [32] A. Manian, *Low-Power Techniques for CMOS Wireline Receivers*, UCLA Ph.D. Thesis, 2016.
- [33] 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, Sep. 2007.