

SYSTEM-DRIVEN CIRCUIT DESIGN FOR
ADC-BASED WIRELINE DATA LINKS

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

Kevin Zheng

August 2018

© 2018 by Kevin Jie Zheng. All Rights Reserved.
Re-distributed by Stanford University under license with the author.



This work is licensed under a Creative Commons Attribution-
Noncommercial 3.0 United States License.
<http://creativecommons.org/licenses/by-nc/3.0/us/>

This dissertation is online at: <http://purl.stanford.edu/hw458fp0168>

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

Boris Murmann, Primary Adviser

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

Amin Arbabian

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

Mark Horowitz

Approved for the Stanford University Committee on Graduate Studies.

Patricia J. Gumpert, Vice Provost for Graduate Education

This signature page was generated electronically upon submission of this dissertation in electronic format. An original signed hard copy of the signature page is on file in University Archives.

Abstract

In the era of connectivity, wireline I/O has been a key technology underpinning the aggressive performance improvements of computer and communication systems. All standards, ranging from electrical to optical, long haul to short reach interconnects, have increased their aggregate bandwidth at a rate of about 2x every 4 years. This trend drives the usage of multi-level modulation schemes, such as PAM4. As a result, ADC-DSP based links are gaining more attention and are now heavily investigated. ADC links also take advantage of the intrinsic bandwidth and area improvements from process technology scaling. It is becoming very difficult for conventional mixed-signal links to meet the bandwidth and performance requirements of next-gen systems. As equalization is moved into the digital domain for ADC-based links, the ADC needs to be very fast, have reasonably high resolution and yet be very power efficient. Consequently, these stringent requirements pose significant challenges in both ADC links' architecture and circuit design.

In this thesis, we focus on a statistical framework for understanding ADC nonidealities, including quantization and nonlinearity, and their impact in a link context. We will then present studies on equalization locations along the link's signal path, motivating the need of pre-equalization before the ADC. With the importance as well as the implementation challenges of pre-equalizers in mind, we present the first inverter-based analog front-end (AFE) for 56Gb/s transceivers for both PAM2 and PAM4

applications. Such AFEs include continuous time linear equalizers (CTLE), programmable gain amplifiers (PGA) and the ADC's track and hold circuits (T/H). The inverter-based AFE is process scaling friendly, consumes smaller power and achieves significant area reduction over prior art. Finally, system identification (SID) methodologies are used with silicon measurements to show validity of the before-mentioned analysis. SID-based bit error rate (BER) estimation method is also presented to compare measured and predicted link performance for different ADC resolutions.

Acknowledgments

*This dissertation is dedicated to the memory of my beloved grandfather, **Bing Gui Zheng (1926-2018)**. Even though you passed away just before I could finish this thesis, every bit of your lifelong influence on me is reflected in this work and I wish to make you proud.*

Throughout my Ph.D. career, I was so fortunate to have the support, encouragements and inspirations from countless mentors, colleagues, friends and family. I never felt alone or helpless during my pursuit of a doctoral degree. It was a once-in-a-lifetime journey full of discoveries and excitements in addition to the unexpected turns and sleepless nights.

First and foremost, I would like to express my utmost gratitude to my advisor, Professor Boris Murmann. My achievements would not be possible without his teaching, mentoring, and guidance. I also thank him for being a humorous and optimistic figure who constantly provides the positive energy any student needs during the up and downs of a Ph.D. I feel privileged to have conducted my research in his group, and look forward to any future opportunities to work together again. Thank you, Boris.

I would like to thank Professor Mark Horowitz for co-advising my research and making me part of the SerDes family tree. My gratitude also goes to Professor Amin Arbabian for being on my reading committee and giving valuable advice both on research and life as a Ph.D. student. I thank Professor Juan Rivas for serving in my

defense committee and also Professor Gordon Wetzstein for chairing my defense.

All circles back to my undergraduate years in M.I.T. I thank my dear friends, especially Jorge Simosa and Krishna Settaluri, for providing the escapes I needed outside of school. I thank my Masters advisor Professor Vladimir Stojanović for introducing me to high speed research, broadening my perspectives and helping me enroll in Stanford University.

Internships and industry collaborations have been a highlight for my graduate career. No words can begin to describe the sheer amount of wisdom that my industry mentors have shared with me. I am thankful to all my former mentors and colleagues from Analog Devices, Broadcom, Texas Instruments and Futurewei for the great projects and learning experiences. I want to thank Stefanos Sidiropoulos from Cadence especially for his constant support and engaging conversations that sparked many ideas. Lastly, my special thanks goes to Ken Chang, Yohan Frans, Geoff Zhang and many others from Xilinx who provided me with the resources, leadership and freedom to succeed in my graduate program.

My research group members, past and present, have made my Ph.D. years educational, diverse and fun. Alex Guo, Bill Chen, Jon Spaulding, Doug Adams, Martin Kraemer, Man-Chia Chen, Ryan Boesch and Vaibhav Tripathi were sources of knowledge, patience and relief when I faced obstacles. Thanks to Danny Bankman, Lita Yang, Nikolaus Hammler, Daniel Villamizar, Sean Fischer, Stephen Weinreich, Nishit Shah, Po-Hsuan Wei, Alex Omid-Zohoor, Chenxin Zhu and Danta Muratore, I have gained so many once-in-a-lifetime memories when we worked, commiserated, travelled, ate or played soccer together.

I thank my in-laws who helped me push through the last phase of my Ph.D. I thank my family in China who have nurtured me and given me the best opportunities to excel, especially my grandparents for their early education, love and indulgence when I was growing up.

To my brother Richard, you never cease to impress me on how much you have learned and grown. I am so proud of your accomplishments so far. In many ways, I look up to you for different aspects in school, family and life. Thank you for being the best brother I can wish for and I am excited about both of our futures.

To my wife Lin, thanks for giving me the reasons to be happy, the strength to keep working, and the courage to move forward. My graduate years have been as much of a challenging journey for you, and thanks for the understanding and patience that I perhaps did not deserve when I was absent. Thanks for being my best friend and partner, and I could not wait to continue our adventurous journey ahead with you.

To my parents, I am honored and blessed to be your son. All of these started with your hard work, sacrifice and unconditional love. To my dad, you have always believed in me and taught me how to be a good person, husband and son, just like you. To my mom, you have always been the anchor that kept our family united and allowed me to grow in the most fertile environment. I owe you the world for everything that you have done, from providing the necessities in life to the important lessons that shaped who I am today. Thank you.

Contents

Abstract	iv
Acknowledgments	vi
1 Introduction	1
1.1 Background and motivation	1
1.2 Organization	6
2 Statistical Framework for ADC Nonidealities	8
2.1 ADC nonidealities and performance metrics	8
2.1.1 Ideal ADC and quantization error	8
2.1.2 ADC nonidealities and ENOB	11
2.1.3 ENOB's statistical implications	14
2.2 Statistical framework for ADC quantization	17
2.2.1 Quantization noise and PDF	17
2.2.2 BER estimation with uniform PDF	21
2.3 Statistical framework for other ADC errors	25
2.3.1 DNL	25
2.3.2 Nonlinearity/INL	29
3 Comparative Study on Equalizer Position	36

3.1	TX vs. RX FFE	37
3.1.1	Fundamental SNR analysis	37
3.1.2	Practical concerns for TX and RX FFE	43
3.2	Equalization before vs. after ADC	45
3.2.1	Design equation for ADC resolution	45
3.2.2	Channel PMR, FFE and ADC resolution	48
4	Inverter-based Pre-ADC Equalizers	53
4.1	Inverter as analog elements	54
4.1.1	Inverter transconductor and diode load	54
4.1.2	Inverter active inductor	56
4.1.3	Inverter linearity	61
4.1.4	Inverter biasing voltage	63
4.1.5	Inverter cell layout	64
4.2	Inverter-based CTLE for PAM2 application	66
4.2.1	Additive two-path CTLE	67
4.2.2	Replica ring oscillator based biasing	68
4.2.3	Transceiver testbed and measurements	70
4.3	Inverter-based AFE for PAM4 application	75
4.3.1	CTLE topologies	76
4.3.2	Transceiver and receiver AFE architecture	77
4.3.3	Inverter-based hybrid CTLE and ADC T/H	79
4.3.4	AFE and transceiver measurements	82
5	System Identification of ADC-based Links	88
5.1	System identification	89
5.1.1	Working principles	89
5.1.2	SID data collection and pre-processing	90

5.1.3	Example SID outputs	91
5.1.4	SID with FFE	93
5.2	BER estimation using SID	95
5.2.1	Residual ISI PDF	95
5.2.2	Conditional PDFs for circuit errors	96
5.2.3	BER estimation method and results	97
5.2.4	BER prediction for lower resolution ADCs	99
5.3	Pre-equalization and ADC resolution trade-off	101
6	Conclusions and Future Work	104
6.1	Summary and Conclusions	104
6.2	Future Work	105
A	Pseudo-Independent Quantization Noise Proof	107
B	Nonlinearity PDF	111
C	ADC Resolution Requirement	112
D	Inverter-based Active Inductor Impedance	114
E	Inverter-based Active Inductor Buffer	116
F	Inverter-based Active Inductor Buffer Noise	117
G	Inverter-based Buffer DC Characteristics	120
G.0.1	Square law device	121
G.0.2	Velocity saturated device	122

List of Tables

3.1	Qualitatively comparison for TX vs. RX FFE practical concerns	43
3.2	Example ADC resolutions for different link scenarios	51
4.1	Comparison table for inverter-based additive CTLE	74
4.2	Additive and subtractive CTLE required g_m comparison	77
4.3	Comparison table for PAM4 transceivers	84

List of Figures

1.1	Block diagram for a simple link model	2
1.2	Examples of (a) an open eye diagram for PAM2, (b) a closed eye diagram for PAM2 and (c) an open eye diagram for PAM4	3
1.3	Bit error rate in wireline links	4
1.4	Equalization in conventional mixed-signal links	5
1.5	Block diagrams for (a) conventional mixed-signal link, (b) conventional ADC-based link, and (c) proposed ADC-based link with pre-equalization	6
2.1	(a) Ideal ADC model, (b) example 3b ADC static transfer function and (c) example 3b ADC quantization error function	9
2.2	FFT of example ideal 8b ADC	11
2.3	(a) ADC model with nonidealities added and (b) equivalent model for ADC with nonidealities	12
2.4	FFT of example nonideal 8b ADC	13
2.5	Simplified ADC model with noise and distortion	14
2.6	Gaussian distribution and BER	16
2.7	BER vs. SNR under Gaussian assumption for PAM2	16
2.8	Discrete time model of an ADC-based link	18
2.9	Quantization error PDFs	20
2.10	(a) ENOB model and (b) statistical model for quantization	21

2.11	Convolution of thermal noise and quantization PDFs	22
2.12	BER estimation parameters illustrated in an eye diagram	24
2.13	Comparison of ENOB model vs. statistical BER estimation curves . .	24
2.14	(a) Ideal quantization error function and (b) quantization error func- tion with DNL	26
2.15	Statistical model incorporating DNL noise	26
2.16	Two ideal ADC approximations with different LSB sizes to estimate DNL effects	27
2.17	BER estimation curves comparison between $Z(x)$ method and two ap- proximation methods	28
2.18	Statistical model for nonlinearity	30
2.19	Example PDF of PAM4 receiver input signal	31
2.20	(a) Nonlinearity error dependency on input PDF and (b) example non- linearity error PDFs from input PDF in equation 2.28 and $c = 0.1$. .	32
2.21	Gain compression on data levels (blue/red dashed lines to solid levels) due to nonlinearity and adjusted decision levels (black dashed lines) .	33
2.22	BER curves comparing ENOB and statistical nonlinearity model for different c and σ_{ISI}	34
2.23	Full scale clipping and input PDF folding	35
3.1	Equalization positions in an ADC-based link	36
3.2	Discrete time models for links with (a) TX side FFE and gain block modeling peak power constraint and (b) RX side FFE with a gain control block	37
3.3	TX FFE operation principle	38
3.4	RX FFE operation principle	39
3.5	Three channels used for FFE location study	41

3.6	Total SNR vs. RX input noise for three channels and three FFE length settings	41
3.7	Sampled eye diagram of link 2 RX output using 10 pre-cursor and 20 post-cursor FFE	42
3.8	TX+RX FFE including nonlinearity and circuit noise	44
3.9	RX FFE model with ADC quantization noise	46
3.10	ADC resolution requirement components	47
3.11	(a) Normalized channels with 2x difference in PMR and (b) corresponding transient waveforms for PAM2 signaling [25]	49
3.12	Quantization noise PDF's propagation through an FFE	50
3.13	Each block's role in an ADC-based link with pre-equalization	52
4.1	(a) CML-based CTLE and (b) inverter-based filter	54
4.2	(a)Inverter transconductor (b)Inverter resistive load	55
4.3	Various inverter configurations for different analog blocks	56
4.4	Inverter active inductor	57
4.5	Unity gain buffer with (a)diode connected load and (b)active inductor load	58
4.6	Unity gain buffer noise models with (a)diode connected load and (b)active inductor load	59
4.7	Unity gain buffer large signal bias point	62
4.8	Example layout diagrams for unit inverter cells	65
4.9	Example layout styles for abutted inverter cells	65
4.10	Short-reach link application block diagram with channel responses . .	66
4.11	(a) Single-ended CTLE schematic and (b) simulated frequency responses	67
4.12	(a) Simulation of parameter α vs. V_{DD} and (b) simulated inverter f_u at different process and temperature corners	70

4.13	PAM2 short reach receiver testbed block diagram	71
4.14	Bathtub curve comparison between CML-based and inverter-based CTLEs under nominal conditions	72
4.15	Bathtub curve comparison between overwrite and auto LDO modes for different temperatures	73
4.16	Bathtub curve comparison between CML-based and inverter-based CTLEs under extreme temperatures	73
4.17	(a) Eye widths vs. temperature and (b) regulated ground voltage vs. temperature in different LDO modes	74
4.18	Chip photos	75
4.19	Pros and cons of additive and subtractive CTLEs	76
4.20	PAM4 ADC-based transceiver architecture	78
4.21	PAM4 ADC-based receiver AFE block diagram	79
4.22	(a) Half circuit schematic for hybrid CTLE and (b) schematic for offset calibration DAC with common mode feedback	80
4.23	Simulated two-stage CTLE frequency response for (a) mid LF code and different HF codes, and (b) mid HF code and different LF codes	81
4.24	Schematics of PGA and ADC interface circuits	82
4.25	(a) ADC output scans and post DSP equalization eye scans of VSR and LR channels without crosstalk and (b) bathtub curves for LR channel under different crosstalk levels.	83
4.26	PAM4 ADC-based transceiver die photo	84
4.27	Transceiver BER performance for all TX settings	85
4.28	Transceiver BER for different voltage and temperature corners	86
4.29	CTLE/PGA codes for different voltage and temperature corners	86
5.1	Block diagram for SID engine	89

5.2	Example snapshot of PRBS13 ADC output stored in memory	91
5.3	(a) Estimated channel and (b) estimated error correlation with estimated linear output	92
5.4	SID pulse responses at $\pm 8\%$ UI phase offsets	93
5.5	Block diagram for SID engine with FFE	94
5.6	(a) Estimated channel and (b) estimated error correlation with estimated linear output with FFE	94
5.7	Numerical calculation of residual ISI PDF by convolution	96
5.8	Clipped ADC SID outputs. (a) Error scatter plot and (b) conditional PDFs from error histograms	97
5.9	BER estimation flow from SID results	98
5.10	BER estimation comparisons with measured BER with 7b data . . .	99
5.11	Truncated ADC output quantization PDF propagation through FFE	100
5.12	BER estimation comparisons with measured BER with 6b and 5b data	101
5.13	BER curves for different ADC resolutions at fixed pre-equalizer settings	102
5.14	(a) BER curves for different ADC resolutions at different pre-equalizer settings and (b) SID estimated channels and PMRs	103
A.1	Model for quantization's pseudo-independence	107
A.2	Quantization as PDF interval sampling	108
D.1	Test voltage to find active inductor impedance	114
F.1	Small signal model for resistor noise	118
G.1	DC currents in an inverter buffer	120

Chapter 1

Introduction

1.1 Background and motivation

As the demand for high data rate continues to rise, driven by evolving applications and larger user base, it has become increasingly difficult to develop power- and area-efficient wireline links needed to support such an infrastructure. Both single lane data rate and density has to increase within the limited printed circuit board (PCB) real estate. Recent trends of various wireline standards show a consistent 2x increase in aggregate bandwidth requirements approximately every three to four years [1]. Current standards such as 400 Gigabit Ethernet (IEEE802.3bs [2]) and OIF Common Electrical I/O 56G (CEI-56G [3]) are pushing the limits of conventional wireline links based on mixed-signal processing with a per-lane bandwidth requirement of >56Gb/s, which motivates the need for new wireline architectures.

Any wireline link over a typical communication channel can be modeled as a system shown in Figure 1.1. The link consists of a transmitter (TX), a channel, and a receiver (RX). The TX starts with a sequence of bits that need to be sent, and typically uses non-return-to-zero (NRZ) pulse amplitude modulation (PAM) schemes to transmit signals in the voltage domain. In this example, PAM2 is used with only

two distinct voltage levels. The channel, which includes PCB traces and connectors, is a low-pass filter in the frequency domain, creating signal loss up to and beyond the Nyquist frequency, which is half of the link's symbol rate. Equivalently, the channel has a corresponding time-domain pulse response, which convolves with the transmitted data. The pulse response of the channel contains inter-symbol interference (ISI), which alters the signal and the receiver is now prone to errors when recovering the received data.

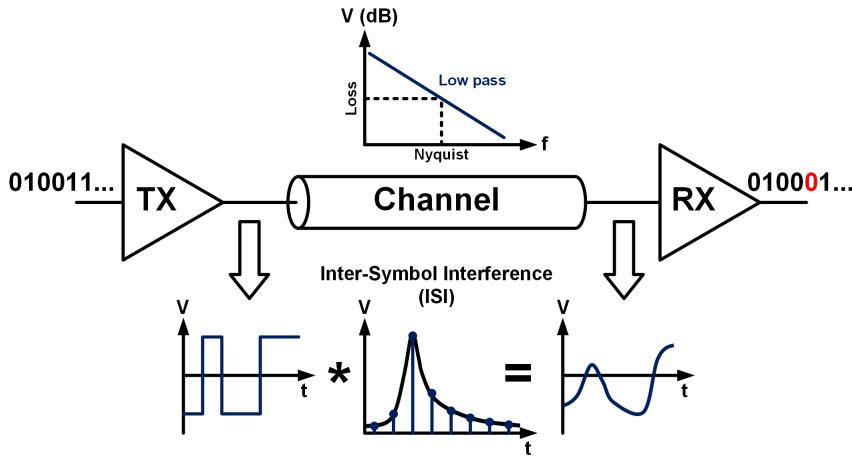


Figure 1.1: Block diagram for a simple link model

Eye diagrams provide a pictorial way to judge whether a channel/link is healthy. Eye diagrams are generated by overlapping unit intervals (UI) of the signal of interest in the time domain. Figure 1.2 shows examples of open and closed eye diagrams. When channel ISI and noise in the link system is much smaller than the actual data signal, we obtain an open eye diagram (Figure 1.2(a)) in which the eye has a vertical opening, eye height (E.H) and a horizontal opening, eye width (E.W). The dither around the data levels are due to residual ISI and noise. Figure 1.2(b) shows an example of a closed eye diagram. In such a system, ISI and noise overwhelm the transmitted data, therefore making the high and low data levels indistinguishable. For a given PAM scheme, there will be PAM-1 eyes in the eye diagram. As an example,

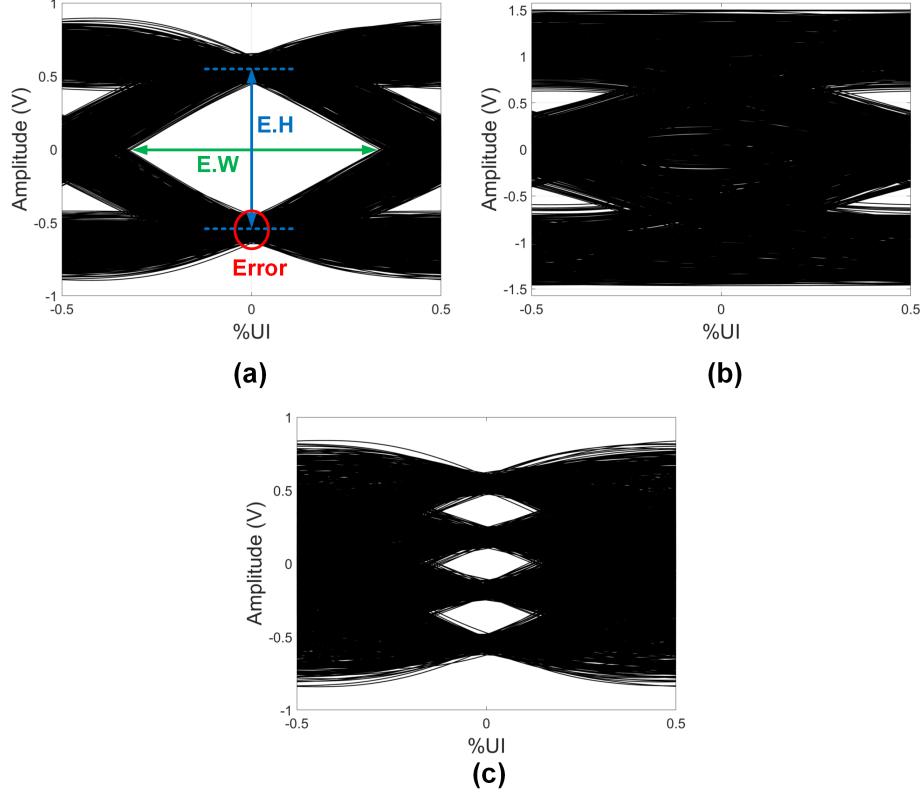


Figure 1.2: Examples of (a) an open eye diagram for PAM2, (b) a closed eye diagram for PAM2 and (c) an open eye diagram for PAM4

Figure 1.2(c) shows an open eye diagram with 3 eyes when PAM4 signaling is used and sent through the same channel as (a). Both eye height and eye width decreased by simply using a higher PAM scheme. To first order, the relationship between eye height and error spread determine the probability that the receiver makes a mistake.

Bit error rate (BER) is the metric that ultimately quantifies a link's performance. As illustrated by the PAM2 example in Figure 1.3, the TX only sends voltages representing either a “0” or “1” in the probability domain. After channel ISI and circuit/environment noise are added, the receiver sees a signal whose probability density function (PDF) is a sum of two Gaussian-like peaks. The area under the curves’ crossed-over portions gives the probability of a wrong bit decision, thus BER.

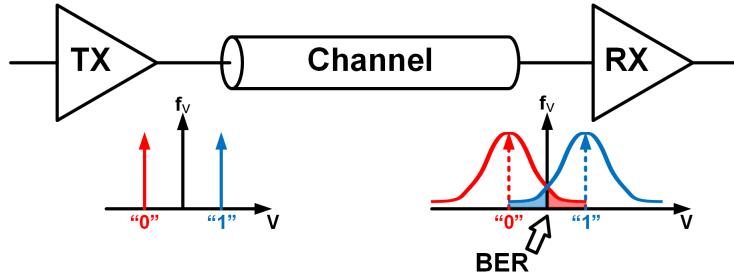


Figure 1.3: Bit error rate in wireline links

In order to compensate for ISI due to channel loss and meet the required BER specifications, different equalization techniques are used in link systems. Figure 1.4 shows conventional mixed-signal equalization blocks on both the TX and RX end [4]. The main cursor in a channel's pulse response is the signal of interest. Any ISI cursors before the main cursor are called pre-cursors. A feed-forward equalizer (FFE), which is typically implemented on the TX side, normally cancels pre-cursors. Any ISI cursors following the main cursor are considered post-cursors, which are often taken care of by the receiver with a continuous-time linear equalizer (CTLE) and decision feedback equalizer (DFE). The CTLE acts as high-pass filter to compensate for the channel's low-pass action. Due to the high-pass nature of CTLEs, any high frequency receiver input noise will be boosted similar to the actual signal. To avoid excessive noise amplification, DFEs pass the noise-less recovered bits through a finite impulse response (FIR) filter that matches the post-cursor portions of the channel to achieve ISI cancellation. The TX FFE can also provide some coarse equalization for post-cursors.

As both the links' speed and area requirements become more stringent, such a mixed-signal architecture, shown in Figure 1.5(a), poses significant implementation challenges. Specifically, the CTLE and the DFE's summing node are the bottlenecks. On the other hand, links based on analog-to-digital converters (ADCs) are more robust due to their digital nature, but such performance comes at the cost of higher

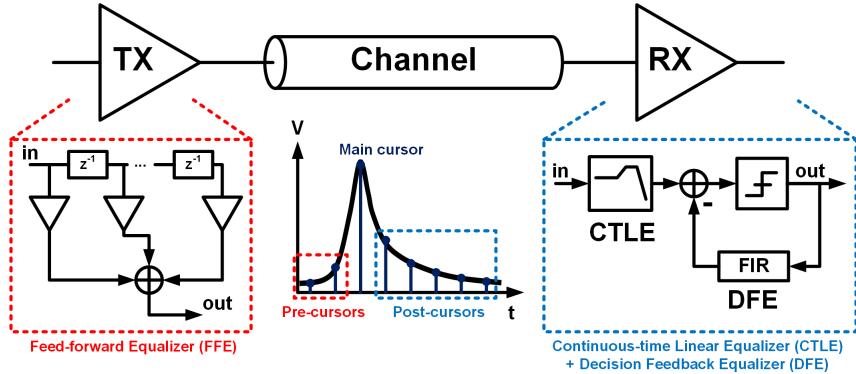


Figure 1.4: Equalization in conventional mixed-signal links

power. As process technology continues to scale, raw BER specification relaxes with forward error correction (FEC), and the use of higher PAM (e.g., PAM4) gains popularity, ADC-based links have gained further traction [5, 6]. Figure 1.5(b) shows a block diagram of a traditional ADC-based link. Here, the ADC is directly connected to the channel output in order to convert the received signal into the digital domain. The digital signal processing (DSP) unit typically includes FFE, DFE and a FEC (with Reed-Solomon codes being currently one of the most popular choices [7]). Despite the flexible, scalable and robust nature of DSP, there are many implementation challenges for both the ADC and the digital circuits at the data rate of interest, demonstrated by first generation works such as [8, 9]. For example, 10pJ/bit is the typical energy efficiency goal for a wireline link, but state-of-the-art ADCs sampling near 50GS/s will consume about 300mW [10], which translates to more than 5pJ/bit, already more than half of the total link energy budget. This motivates a systematic analysis and understanding of ADC-based links and their power and performance tradeoffs from both system and circuit perspectives.

This work explores an architecture in which a pre-equalizer is placed in front of the ADC, as shown in Figure 1.5(c). In essence, a pre-equalizer conditions the input signal first so that the ADC's requirements can be relaxed. However, a thorough

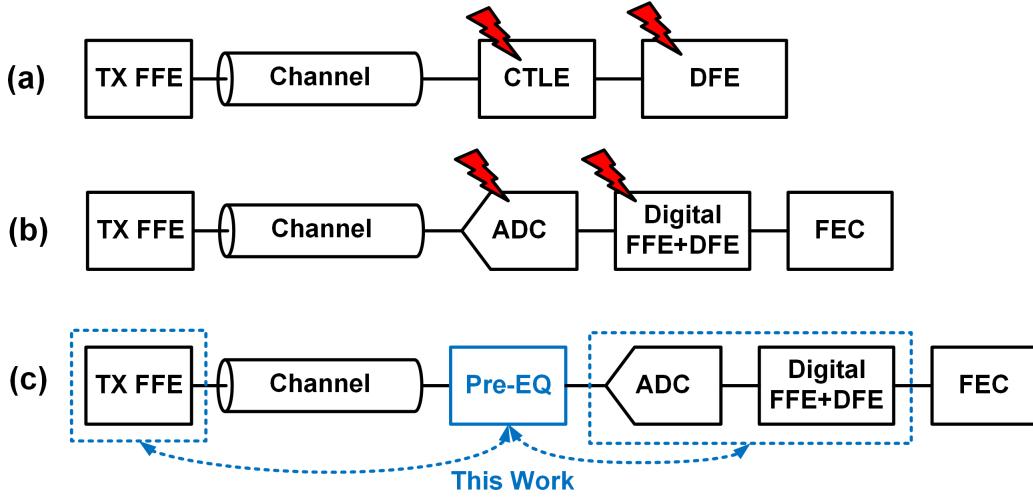


Figure 1.5: Block diagrams for (a) conventional mixed-signal link, (b) conventional ADC-based link, and (c) proposed ADC-based link with pre-equalization

understanding of such tradeoffs is still lacking. Thus, this work first presents a statistical framework for correctly understanding ADC nonidealities in a link context. A further discussion of equalizer position will follow, highlighting the necessity of a pre-equalizer before the ADC. With a better system understanding, custom-designed inverter-based CTLEs and analog front-end (AFE) circuits will be presented within complete transceivers to show their feasibility. Finally, a system identification (SID) methodology is used to further analyze silicon data and conduct experiments to validate the tradeoff between pre-equalization and ADC resolution.

1.2 Organization

The remainder of the thesis is organized as follows:

- Chapter 2 discusses different ADC nonidealities in the statistical domain. Section 2.1 introduces several dominant ADC impairments and conventional performance metrics, such as effective number of bits (ENOB). After understanding

limitations of this oversimplified metric, Section 2.2 uses quantization as an example to show the correct statistical model for thinking about ADC errors and discusses the impact on setting specifications. In Section 2.3, other ADC nonidealities will be investigated under the statistical framework developed.

- Chapter 3 presents a comparative study of different equalizer locations along the signal path and their effects on the overall system and the ADC. Section 3.1 uses a simple model to compare putting the FFE on either the TX or RX side. Section 3.2 expands the model by bringing the ADC into the picture, and derives the relationship between equivalent channel response and the ADC’s resolution requirement. It also focuses on a metric that measures the ADC resolution increase due to channel ISI, therefore leading to the need of pre-equalization.
- Chapter 4 presents two inverter-based AFEs for both PAM2 and PAM4 transceivers. Section 4.1 discusses inverters as analog elements and their layouts as a critical design aspect. Two inverter-based CTLE and AFE are implemented and verified, one for a short-reach PAM2 application in Section 4.2 and one for a flexible-reach PAM4 application in Section 4.3.
- Chapter 5 presents analysis from applying SID on silicon data. Section 5.1 introduces working principles of SID with example results on collected data. Section 5.2 introduces statistical BER estimation from SID and compares estimated BERs with measured BERs. Section 5.3 uses the SID and BER estimation tools to conduct experiments showing the tradeoffs between pre-equalization and ADC resolution.
- Chapter 6 summarizes this thesis and proposes possible future works and research directions.

Chapter 2

Statistical Framework for ADC Nonidealities

2.1 ADC nonidealities and performance metrics

2.1.1 Ideal ADC and quantization error

An ideal ADC uses B bits to represent an analog value that it samples. B is also known as the resolution of the ADC, which in principle translates to 2^B possible levels at the ADC output. As an example, Figure 2.1 shows the static characteristics of an ideal 3b ADC. The ADC adds quantization error, e_q , to the input analog value as shown in Figure 2.1(a). e_q is the output of a quantization error function $q(\cdot)$. Figure 2.1(b) is the static transfer function between analog input and digital output, which has a total of 8 possible output digital levels. Analog values within the full scale range (FSR) of the ADC will be quantized into a corresponding step in the staircase shaped function. The step size (a.k.a. least significant bit (LSB) size) is determined by $\Delta = FSR/2^B$. Analog inputs that exceed $\pm FSR/2$ will be clipped and assigned with the minimum/maximum digital output values. The quantization error in the

digital outputs then can be represented by the sawtooth shaped function in Figure 2.1(c). The quantization error is bounded by $\pm\Delta/2$, and grows indefinitely beyond of ADC's FSR.

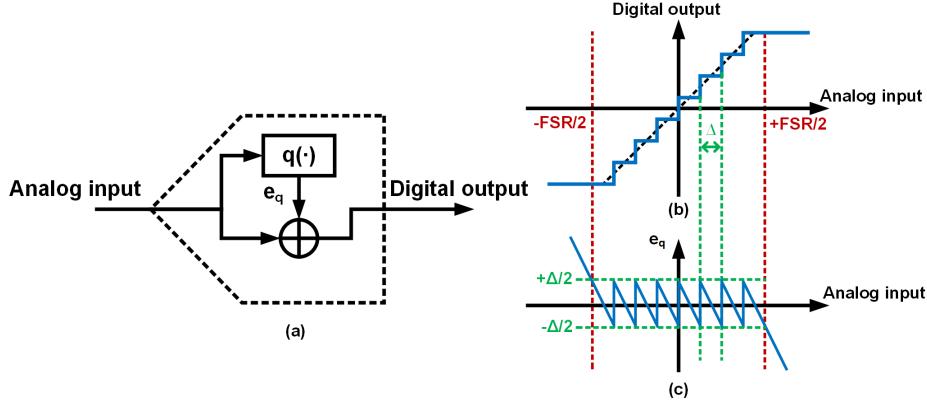


Figure 2.1: (a) Ideal ADC model, (b) example 3b ADC static transfer function and (c) example 3b ADC quantization error function

One way to think about the quantization error's effect on the signal of interest is by looking at the signal to quantization noise ratio (SQNR). In this framework, only the error source's average power is considered, and several assumptions must be made to simplify the underlying model. First of all, the input signal to the ADC needs to cover enough quantization steps. The signal also needs to be active and random enough so that it exercises different parts of the quantization error curve with almost equal chances. The ADC also cannot be clipped beyond its full scale range (at least not often) [11]. In general, these assumptions hold reasonably well in practice, but we will revisit these in a link context in later sections. The quantization noise power, σ_q^2 , can be derived from

$$\sigma_q^2 = \frac{1}{FSR} \int_{-FSR/2}^{FSR/2} q^2(x)dx = \frac{1}{\Delta} \int_0^\Delta q^2(x)dx \quad (2.1)$$

Since the quantization sawtooth function is periodic, the average power can be found

by integrating $q(x) = \frac{\Delta}{2} - x$ over only one period. The result is

$$\begin{aligned}\sigma_q^2 &= \frac{1}{\Delta} \int_0^\Delta \left(\frac{\Delta}{2} - x\right)^2 dx \\ &= \frac{1}{\Delta} \int_0^\Delta \left(\frac{\Delta^2}{4} - \Delta x + x^2\right) dx \\ &= \frac{1}{\Delta} \left(\frac{\Delta^3}{4} - \frac{\Delta^3}{2} + \frac{\Delta^3}{3}\right) \\ \sigma_q^2 &= \frac{\Delta^2}{12}\end{aligned}\tag{2.2}$$

Single tone sinusoidal signals are typically used as an ADC's test input signal since they are well-suited for metrics like SQNR. The sine wave's signal power is well known to be $A^2/2$, with A being the amplitude. If we send a sine wave with amplitude of $FSR/2$ to the ADC, the following SQNR is obtained

$$\text{SQNR} = \frac{\frac{1}{2}(FSR/2)^2}{\Delta^2/12} = \frac{3}{2} \frac{(2^B \Delta)^2}{\Delta^2} = \frac{3}{2} \cdot 2^{2B}\tag{2.3}$$

If SQNR is converted to decibel (dB), it has a linear relationship with ADC's resolution B as shown below

$$\text{SQNR (dB)} = 6.02 \text{dB} \cdot B + 1.76 \text{dB}\tag{2.4}$$

For example, the SQNR for an ideal 3b ADC is 19.8dB and about 50dB for an 8b ADC. A 6.02dB reduction in SQNR means loss of 1 bit in resolution. Another benefit of using sine wave as test signal is to visualize and quantify the ADC performance in the frequency domain. Figure 2.2 shows a Fast Fourier Transform (FFT) of an ideal 8b ADC's output. The input sine wave manifests itself as large single tone spur at the corresponding frequency. The rest of the frequency contents are all due to quantization and sampling, which almost resembles a flat noise floor. If the ratio of

tone power and the noise floor power is evaluated, we also arrive at 50dB as before. As another step of simplification, quantization errors can be treated as “white” noise due to its near-constant spectrum in the frequency domain [12]. This justifies the use of SQNR again as a reasonable metric. While these first-order analyses provide simple

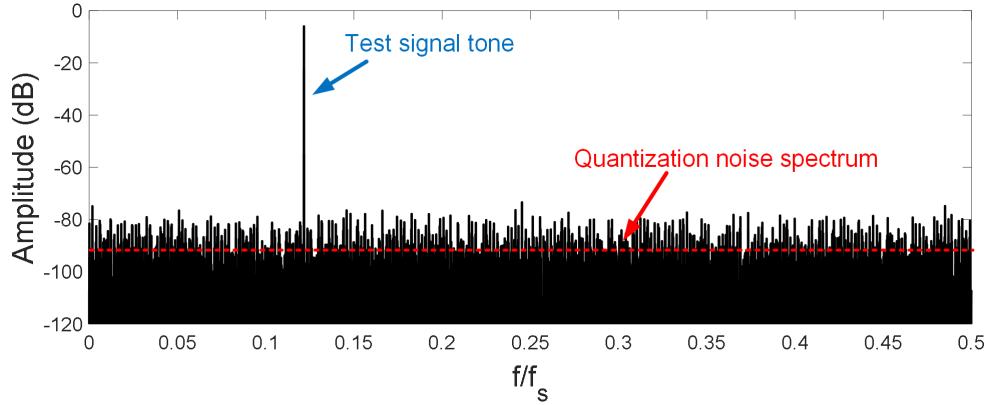


Figure 2.2: FFT of example ideal 8b ADC

ways for engineers to communicate about an ADC’s resolution and performance, we will expand on these equations in later sections to see their limitations in a link context.

2.1.2 ADC nonidealities and ENOB

Even though quantization is an ADC’s most fundamental error source, all real ADCs introduce other nonidealities, such as noise and distortion. Some of the more dominant ADC impairments include thermal noise, front-end static nonlinearity, and differential/integral nonlinearity (DNL/INL). Figure 2.3(a) shows a more complete ADC model with these nonidealities added. Any real circuits have thermal noise that adds a random voltage value to the input for every sample. The static nonlinearity of the ADC front-end will compress large amplitude signals. DNL and INL changes the

quantization error function shape. DNL measures an ADC's quantization step uniformity; in other words, an ADC with DNL will not have equal step sizes given by Δ , and this can be modeled as another error source in addition to the ideal quantization error. INL is a measure of how much the ADC deviates from an ideal quantizer at each output code, which is similar to static nonlinearity. Thus, Figure 2.3(b) shows an equivalent model for such a nonideal ADC, which will come in handy for subsequent analysis.

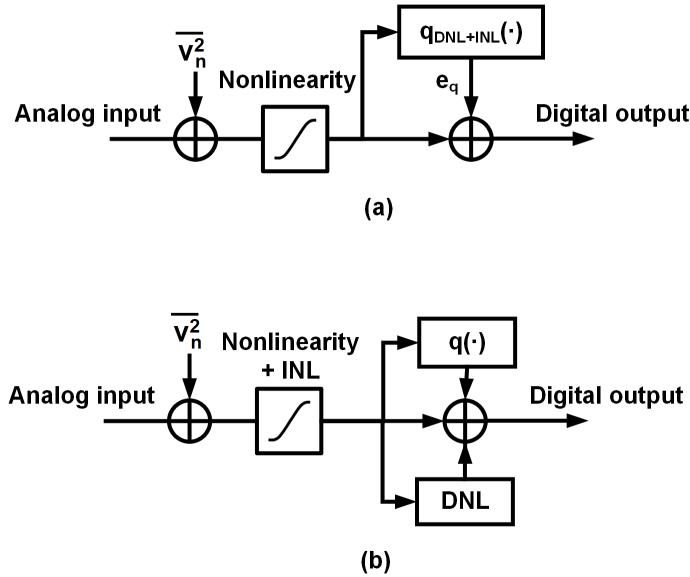


Figure 2.3: (a) ADC model with nonidealities added and (b) equivalent model for ADC with nonidealities

These nonidealities also affect the ADC's output spectrum in the frequency domain differently. Figure 2.4 shows a FFT of an example nonideal 8b ADC. Nonlinearity and INL creates distortion and harmonics in the frequency domain, shown by the extra spurs at multiples of the signal's fundamental frequency. The DNL and thermal noise adds to quantization noise and elevates the flat noise floor. Because sinusoidal inputs are typically sampled coherently when testing ADCs and quantization noise [13], thermal noise and DNL become indistinguishable, as shown in this example plot. As

a simplification, different “noise” sources’ power are combined together and compared with the signal power.

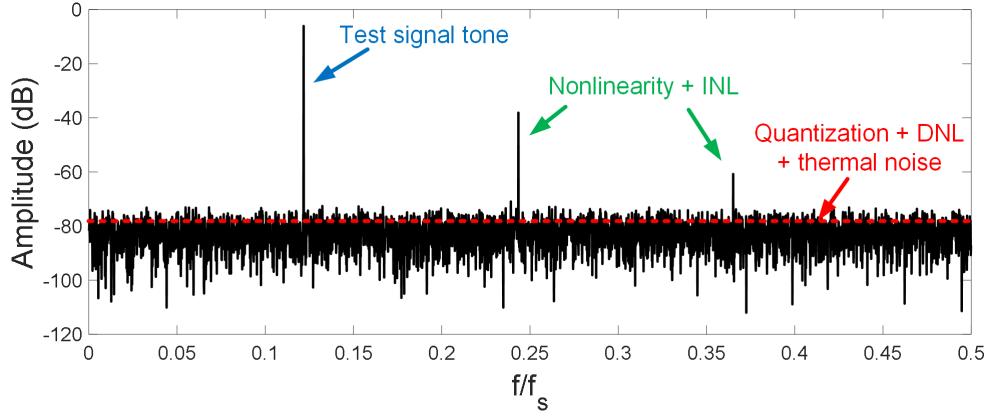


Figure 2.4: FFT of example nonideal 8b ADC

In the shown plot case, the ADC’s imperfections include both noise and distortion, therefore the metric typically used is called signal-to-noise and distortion ratio (SNDR). Anything other than the fundamental tone is considered either noise or distortion. Therefore, given the magnitude of an ADC’s FFT, $Y(f)$, and fundamental frequency, f_0 , SNDR can be numerically calculated by

$$\text{SNDR} = \frac{Y^2(f_0)}{\sum_{k \neq f_0} Y^2(k)} \quad (2.5)$$

Similar to SQNR, SNDR can also be expressed in dB. For this particular example, the ADC’s SNDR is 31dB.

To take the abstraction one step further, effective number of bits (ENOB) was introduced. ENOB equates a nonideal ADC’s SNDR to that of an ideal ADC to provide a expression that is intuitive and convenient. By simply replacing SQNR in

(2.4) with SNDR and B with ENOB, we obtain

$$\begin{aligned} \text{SNDR(dB)} &= 6.02\text{dB} \cdot \text{ENOB} + 1.76\text{dB} \\ \text{ENOB} &= \frac{\text{SNDR(dB)} - 1.76\text{dB}}{6.02\text{dB}} \end{aligned} \quad (2.6)$$

2.1.3 ENOB's statistical implications

One might think that using ENOB is sufficient for determining ADC requirements in link applications. To show why this is not the case, we will follow through an analysis under the ENOB assumptions in this section. We will demonstrate that such oversimplifications could and lead to pessimistic results and impact a link system's design decisions. The importance of separating noise sources that are bounded (e.g. ISI, nonlinearity, etc.) and unbounded (e.g. circuit thermal noise) is well-known [14, 15, 16]. However, ENOB as a metric models each error source in Figure 2.3 as a unbounded Gaussian white noise, as shown in Figure 2.5, when in fact most of these nonidealities generate errors that are bounded.

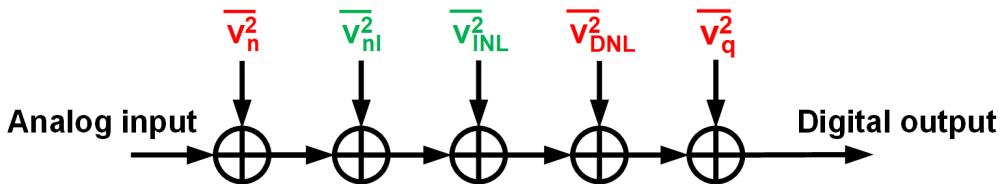


Figure 2.5: Simplified ADC model with noise and distortion

Because of the Gaussian assumption, SNDR and ENOB essentially sums all noise power together, denoted by σ 's in (2.7), and compares it with the signal power. In other words, ENOB treats all noise/error sources the same way, ignores their different natures and only considers noise power. For example, we have already qualitatively established that quantization noise presents bounded errors set by ADC's step size, but thermal noise exhibits an unbounded nature in its noise values. Both

nonlinearity and INL generates harmonics in the frequency domain, which shows their input dependency. On the other hand, thermal noise, quantization and DNL can be approximated as input independent noise sources.

$$\text{SNDR} = \frac{P_{sig}^2}{\sigma_n^2 + \sigma_{nl}^2 + \sigma_{INL}^2 + \sigma_{DNL}^2 + \sigma_q^2} \quad (2.7)$$

Link system designers need to estimate BER when specifying requirements for different blocks. As discussed in Section 1.1, BER is calculated by integrating the tail portion of the error PDF that crosses over the decision boundary. A Gaussian distribution (normal distribution) with certain noise power, σ is assumed for such an error PDF. A Gaussian distribution with mean offset μ and standard deviation σ is described as

$$N(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (2.8)$$

This assumption has served the purpose reasonably well until now since Gaussian distributions are good approximations for major noise sources, such as circuit thermal noise. Using Gaussian PDFs also relates the system's BER estimate directly with SNR in conventional mixed-signal links. Figure 2.6 shows a Gaussian distribution and the portion of the tail that contributes to BER. The integral of a Gaussian function is known as the normal cumulative density function (CDF), $\Phi(x)$, defined as

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{u^2}{2}} du \quad (2.9)$$

The estimated BER can then be calculated as

$$\begin{aligned} \text{BER} &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^0 e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-\frac{\mu}{\sigma}} e^{-\frac{u^2}{2}} du \\ \text{BER} &= \Phi\left(-\frac{\mu}{\sigma}\right) = Q\left(\frac{\mu}{\sigma}\right) \end{aligned} \quad (2.10)$$

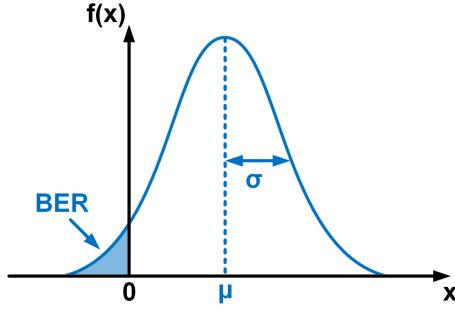


Figure 2.6: Gaussian distribution and BER

The Q-function, $Q(x)$ is interchangeable with $\Phi(x)$ for the purpose of BER estimation, and makes the argument positive. It becomes clear that the term μ/σ is precisely the SNR of the system. In the context of a link using PAM2, μ is half of the received signal's eye height or main cursor of the channel, and σ is the noise power of all ISI and circuit impairments. Figure 2.7 shows this relationship between SNR and BER for PAM2, which has a very steep slope, especially for the high SNR region.

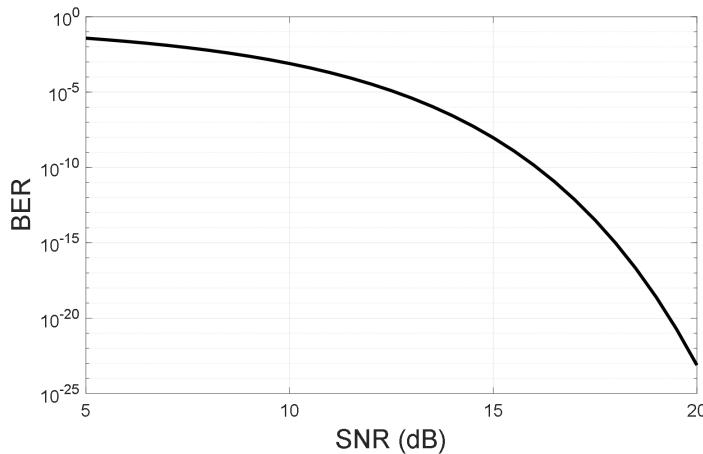


Figure 2.7: BER vs. SNR under Gaussian assumption for PAM2

The next “natural” step is substituting SNDR into the Q-function to obtain a BER estimation for an ADC-based link. Since ENOB is derived from testing the ADC with a sine wave of amplitude $A = FSR/2$, while a link’s signal of interest is

the half eye height μ , a signal power correction term, $\mu/(A/\sqrt{2})$ needs to be applied and we obtain the following equation

$$\text{BER} = Q(\text{SNDR} \cdot \frac{\mu}{A/\sqrt{2}}) = Q\left(\sqrt{3} \cdot \frac{\mu}{FSR/2} \cdot 2^{\text{ENO}}\right) \quad (2.11)$$

In the ideal case of an open eye that completely saturates the ADC's FSR , μ will be equal to $FSR/2$. While the Q-function provides a very straightforward way of estimating BER using ADC metrics such as ENOB, it also makes strong assumptions that are not suitable for ADC-based links. The Q-function assumes that all noise sources have Gaussian PDFs, and only deals with their collective noise power (2.7). As we have briefly touched on previously, this is a wrong assumption due to the different nature of the ADC's nonidealities even though it provides a simpler way of calculating BER. The following sections will discuss how using the Q-function and ENOB (SNDR) can lead to over-specification and present the correct statistical models for each ADC nonideality.

2.2 Statistical framework for ADC quantization

This section focuses on ADC quantization first to build up a statistical framework using PDFs, and compares the BER estimation results with that of the ENOB model.

2.2.1 Quantization noise and PDF

As a start, the first assumption in the ENOB model that quantization acts as “white” noise needs to be validated. Figure 2.8 shows a simplified discrete time model of an ADC-based link with only thermal noise and quantization considered. The data sequence $d[n]$ is sent through the channel, modeled by a filter vector \vec{h} . For this analysis, we assume a normalized \vec{h} with main cursor (normalized to 1), P pre-cursors

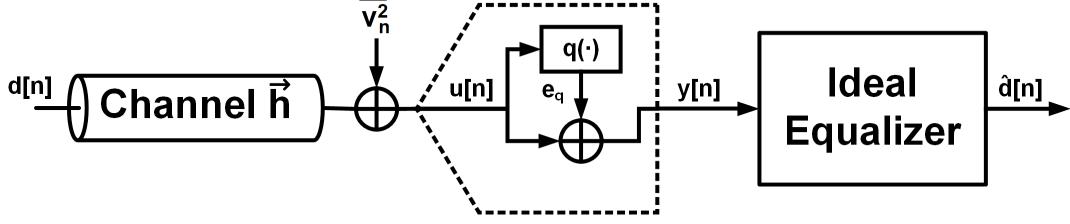


Figure 2.8: Discrete time model of an ADC-based link

and Q post-cursors.

$$\vec{h} = \langle h_{-P}, h_{-P+1}, \dots, h_{-1}, 1, h_1, h_2, \dots, h_{Q-1}, h_Q \rangle \quad (2.12)$$

$u[n]$ is the channel output (convolution of data and channel vectors) with thermal noise $\overline{v_n^2}$ added, and gets quantized by the ADC. Output of the ADC $y[n]$ is further processed by an equalizer. The following equation describes $y[n]$ and shows its relationships with other variables in this model

$$y[n] = u[n] + q(u[n]) = \sum_{k=-P}^Q \vec{h}[k]d[n-k] + v_n[n] + q(u[n]) \quad (2.13)$$

Let's assume there is an ideal equalizer that cancels all pre-cursors and L post-cursors without adding extra noise to the system. The recovered data, $\hat{d}[n]$ can then be expressed as

$$\hat{d}[n] = d[n] + e[n] = d[n] + \sum_{k=L}^Q \vec{h}[k]d[n-k] + v_n[n] + q(u[n]) \quad (2.14)$$

which includes the original $d[n]$ and three error terms - residual ISI shown by the partial convolution of data and channel, random thermal noise and quantization error that is dependent on $u[n]$. If these errors terms are expanded and studied further as

below,

$$e[n] = \sum_{k=L}^Q \vec{h}[k]d[n-k] + v_n[n] + q \left(\sum_{k=-P}^Q \vec{h}[k]d[n-k] + v_n[n] \right) \quad (2.15)$$

$$= e_{\text{resISI}} + v_n[n] + q(\text{cancelled ISI} + e_{\text{resISI}} + v_n[n]) \quad (2.16)$$

both residual ISI, e_{resISI} , and thermal noise $v_n[n]$ show up in the argument of quantization error function $q(\cdot)$. This signifies a correlation/dependency among these error sources. However, the canceled ISI term is also a part of q 's argument and becomes crucial in reducing these error sources' correlation/dependency significantly. For example, when the “noise” power of the canceled ISI is much larger than both the residual ISI and thermal noise, which is true for most practical cases, $e[n]$ can be rewritten as

$$e[n] = e_{\text{resISI}} + v_n[n] + q(\text{ISI}) \quad (2.17)$$

and in essence becomes the sum of three independent error sources. In addition, ISI also provides enough signal activity and randomizes the ADC's input so that quantization can be treated as noise instead of a deterministic error as discussed in the previous section.

One way to visualize this effect is depicted in Figure 2.9. A typical ADC input PDF (for PAM2) is projected onto the rotated quantization sawtooth functions of three ADCs with different resolutions. The outputs are the quantization error PDFs resulting from ADCs of low to moderate resolutions. In the 2-bit case, the input PDF mostly exercises the “linear” portion of the sawtooth function, and rarely gets folded within the $\pm\Delta/2$ bound. As a result, the corresponding error PDF still resembles the input PDF shape. In the 3-bit case, the input PDF starts to get folded more, but there is still some correlation between the input PDF and error PDF. It's only when the LSB size becomes sufficiently small compared to the span of the input PDF, as shown in the

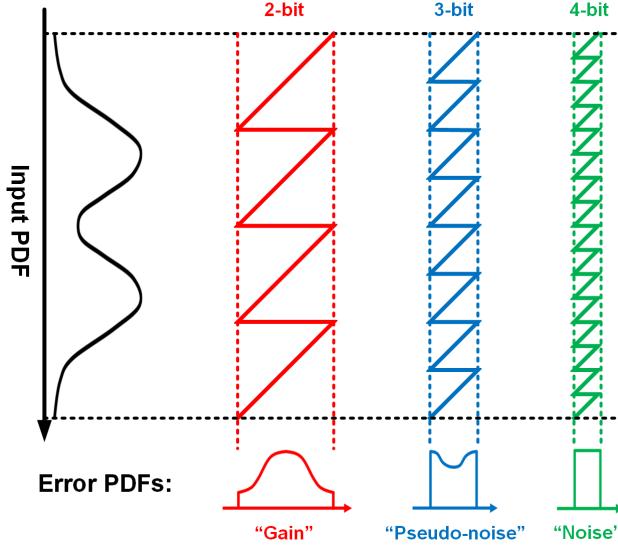


Figure 2.9: Quantization error PDFs

4-bit case, that the error PDF has a uniform distribution. When this condition is met, quantization error can be approximated as an independent noise source (Appendix A shows a more rigorous derivation for quantization's pseudo-independent nature). It is also important to note that the variance of a uniform PDF of width Δ is exactly $\Delta^2/12$, same as what was derived before.

In conclusion, when channel ISI is dominant compared to an ADC's LSB size, quantization can be modeled as an independent noise source in the system with a uniform PDF. Therefore, the correct statistical model becomes similar to the ENOB model except that quantization has a very different distribution, shown in Figure 2.10. The ENOB model in Figure 2.10(a) incorrectly assigns a Gaussian distribution to quantization noise when in reality it has a uniform distribution, depicted by the statistical model in Figure 2.10(b). One thought experiment can demonstrate how the ENOB model is overly pessimistic. Due to the unbounded nature of Gaussian distributions, thermal noise is the main source of BER. Let's imagine an ADC's LSB size is much smaller than the channel's main cursor value and there is no thermal

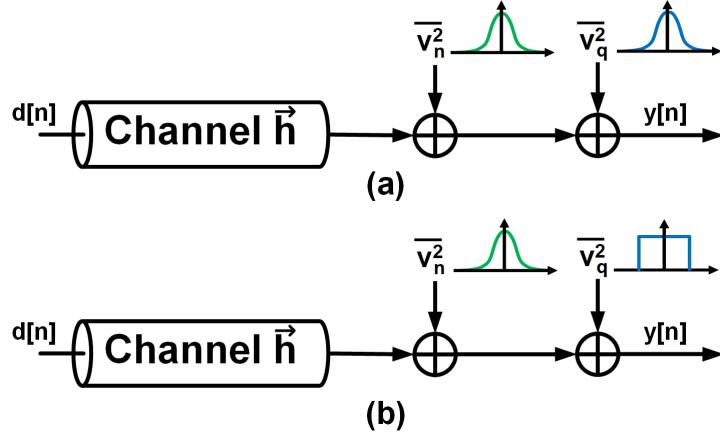


Figure 2.10: (a) ENOB model and (b) statistical model for quantization

noise; such a system should have zero BER since the quantization error is not large enough to incur any wrong bit decision. However, the ENOB model will still estimate a non-zero BER, albeit small, because quantization has the infinite tail of a Gaussian PDF under this model instead of having a maximum error bound. The next subsection uses the statistical model to derive a new method of BER estimation and compares it with the ENOB model quantitatively.

2.2.2 BER estimation with uniform PDF

Once quantization noise can be viewed as an independent noise source, its PDF will convolve with the thermal noise PDF. There is a closed-form result when convolving a uniform distribution with a Gaussian distribution. A uniform box, $\Pi(x)$ can be viewed as the difference of two step functions, $u(x)$. Thus, a uniform distribution bounded by $\pm\Delta/2$ can be expressed as

$$\Pi(x|\Delta) = \frac{1}{\Delta} \left(u\left(x + \frac{\Delta}{2}\right) - u\left(x - \frac{\Delta}{2}\right) \right) \quad (2.18)$$

Convolving with a step function is equivalent to integration. Therefore, the final error PDF $f_E(e)$, which is a convolution between a Gaussian and a uniform distribution is

$$f_E(e|\sigma, \Delta) = N(e|\sigma) * \Pi(e|\Delta) = \frac{1}{\Delta} \left(\Phi \left(\frac{e + \Delta/2}{\sigma} \right) - \Phi \left(\frac{e - \Delta/2}{\sigma} \right) \right)$$

The error PDF used for BER estimation is derived to be a difference of two normal CDFs with mean shifted by $\pm\Delta/2$. Figure 2.11 shows the shape of f_E given thermal noise σ and ADC LSB size Δ . It is intuitive that in the extreme case of $\sigma \gg \Delta$, the thermal noise PDF dominates and f_E still resembles a Gaussian shape. The more interesting case is when $\sigma \ll \Delta$, and thermal noise only contributes to adding a “skirt” to the uniform box. Contrary to the ENOB model’s assumptions, the error PDF is very different from a Gaussian distribution. As a result, this will lead to

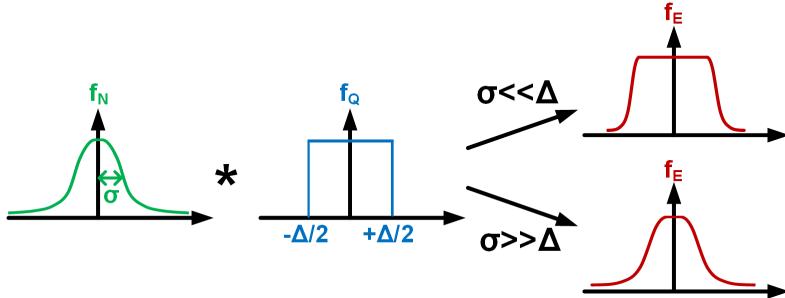


Figure 2.11: Convolution of thermal noise and quantization PDFs

very different BER estimation results, especially for low ADC resolution. Using the statistical model, the BER estimation should be the tail integral of $f_E(e)$, which involves integrating the normal CDF function. There is also a closed-form formula for such an integration, denoted as $Z(x)$ shown below

$$Z(x|\mu, \sigma) = \int_{-\infty}^x \Phi \left(\frac{u - \mu}{\sigma} \right) du = (x - \mu)\Phi \left(\frac{x - \mu}{\sigma} \right) + \sigma N \left(\frac{x - \mu}{\sigma} \right) \quad (2.19)$$

Therefore, a closed-form equation for statistical BER estimation can also be obtained by using (2.19)

$$\begin{aligned} \text{BER} &= \frac{1}{\Delta} \int_{-\infty}^0 \left(\Phi\left(\frac{x - \mu + \Delta/2}{\sigma}\right) - \Phi\left(\frac{u - \mu - \Delta/2}{\sigma}\right) \right) dx \\ &= \frac{1}{\Delta} \left(Z\left(0 \middle| \mu - \frac{\Delta}{2}, \sigma\right) - Z\left(0 \middle| \mu + \frac{\Delta}{2}, \sigma\right) \right) \end{aligned} \quad (2.20)$$

This equation will provide a much more accurate BER estimate and help avoid overdesign and also become helpful when we investigate other ADC nonidealities in later sections.

With both (2.11) and (2.20), a quantitative comparison can be made to evaluate these two BER estimation methods. Figure 2.12 serves as a visual aid for setting parameters used in these BER estimation methods. Again, the main signal of interest is half the eye height μ . There is Gaussian-distributed thermal noise around the eye's data levels with noise power σ , and the eye is quantized with step size Δ given a B_{eye} -bit ADC. By changing B_{eye} and sweeping σ , Figure 2.13 is generated using the two models' BER estimation equations for comparison. The BER curve for infinite resolution is also plotted as a reference (which is the same curve as in Figure 2.7). For the $B_{eye} = 4$ case, both methods overlap for low SNR. This means that thermal noise is the dominant noise source and $B_{eye} = 4$ gives sufficiently high resolution that quantization noise is negligible for most applications' target BER ($< 10^{-15}$). The difference is much more pronounced for $B_{eye} = 2$. Even at a moderate BER target of 10^{-8} , the ENOB method will over specify the SNR requirement by 3dB compared to the statistical method. This roughly translates to a 2x increase in circuit power for a 3dB increase in SNR. One other scenario in which the ENOB method will lead to overdesign is by looking at the vertical line at a fixed system SNR of 18dB. For this example, the ENOB method concludes that $B_{eye} = 2$ is not a feasible solution while statistical method predicts that $B_{eye} = 2$ gives acceptable performance that is similar

to the ENOB method's $B_{eye} = 3$ curve. Consequently, 1-bit can be saved simply by using the correct statistical model for BER estimation. Therefore, these curves demonstrate that the ENOB model is not suitable for ADC-based link applications and a full statistical framework is necessary in order to avoid over specifications.

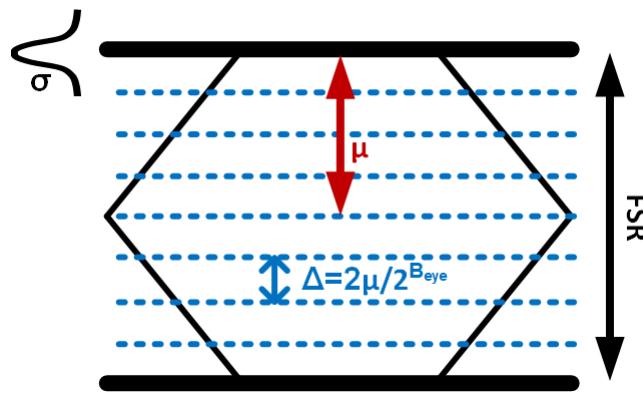


Figure 2.12: BER estimation parameters illustrated in an eye diagram

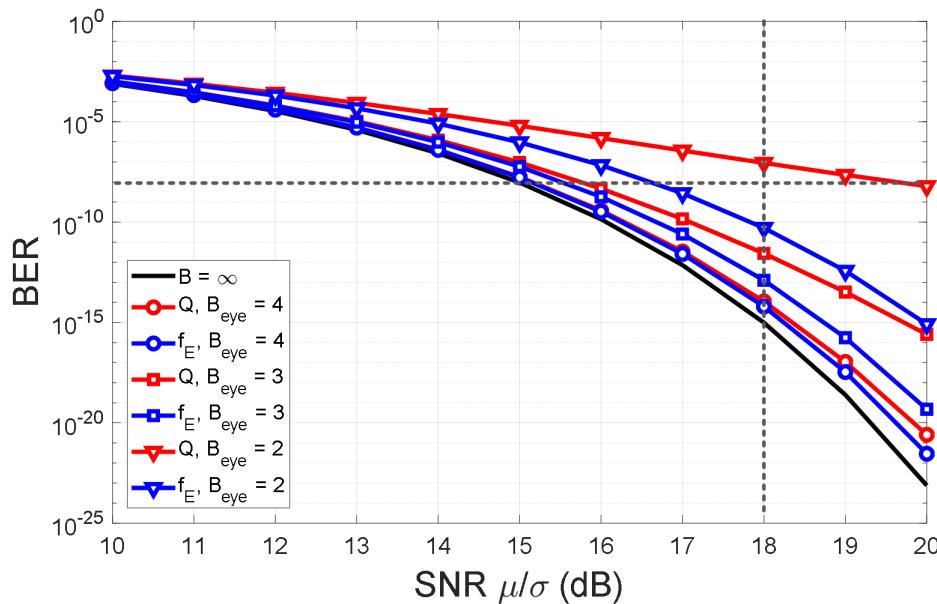


Figure 2.13: Comparison of ENOB model vs. statistical BER estimation curves

2.3 Statistical framework for other ADC errors

In this section, we will extend the statistical framework to include other ADC nonidealities such as DNL and nonlinearity/INL. These will serve as examples to gain further insights into how other ADC errors contribute to system performance degradation. Rules of thumb will also be developed to aid design decisions.

2.3.1 DNL

As discussed in previous sections, DNL quantifies the non-uniformity in an ADC's quantization step. DNL is typically measured in LSBs; a DNL of $\pm D/2$ LSB generates extra quantization error with a range of $\Delta_D = \Delta \cdot D$. Therefore, DNL effect can be directly viewed in the quantization error function of an ADC with DNL, as shown in Figure 2.14. We have already established that the quantization noise PDF for an ideal ADC is a uniform distribution. The quantization PDF with DNL can then assume a different shape such as a triangular distribution, shown in Figure 2.14(b), when DNL adds another $\pm\Delta/2$ to the original quantization and is uniformly distributed.

The new quantization noise PDF can be approximated by giving DNL its own noise PDF and convolving with the ideal quantization uniform PDF. It is important to note that we are using this approximation to gain some insight into DNL's impact on system performance and arrive at a heuristic that is useful for designers. The exact distribution that DNL noise depends on the nature of what generates DNL in an ADC. For example, in a typical Flash ADC in which each comparator offset is calibrated to be within a certain residual, a good approximation for DNL PDF is another uniform box with its own bounds. In a SAR ADC in which the major source of DNL comes from DAC capacitor mismatch, the DNL PDF will have a bounded Gaussian shape.

We will use the model in Figure 2.15 and a uniform DNL PDF as an example to

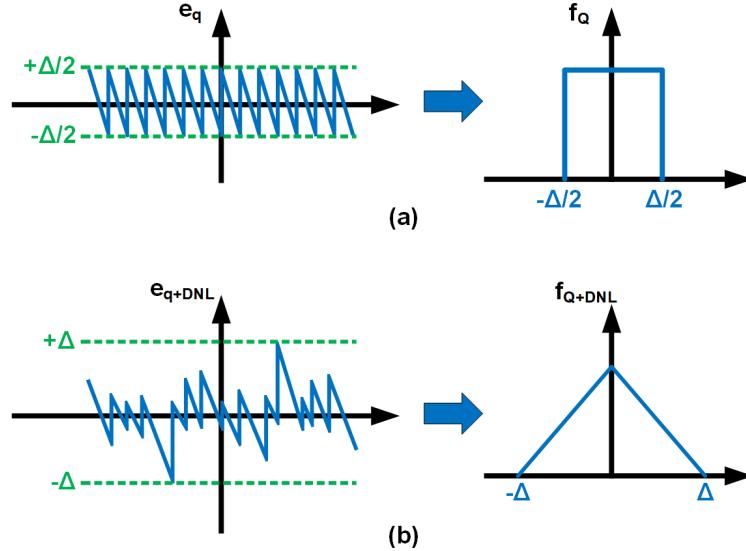


Figure 2.14: (a) Ideal quantization error function and (b) quantization error function with DNL

further study its effect on BER within our statistical framework. To recap, we have found a closed-form error PDF when convolving quantization's uniform distribution with thermal noise' Gaussian distribution by taking the difference of two normal CDFs. For a uniform DNL PDF bounded by $\pm\Delta_D/2$, it is essentially performing

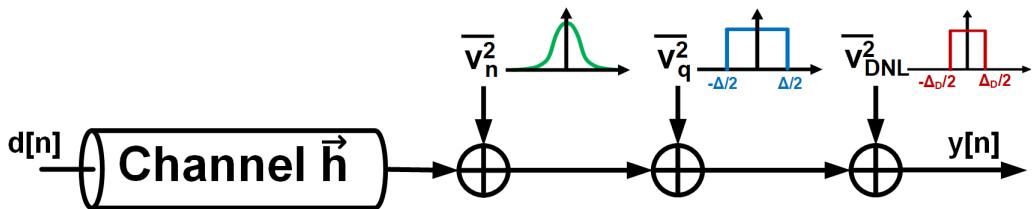


Figure 2.15: Statistical model incorporating DNL noise

a difference of integration operation again similar to ideal quantization. We have already done such an integration, denoted by function $Z(x)$, in Section 2.2.2 when deriving an equation for statistical BER estimation. Therefore, the error PDF with

DNL can be found to be

$$\begin{aligned} f_{E,DNL}(e|\Delta, \Delta_D, \sigma) &= \Pi(e|\Delta_D) * f_{E,ideal}(e|\Delta, \sigma) \\ &= \int_{-\infty}^{e+\frac{\Delta_D}{2}} f_{E,ideal}(x|\Delta, \sigma) dx - \int_{-\infty}^{e-\frac{\Delta_D}{2}} f_{E,ideal}(x|\Delta, \sigma) dx \end{aligned}$$

$$\begin{aligned} f_{E,DNL}(e|\Delta, \Delta_D, \sigma) &= \frac{1}{\Delta\Delta_D} \left(Z\left(x \left| -\frac{\Delta + \Delta_D}{2}, \sigma\right.\right) - Z\left(x \left| -\frac{\Delta - \Delta_D}{2}, \sigma\right.\right) \right. \\ &\quad \left. - Z\left(x \left| \frac{\Delta - \Delta_D}{2}, \sigma\right.\right) + Z\left(x \left| \frac{\Delta + \Delta_D}{2}, \sigma\right.\right) \right) \quad (2.21) \end{aligned}$$

Similar to how uniform quantization PDF turns thermal Gaussian into a difference of $\Phi(x)$, uniform DNL PDF turns each normal CDF into a difference of $Z(x)$. The BER estimate is then again calculated from the tail integral of this error PDF with DNL.

$$\text{BER}_{DNL} = \int_{-\infty}^0 f_{E,DNL}(x - \mu|\Delta, \Delta_D, \sigma) dx \quad (2.22)$$

This closed-form error PDF and its tail integral might be too complicated and not quite intuitive despite being accurate. As illustrated in Figure 2.16, we propose two ways of approximating (2.21) to gain insight into the DNL's effect on an ADC's effective resolution with respect to BER performance.

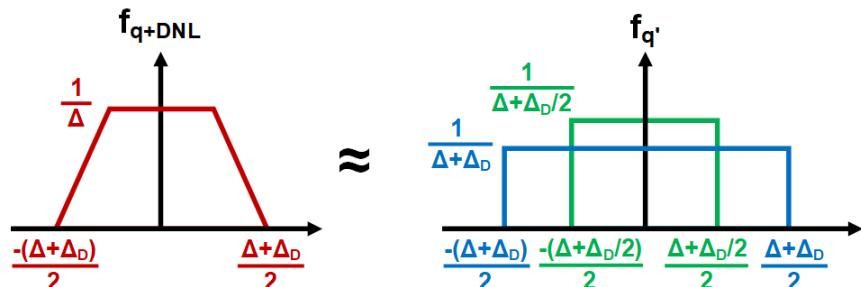


Figure 2.16: Two ideal ADC approximations with different LSB sizes to estimate DNL effects

The two approximate ADCs have effective LSB sizes given by

$$\Delta_{upper} = \Delta + \Delta_D \quad (2.23)$$

$$\Delta_{lower} = \Delta + \frac{\Delta_D}{2} \quad (2.24)$$

We now compare the BER estimation curves between BER_{DNL} and those from approximating with ADCs of equivalent step sizes. Figure 2.17 shows a group of BER curves for different base resolution B_{eye} and DNL values. It is expected that for the high-resolution $B_{eye} = 4$ case, all estimation methods give similar results and the performance degradation is insignificant, since thermal noise is the dominant noise source in this regime. For small DNL (± 0.3 LSB), the ideal ADC estimation using $\Delta + \Delta_D/2$ LSB size is more accurate in terms of approximating $Z(x)$ method. Differences start to show when base resolution starts to decrease and DNL increases.

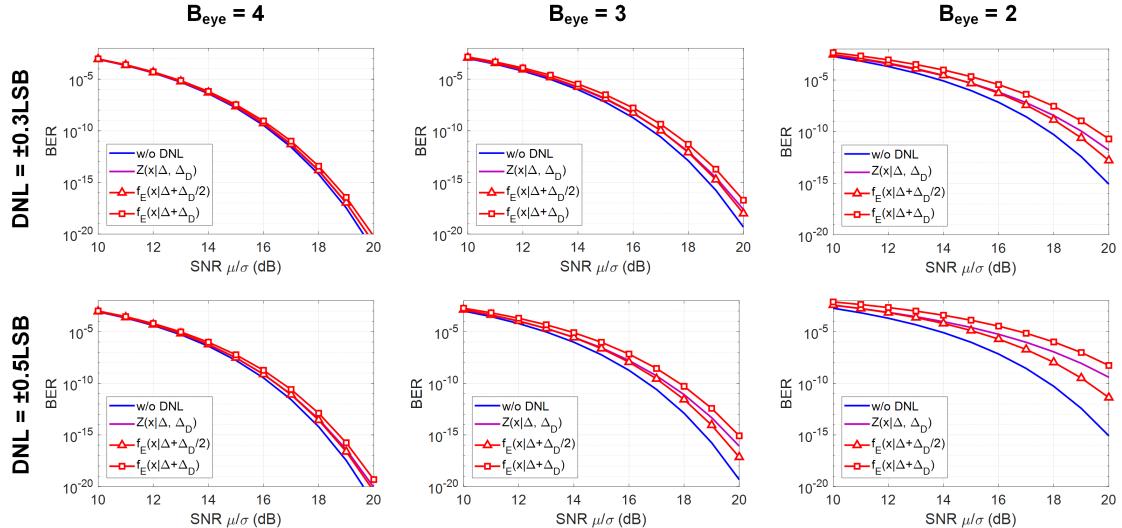


Figure 2.17: BER estimation curves comparison between $Z(x)$ method and two approximation methods

In all cases, the proposed two approximations serve as lower and upper bounds

for the $Z(x)$ estimation as described by the equation below, even though $\Delta + \Delta_D/2$ still seems to be a better approximation for the lower BER space. We can use this empirical result to devise a rule of thumb that lumps the DNL specification into an equivalent resolution, which we are now confident about from statistical BER analysis.

$$\int_{-\infty}^0 f_{E,ideal}(x - \mu|\Delta + \Delta_D/2, \sigma)dx < \text{BER}_{DNL} < \int_{-\infty}^0 f_{E,ideal}(x - \mu|\Delta + \Delta_D, \sigma)dx \quad (2.25)$$

To reiterate, an ADC with a uniformly distributed DNL of $\pm D/2$ LSB can be approximated by an ADC with equivalent LSB size of $\Delta + \Delta \cdot D/2$. To translate this statement into a resolution requirement, an effective resolution B_{eff} has the following relationship with ADC's original resolution B

$$B_{eff} = B - \log_2 \left(1 + \frac{D}{2} \right) \quad (2.26)$$

For example, any ADC with a ± 0.5 LSB DNL has an equivalent resolution degradation of $\log_2(1.5) \approx 0.6$ bits. This rule of thumb is helpful when accounting for the DNL's performance impact and will be applied in future chapters when studying ADC resolution requirements.

2.3.2 Nonlinearity/INL

As higher PAM schemes are employed, circuit nonlinearity has become a much more crucial aspect when considering overall system performance. Even though nonlinearity has been heavily studied and various linearization techniques have been developed within many years of ADC research, nonlinearity requirements remain unclear for ADC-based links. Nonlinearity, including ADC INL, has only been investigated in the frequency domain as harmonics because of sine wave test inputs, but we have concluded that a PDF domain framework provides a more direct correlation with the

system's final BER performance. Therefore, this section explores how nonlinearity can be modeled statistically.

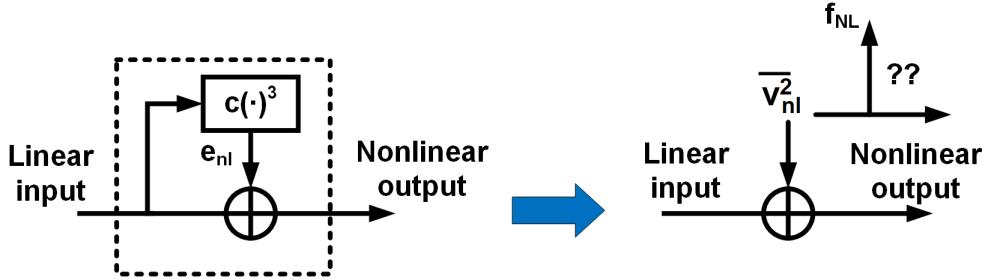


Figure 2.18: Statistical model for nonlinearity

One of the most prevalent nonlinearity error source is from front-end's static nonlinearity, including INL. Third-order compressive nonlinearity can serve as a realistic model since it is present in most circuits without any calibration schemes. Figure 2.18 shows the statistical model conversion for nonlinearity error. The compressive coefficient c determines how much nonlinear error is generated, and in this example c is normalized to an input signal of $FSR/2 = 1$. The static model for nonlinearity looks similar to that of quantization, and using the same assumptions as before, we can also treat nonlinearity as an independent noise source. However, the question remains as to what the nonlinearity noise PDFs are. Appendix B derives the PDF of the nonlinearity error $y = -cx^3$ to be

$$f_Y(y) = \frac{1}{3c} \left(-\frac{y}{c} \right)^{-2/3} f_X \left(\sqrt[3]{-\frac{y}{c}} \right) \quad (2.27)$$

As expected, the nonlinearity error PDF $f_Y(y)$ is dependent on the input PDF $f_X(x)$. In other words, $f_Y(y)$ will not only be dependent on what data is transmitted, but it also is a strong function of the channel that determines the input PDF.

To illustrate this point, an approximate input PDF for a PAM4 system is synthesized in Figure 2.19. The overall PDF $f_X(x)$ is the average of four conditional PDFs

when different data $D \in [-1, -\frac{1}{3}, +\frac{1}{3}, +1]$ is transmitted. In this particular example,

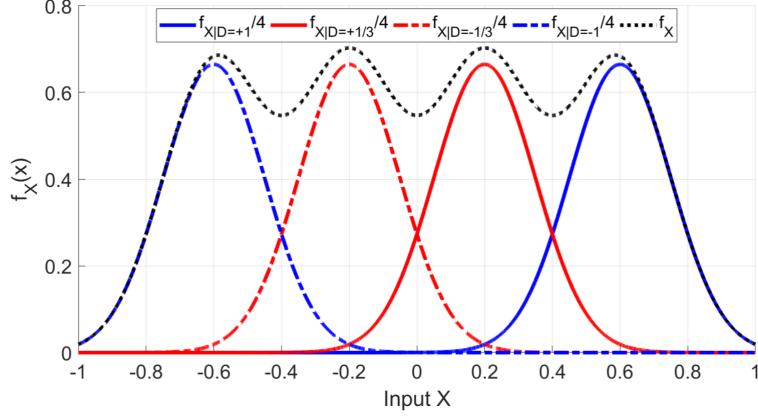


Figure 2.19: Example PDF of PAM4 receiver input signal

the transmitter and channel incur some amplitude loss on transmitted data and the received data are centered around ± 0.6 for $D = \pm 1$ and ± 0.2 for $D = \pm \frac{1}{3}$. Each conditional PDF, which is the result after ISI and noise are added to the transmitted data, is modeled as a Gaussian distribution for simplicity, even though in reality it strongly depends on the actual channel response and noise environment. Nevertheless, this will still provide us with a reasonable approximation to study the nature of the nonlinearity. Therefore, the following expression is used for $f_X(x)$ for subsequent analysis

$$f_X(x) = \frac{1}{4} \left(N(x| -0.6, \sigma_{ISI}) + N(x| -0.2, \sigma_{ISI}) + N(x| +0.2, \sigma_{ISI}) + N(x| +0.6, \sigma_{ISI}) \right) \quad (2.28)$$

The ISI and noise term σ_{ISI} , which in this example is set to 0.15, is a variable that will be changed to see how nonlinearity errors are affected. It is very important to note that so far in our analysis we have assumed that quantization and DNL error has the same PDF regardless of what data is transmitted for simplicity. However, this assumption is no longer acceptable for this noise source due to the nature of

nonlinearity; each data Gaussian will yield different nonlinearity errors, demonstrated by the diagram in Figure 2.20(a), because the conditional input PDFs are projected onto different portions of the static nonlinear error function $-cx^3$. By substituting each conditional input PDF $f_{X|D}$ into (2.27), the conditional nonlinearity error PDFs $f_{Y|D}$ are obtained in Figure 2.20(b). When the PDFs are plotted on a log scale, we

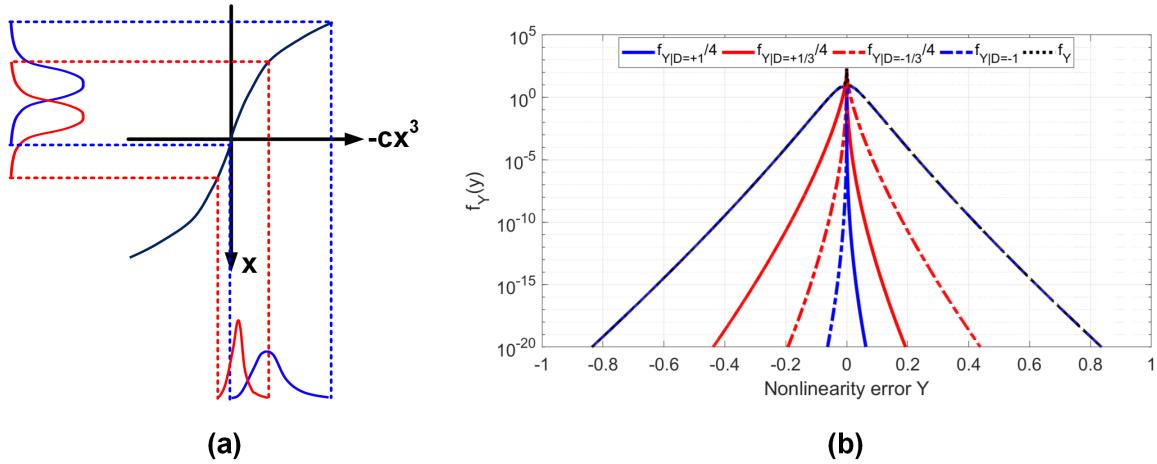


Figure 2.20: (a) Nonlinearity error dependency on input PDF and (b) example nonlinearity error PDFs from input PDF in equation 2.28 and $c = 0.1$

see that larger and more errors are generated for $D = \pm 1$ than for $D = \pm \frac{1}{3}$, which means it is more likely to have a bit error for large value data when nonlinearity is present. The nonlinearity errors also have non-zero mean also because their PDFs are “biased” at different points on the nonlinearity curve. To first order, this results in a linear gain compression for the data signal. By finding the new compressed data levels, new decisions levels, such as s in Figure 2.21, can be used to lower BER. It is also important to note that there are two ways for the small-valued data to make a decision error while large valued data only makes a mistake in one direction. These are aspects of nonlinearity error that an ENOB model will not catch.

With the example conditional PDFs and decision levels s , we can exercise the same comparison between ENOB and statistical models for nonlinearity. Besides, σ_{ISI} and

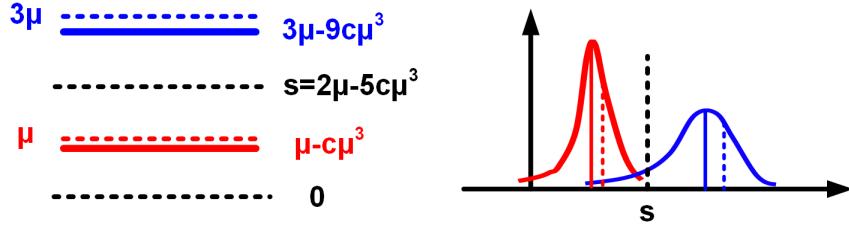


Figure 2.21: Gain compression on data levels (blue/red dashed lines to solid levels) due to nonlinearity and adjusted decision levels (black dashed lines)

compression coefficient c 's effects on BER can also be studied. For simplicity, we will assume $f_{Y|D}$ PDFs are symmetrical with respect to the transmitted data, and thus we only focus on the positive data. By adding and sweeping the thermal noise level σ with respect to μ , the BER with nonlinearity can be numerically calculated as

$$BER_{NL} = \frac{1}{2} \left(\int_{-\infty}^{\mu} N(x|0, \sigma) * f_{Y|D=1/3}(x) dx + \int_{s-\mu}^{+\infty} N(x|0, \sigma) * f_{Y|D=1/3}(x) dx + \int_{-\infty}^{3\mu-s} N(x|0, \sigma) * f_{Y|D=1}(x) dx \right) \quad (2.29)$$

Figure 2.22 shows the BER curves comparing ENOB and statistical nonlinearity model for different c and σ_{ISI} . The different c values and their corresponding signal to distortion ratios (SDR) in a sine wave test are presented. The ENOB model in which nonlinearity distortion is also treated as Gaussian noise gives BER curves that don't change with respect to input signal spread (σ_{ISI}), while on the other hand, the statistical model correctly predicts that BER is also heavily dependent on the input PDF in addition to how nonlinear the system is itself. This motivates the need to condition the input PDF to minimize the effect of nonlinearity.

Furthermore, we see that nonlinearity becomes negligible especially for low SNR and high BER. This means that there is no need to design an ADC with a linearity requirement on the same order of required resolution, especially when the input PDF

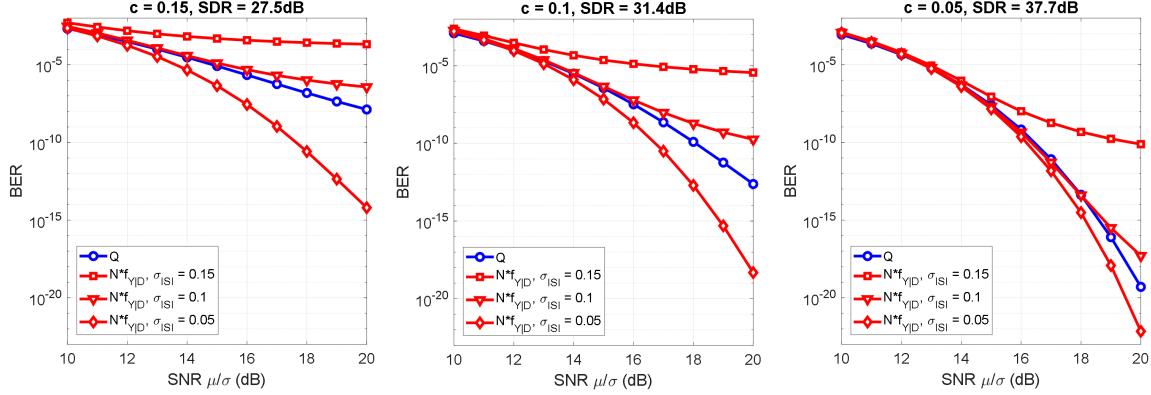


Figure 2.22: BER curves comparing ENOB and statistical nonlinearity model for different c and σ_{ISI}

is benign enough and designing for higher BER standards. Another interesting feature to point out is that for a large input signal spread ($\sigma_{ISI} = 0.15$ compared to $\mu = 0.2$), nonlinearity can be bad enough to create a BER floor. This could be explained by looking at the input PDF in Figure 2.19 again and realize there is a significant portion of the large data Gaussian tail that exceeds the full scale range. Even though this is a soft nonlinear block, the resultant errors could still be large enough to directly cause bit errors.

A similar phenomenon occurs when the ADC is clipped beyond its full scale range. FSR clipping is a hard nonlinearity that can also be investigated in the PDF domain. We revisit the quantization error function in Figure 2.1 again and recognize that the quantization error extends indefinitely beyond $\pm FSR/2$. Figure 2.23 demonstrates the mechanism behind full scale clipping in the PDF domain. For a small transmitted data whose conditional input PDF stays inside the ADC's FSR, the corresponding quantization PDF is the uniform distribution as discussed before. However, for the large value data whose conditional input PDF exceeds the ADC's FSR, the resultant quantization PDF becomes a uniform box with a long tail. This tail is exactly the portion of the input PDF outside of the ADC's FSR that is being folded over and

becomes part of the quantization errors. This negates the bounded nature assumption of quantization error and will contribute to BER significantly depending on the clipping probability. This has implications on the system adaption algorithm when tuning system gain and setting the ADC's input amplitude.

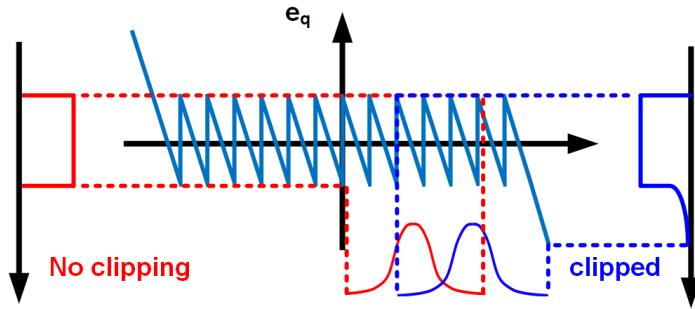


Figure 2.23: Full scale clipping and input PDF folding

To summarize, nonlinearity plays an interesting role in an ADC-based link due to its input PDF dependency. Therefore, BER needs to be estimated using conditional error PDFs. Static nonlinearity and INL has a linear gain compression effect, which can be alleviated by changing data decision level. More importantly, input PDF and its spread need to be controlled well so that large nonlinearity error and ADC FSR clipping occur with a probability well below the required BER. This motivates the need for good signal conditioning blocks before the ADC, which will be discussed in later chapters.

Chapter 3

Comparative Study on Equalizer Position

After many years of proven results, mixed-signal links adopted an architecture with TX FFE, RX CTLE and DFE due to both performance and implementation concerns. On the other hand, emerging ADC-based links allow more possible equalization positions as shown in Figure 3.1. Specifically, it is feasible now to implement a discrete time FFE on the RX side after the ADC, in addition to equalization before the ADC. However, for the first-generation ADC-based RX, most of the attention has been



Figure 3.1: Equalization positions in an ADC-based link

focused on implementation of high-speed and high-performance ADCs due to their various challenges. Even though there has been some systematic analysis, such as [17] and [18], a more fundamental understanding regarding the optimal location of equalizers (like FFEs) to achieve the best system performance is still needed. In this chapter, we present a comparative study on equalizer positions and their relationships

with both system performance as well as ADC requirements.

3.1 TX vs. RX FFE

As the data rate increased beyond 25Gbps, the pre-cursor ISI in a backplane/copper cable system became non-negligible. Thus, the need of a power efficient FFE is ever more important to effectively deal with pre-cursor ISI as well as long tails in the channel pulse response. The emergence of ADC-based RX allows system and circuit designers to re-evaluate the choice of TX FFE vs. RX FFE.

3.1.1 Fundamental SNR analysis

In order to compare the two options, we start with a simple discrete model in Figure 3.2 for both TX and RX FFE. Both the FFE and channel are modeled as a filter with

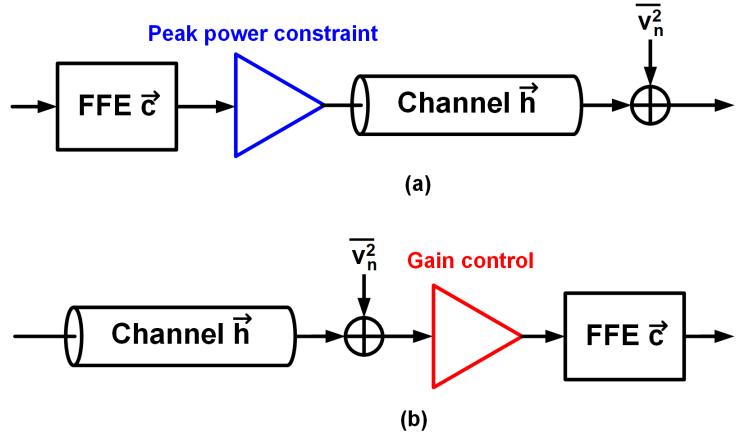


Figure 3.2: Discrete time models for links with (a) TX side FFE and gain block modeling peak power constraint and (b) RX side FFE with a gain control block

vectors \vec{c} and \vec{h} respectively. A noise term modeling cross talk and circuit thermal noise is added at the output of the channel. The critical differences lie in how the different locations of FFEs affect the signal gain in the respective models. The gain

limitation for TX FFE in Figure 3.2(a) is due to peak power constraint where the largest FFE output can not exceed a maximum swing on the TX output. Similarly, the gain block for RX FFE in Figure 3.2(b) controls the incoming signal amplitude.

Figure 3.3 takes a closer look at the operation of the TX FFE. An FFE attempts to compensate for the channel's low-pass nature, thus it is a high-pass filter in the frequency domain that amplifies the transition edges of the transmit data. A normalization factor needs to be applied afterwards, given by the sum of the coefficients' absolute values. This term is also known as the L-1 norm $\|\vec{c}\|_1$ of vector \vec{c} shown below

$$\|\vec{c}\|_1 = \sum_n |\vec{c}_n| \quad (3.1)$$

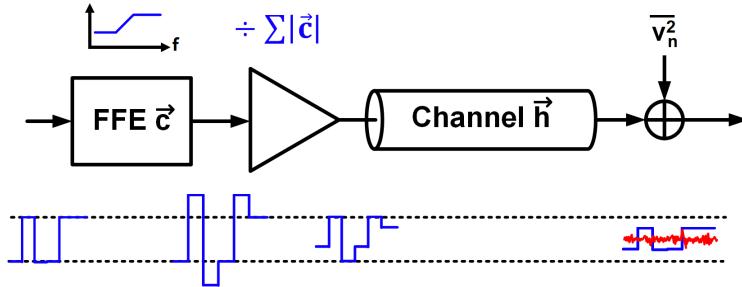


Figure 3.3: TX FFE operation principle

Given the noise power of channel output thermal noise σ_n and residual ISI power σ_{ISI} (which will also be proportional to signal amplitude), the received SNR for TX FFE model can be derived as the following

$$\begin{aligned} \text{SNR}_{TX} &= \frac{1}{\text{PAM} - 1} \frac{(\vec{h} * \vec{c})_0 / \|\vec{c}\|_1}{\sqrt{\alpha^2 \sigma_{ISI}^2 / \|\vec{c}\|_1^2 + \sigma_n^2}} \\ &= \frac{1}{\text{PAM} - 1} \frac{1}{\|\vec{c}\|_1} \frac{(\vec{h} * \vec{c})_0}{\sqrt{\alpha^2 \sigma_{ISI}^2 / \|\vec{c}\|_1^2 + \sigma_n^2}} \end{aligned} \quad (3.2)$$

The main cursor of the equivalent channel $(\vec{h} * \vec{c})_0$ is the main signal of interest. The $1/(\text{PAM} - 1)$ term models the signal reduction due to multi-level PAM's increased number of eyes. For example, the signal is reduced by a factor of 3 in PAM4 applications. The residual ISI scalar α is the data signal RMS strength. For example, $\alpha = 1$ for PAM2 and for PAM4 whose data is ± 1 and $\pm 1/3$, $\alpha \approx 0.745$.

A similar analysis can be done for the RX FFE model in Figure 3.4. The data signal is sent through the channel directly. The RX applies the same gain A to both the received signal and input noise. On the other hand, the FFE performs different functions on the signal and noise. Again, the FFE equalizes the signal identical to the TX FFE. The FFE will also amplify the incoming noise due to its high-pass nature.

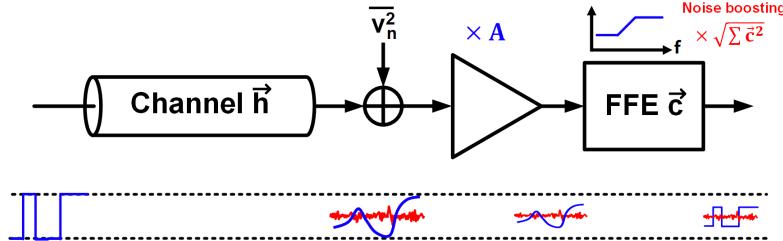


Figure 3.4: RX FFE operation principle

When a noise source of RMS value σ_n is filtered by a vector \vec{c} , the output noise power scaled by the root sum of the squares of the coefficients, also known as the L-2 norm $\|\vec{c}\|_2$ of vector \vec{c}

$$\|\vec{c}\|_1 = \sqrt{\sum_n (\vec{c}_n)^2} \quad (3.3)$$

The output SNR for the RX FFE model can then be expressed as

$$\begin{aligned} \text{SNR}_{RX} &= \frac{1}{\text{PAM} - 1} \frac{(\vec{h} * \vec{c})_0}{\sqrt{\alpha^2 \sigma_{ISI}^2 + \sigma_n^2 \|\vec{c}\|_2^2}} \\ &= \frac{1}{\text{PAM} - 1} \frac{1}{\|\vec{c}\|_2} \frac{(\vec{h} * \vec{c})_0}{\sqrt{\alpha^2 \sigma_{ISI}^2 / \|\vec{c}\|_2^2 + \sigma_n^2}} \end{aligned} \quad (3.4)$$

Assuming enough FFE taps are used and residual ISI at the decision-making point is not the dominant noise source in the system, the SNR expressions can be simplified and their ratio found to be

$$\begin{aligned}\frac{\text{SNR}_{RX}}{\text{SNR}_{TX}} &= \frac{1}{\text{PAM} - 1} \frac{1}{\|\vec{c}\|_2} \frac{(\vec{h} * \vec{c})_0}{\sigma_n} \Big/ \frac{1}{\text{PAM} - 1} \frac{1}{\|\vec{c}\|_1} \frac{(\vec{h} * \vec{c})_0}{\sigma_n} \\ &= \frac{\|\vec{c}\|_1}{\|\vec{c}\|_2} \geq 1\end{aligned}\quad (3.5)$$

It is well known that for any given vector \vec{c} , $\|\vec{c}\|_2 \leq \|\vec{c}\|_1$, therefore the system performance with RX FFE is at least as good as that with TX FFE.

To validate the inequality, we consider three different channels used as shown in Figure 3.5. The channel losses at the Nyquist frequency of interest (28GHz for 56Gbps PAM2 or 112Gbps PAM4) are approximately 16dB, 24dB, and 33dB, respectively. The same FFE coefficients are applied to both the TX and RX FFE, and they are calculated to cancel ISI at sampling phase of interest completely (zero forcing), given the number of pre- and post-cursor taps. The following describes the steps to calculate SNRs for a given channel and selected number of pre-cursor and post-cursor taps:

1. Compute the zero-forcing FFE coefficients.¹
2. Convolve the computed equalizer coefficients with DT channel pulse response to obtain the equalized pulse response.
3. Calculate residual ISI noise power and multiply by α .
4. Find the L1 and L2 norm of the FFE coefficients.
5. Use (3.2) and (3.4) to calculate SNRs for the systems with TX or RX FFE.

¹Minimum mean squared error (MMSE) algorithm are typically used for FFE adaptation, but this analysis still holds since the same coefficients are used for both TX and RX FFEs.

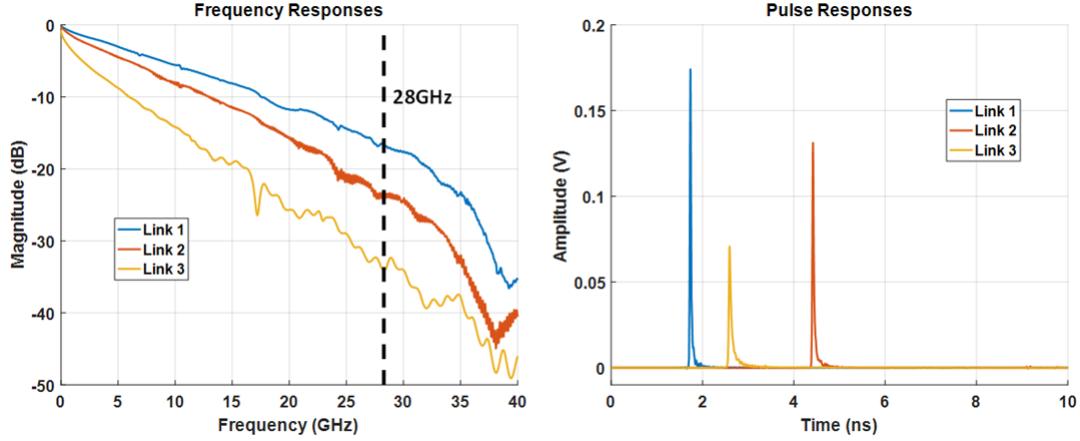


Figure 3.5: Three channels used for FFE location study

The RX input noise strength is swept from 0mV_{rms} to 5mV_{rms} . The maximum TX swing is held at $\pm 400\text{mV}$. The FFE lengths used for this analysis are 5 pre-cursor + 15 post-cursor taps, 10 pre-cursor + 20 post-cursor taps, and 15 pre-cursor + 25 post-cursor taps. Figure 3.6 shows the calculated system SNR with respect to input noise for all three channels and FFE lengths considered. For any channel and FFE

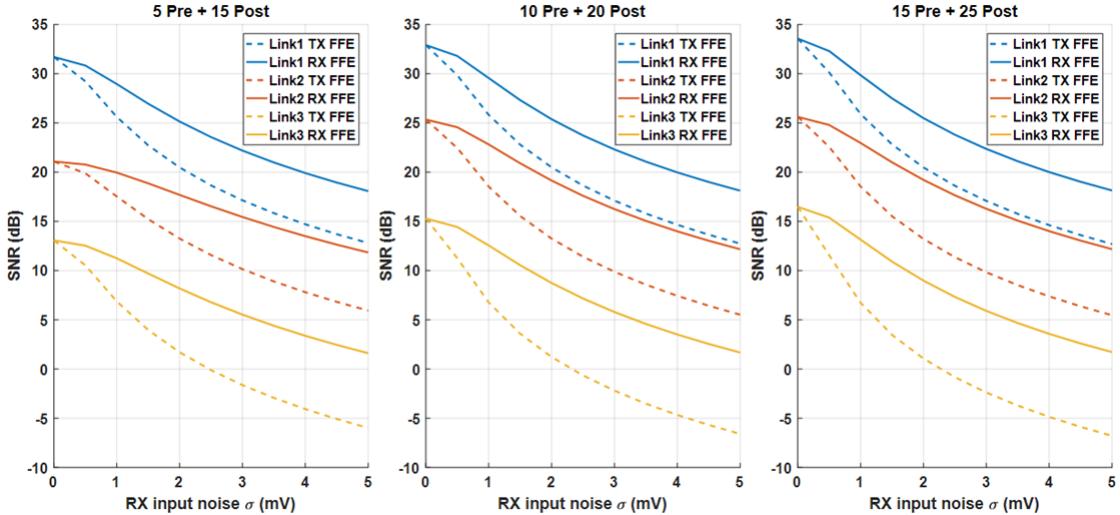


Figure 3.6: Total SNR vs. RX input noise for three channels and three FFE length settings

length, we see that the RX FFE system performs as well as the TX FFE system. Due to the linear nature of the systems under study, FFE has the same effect on the system regardless of its location when there is no RX input noise ($\sigma_n = 0\text{mV}$). However, a large discrepancy starts to manifest when RX input noise is considered. TX FFE performances roll off much faster than that of the RX FFE. The system SNR difference can be more than 6dB when σ is larger than 2mV for the worst-case channel when comparing TX and RX FFEs. For channels with larger losses (link 2 and link 3), increasing the FFE lengths improves performance noticeably. However, there is no significant difference between the 30-tap and 40-tap settings. Link 3 is a particularly difficult channel to operate in the presence of RX input thermal noise. As a result, TX FFE can hardly equalize link 3.

Behavioral transient simulations are also used to verify the system SNR equations. A PRBS13 data pattern is used for acceptable simulation time and sufficient data activities. Figure 3.7 shows an example of sampled eye diagrams for both TX and RX FFE systems. We can visually conclude that the RX FFE has a much higher system SNR given the error spread around the data levels.

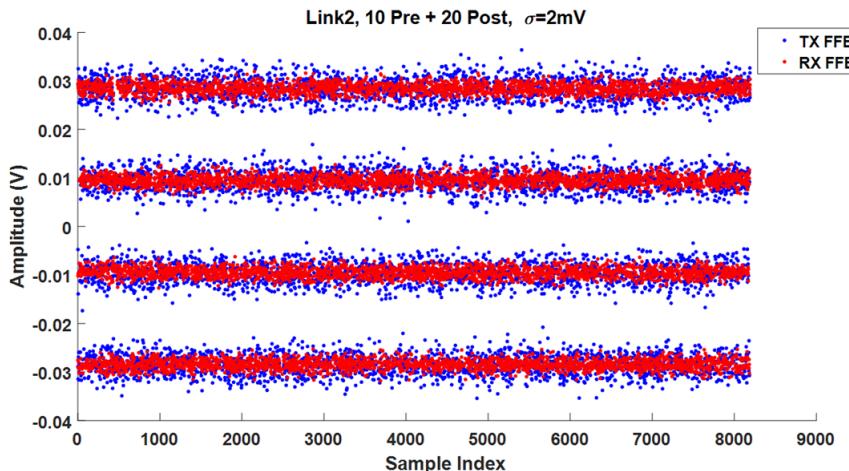


Figure 3.7: Sampled eye diagram of link 2 RX output using 10 pre-cursor and 20 post-cursor FFE

3.1.2 Practical concerns for TX and RX FFE

From the analysis above, RX FFE fundamentally outperforms TX FFE when input noise is present. However, there are many more practical circuit issues that complicates the simple model used thus far. In this section, we will use the statistical framework developed in the previous chapter to discuss the advantages and disadvantages of TX and RX FFE.

Table 3.1 summarizes the various benefits and challenges for TX and RX FFE respectively in a PAM4 link. Even though the FFE coefficients might have their own resolution requirements, TX FFE has the advantage of processing reduced number of data bits. The RX FFE needs to handle the ADC's B-bit output data, thus leading to increased digital complexity. However, state of the art transmitters such as [19, 20, 21, 22, 23] all use voltage/current mode drivers with FFE implemented at the output summing node. This significantly limits the number of taps for TX FFE. The speed bottleneck at the critical TX output node also restricts FFE coefficient tuning, while RX FFE can take advantage of pipelining in digital domain to tradeoff speed, FFE length and coefficient tuning with latency. Transmitters using digital to analog converters (DAC) can also enjoy longer FFE lengths, as shown in [24]. However, it has become more challenging to build an area and power efficient DAC and further research is needed.

	TX FFE	RX FFE
Resolution	2-bit data + N-bit weights	B-bit data + N-bit weights
Implementation	Voltage/current mode	Pipelined digital
FFE length	Limited	Long
Adaptivity	Back channel required	Fully adaptive
Dynamic range	Peak power constraint	Linearity limited
Circuit noise	Insignificant	Needs attention

Table 3.1: Qualitatively comparison for TX vs. RX FFE practical concerns

One of the biggest advantages for RX FFE is its adaptability. Typically the

adaptation engine is on the RX side since it has the post-channel signal and noise information. Therefore, it is natural for RX FFE to be adapted and have more robust performance over process, voltage and temperature (PVT) variations. It is more costly to adapt a TX FFE since a back channel is required to send back the desired coefficients by the RX. As we will demonstrate in later chapters, ADC-based links are almost insensitive to TX FFE settings when the subsequent RX FFE is adapted.

Although we have shown from previous analysis that the TX FFE's peak power constraint becomes the bottleneck in terms of system's SNR performance when RX input noise dominates, the RX front-end's own linearity limit and noise before the FFE in realistic circuits cannot be ignored. A more accurate model is depicted in Figure 3.8, in which nonlinearity and the RX front-end noise are included. Comparing to the TX FFE model now, the RX FFE system will have more error sources that will close the SNR gap shown before. Nevertheless, within the statistical framework that was developed, better equalization partitioning between TX and RX FFE can be realized. The RX circuit noise σ_s , which follows the gain block, will also be boosted

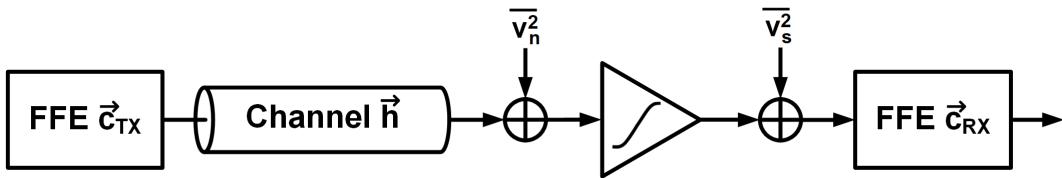


Figure 3.8: TX+RX FFE including nonlinearity and circuit noise

by RX FFE \vec{c}_{RX} . For this model, we still use SNR as a proxy to the final system performance to gain valuable insight regarding partitioning equalization between \vec{c}_{TX} and \vec{c}_{RX} . The following equation describes the final output SNR

$$\text{SNR} = \frac{1}{\text{PAM} - 1} \cdot \frac{1}{\|\vec{c}_{TX}\|_1 \|\vec{c}_{RX}\|_2} \cdot \frac{A(\vec{h} * \vec{c}_{TX} * \vec{c}_{RX})_0}{\sqrt{A^2 \sigma_n^2 + \sigma_{NL}^2 + \sigma_s^2}} \quad (3.6)$$

Even though this thesis does not focus on more detailed analysis on optimal partitioning of TX and RX equalization, intuitive conclusions can still be drawn from this expression that will help with the design of a power efficient link. Specifically, since the overall FFE transfer function stays constant (i.e., $\vec{c}_{TX} * \vec{c}_{RX} = \vec{c}$), the scalar term $1/(\|\vec{c}_{TX}\|_1 \|\vec{c}_{RX}\|_2)$ provides an interesting tradeoff between \vec{c}_{TX} and \vec{c}_{RX} . While TX FFE reduces signal directly, the benefit is that the RX FFE will have less boosting on both the RX input noise and circuit noise. Furthermore, the gain A should ideally be as large as possible, but the maximum allowed gain is limited by the full scale range and linearity limit of the circuit. We have already discussed in Section 2.3.2 that the nonlinearity error σ_{NL} strongly depends on the input PDF, which is shaped by the TX FFE. Therefore, there is another tradeoff among A , \vec{c}_{TX} and σ_{NL} .

The discussions have not explicitly included the ADC so far, but a framework is already in place for further analysis. The next section will expand on this model and bring the ADC into the discussion and emphasize the necessity of pre-equalization before the ADC.

3.2 Equalization before vs. after ADC

3.2.1 Design equation for ADC resolution

When an ADC is included, the model in Figure 3.8 does not change too much if only quantization error is considered. The circuit noise \bar{v}_s^2 simply becomes quantization noise \bar{v}_q^2 as shown in Figure 3.9. For simplicity, TX FFE is eliminated for now and only RX FFE is considered. We will also only focus on quantization noise for this section.

In order to fully exercise the ADC's FSR without clipping, the front end gain should be set to the ratio between $FSR/2$ and the channel's worst case output. The

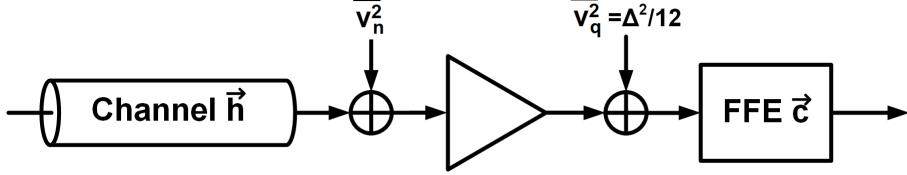


Figure 3.9: RX FFE model with ADC quantization noise

channel's maximum output is determined by the sum of the channel vector coefficients' absolute values, therefore its L1-norm. The gain A required is then

$$A = \frac{FSR}{2 \|\vec{h}\|_1} \quad (3.7)$$

Again, the ADC's LSB size Δ is determined by the FSR and resolution B . Given these parameters, we are able to use SQNR as another proxy to see how the ADC requirement relates to the rest of the system. The SQNR in this context is

$$\begin{aligned}
 \text{SQNR} &= \frac{1}{\text{PAM} - 1} \frac{1}{\|\vec{c}\|_2} \frac{A(\vec{h} * \vec{c})_0}{\sqrt{\Delta^2/12}} \\
 &= \frac{1}{\text{PAM} - 1} \left(\frac{FSR}{2 \|\vec{h}\|_1} (\vec{h} * \vec{c})_0 \Big/ \frac{\|\vec{c}\|_2 FSR}{2\sqrt{3} 2^B} \right) \\
 &= \frac{1}{\text{PAM} - 1} \cdot 2^B \cdot \frac{\sqrt{3}(\vec{h} * \vec{c})_0}{\|\vec{h}\|_1 \|\vec{c}\|_2} \\
 \text{SQNR} &= 2^B \cdot \frac{1}{\text{PAM} - 1} \cdot \frac{\vec{h}_0}{\|\vec{h}\|_1} \cdot F(\vec{c}) \quad (3.8)
 \end{aligned}$$

It is worth noting that the system's SQNR is determined by the ADC resolution B , the specific modulation used, and an extra term that's dependent on both channel \vec{h} and FFE \vec{c} . Although the FFE's tap values also depend on the channel, the equation is rewritten to separate \vec{h} and \vec{c} . The portion that depends on \vec{h} is expanded explicitly

and we use a function $F(\vec{c})$ to represent the effects due to FFE. We also recognize that such an expression is defined as the SNR for a single received eye since it is the ratio between eye height and noise power. By equating SNR_{eye} with the expression above and taking \log_2 on both sides, the following equation for ADC resolution is obtained

$$B = \log_2(\text{SNR}_{eye}) + \log_2(\text{PAM} - 1) + \log_2 \left(\frac{\|\vec{h}\|_1}{\vec{h}_0} \cdot \frac{1}{F(\vec{c})} \right) \quad (3.9)$$

$$B = B_{eye} + B_{PAM} + B_{channel} \quad (3.10)$$

This is an important result since we have just derived a first-order equation for ADC resolution requirement. A more accurate expression is derived in Appendix C. The ADC resolution is shown to have three major components and Figure 3.10 illustrates how each component contributes to the final ADC resolution requirement. First,

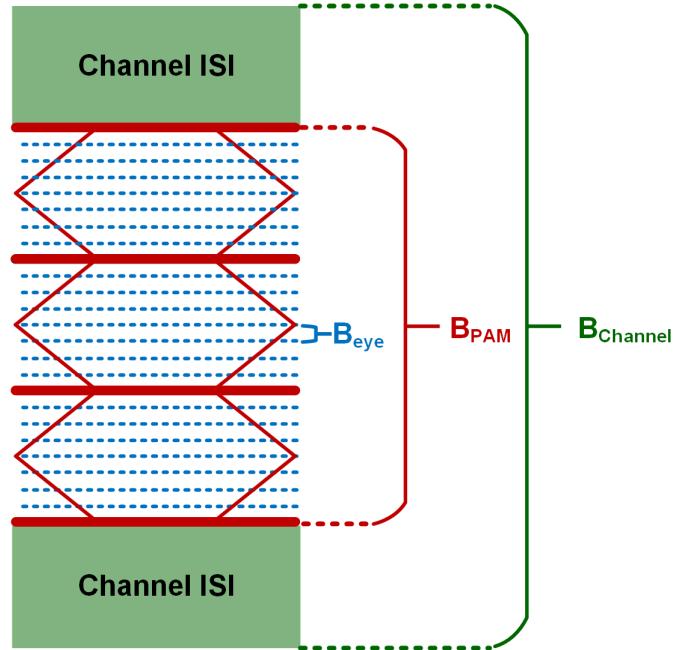


Figure 3.10: ADC resolution requirement components

there is a base resolution requirement per eye determined by B_{eye} . This component was discussed extensively in the previous chapter when the statistical framework was developed. Depending on the PAM scheme used for the system, there is a fundamental increase in resolution simply due to the increased number of eyes. Lastly, there will be extra channel ISI that “hide” the real eyes and increase the total signal dynamic range, thus increasing the number of quantization levels necessary to capture the whole signal range.

3.2.2 Channel PMR, FFE and ADC resolution

Let’s now investigate the $B_{channel}$ term in (3.10) further to understand how the channel affects the ADC requirement. The term that involves the channel, $\|\vec{h}\|_1 / \vec{h}_0$ is the ratio of the worst channel output and main cursor value. This is defined as the channel’s peak to main ratio (PMR) and studied extensively in [25]. There is also an eye opening equivalence without other noise present shown in [25]. Here we extend this equivalence to incorporate higher modulations, described by the equation below

$$\text{eye opening} = 2\vec{h}_0 \left(\frac{\text{PAM}}{\text{PAM} - 1} - \text{PMR} \right) \quad (3.11)$$

For an ideal channel, PMR equals to 1 and the equation above reduces to $2\vec{h}_0 / (\text{PAM} - 1)$, which is the expected eye height for an ideal system. For higher PAM schemes, the minimum PMR required to maintain an open eye becomes smaller (e.g. PMR needs to be < 2 for PAM2 but $< 4/3$ for PAM4 to reach an open eye). In order to reduce $B_{channel}$, the PMR of the channel before the ADC needs to decrease; in other words, the eye needs to be as ”open” as possible. Therefore, it is crucial to make sure that the pre-ADC channel’s PMR is well controlled by equalization so that the ADC requirements can be relaxed. Figure 3.11 illustrates how a smaller PMR can lead to ADC resolution reduction. The two channels’ main cursors are both normalized to 1,

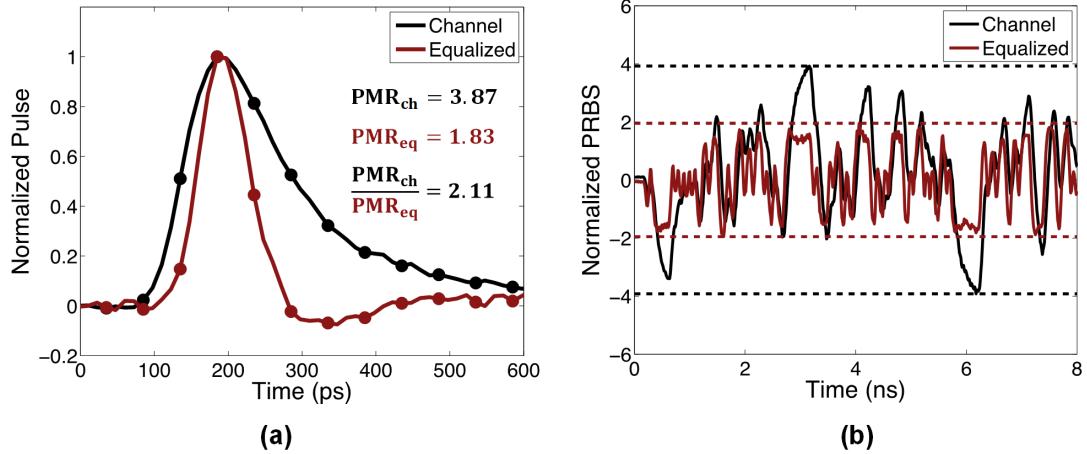


Figure 3.11: (a) Normalized channels with 2x difference in PMR and (b) corresponding transient waveforms for PAM2 signaling [25]

and their respective PMRs are also shown in Figure 3.11(a). The equalized channel has a 2x smaller PMR than the original channel, which leads to a 2x reduction in the transient waveform's full scale range in Figure 3.11(b). As a result, only half of the ADC quantization levels are needed for the equalized channel while capturing the same main cursor signal, which means the equalized channel will require one fewer bit in ADC resolution for a 2x PMR reduction.

In addition, the FFE \vec{c} also affects B_{channel} . Even though we have lumped all its effects into a function $F(\vec{c})$, the concept of FFE noise boosting has already been explained in Section 3.1. Figure 3.12 shows how quantization noise propagates through an FFE with coefficient vector \vec{c} . The assumption that each quantization noise sample is independent from others still holds true in this case, and thus the quantization PDF simply propagates through the delay line in the FFE. The FFE coefficients scale the quantization PDF bounds to be $\pm|c_n|\Delta/2$. Finally, the independent nature of the delayed quantization samples leads to an output PDF from the convolution of all the scaled uniform PDFs.

Several properties of the output PDF need to be highlighted:

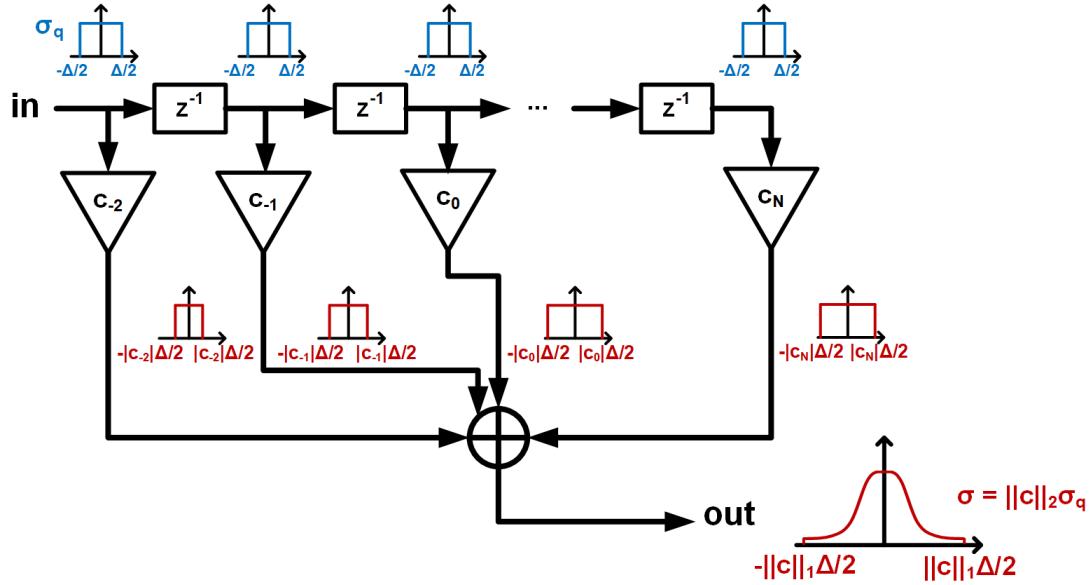


Figure 3.12: Quantization noise PDF's propagation through an FFE

1. The central limit theorem establishes that when independent random variables add, their sum tends toward a normal distribution. In this case, the output PDF indeed starts to become more Gaussian, however still bounded due to the limited number of FFE taps.
2. The output PDF is bounded by $\pm\Delta/2$ scaled by the L-1 norm of FFE coefficients, $\|\vec{c}\|_1$.
3. The output PDF's standard deviation is σ_q scaled by the L-2 norm of FFE coefficients, $\|\vec{c}\|_2$.

To alleviate the FFE's noise boosting effect, pre-equalization can also help reduce the necessary FFE coefficients and filter length. Therefore, both the \vec{h} and \vec{c} dependent terms in the $B_{channel}$ ADC resolution component call for a pre-equalizer before the ADC. Such a pre-equalizer will not only directly reduce the channel's PMR by canceling ISI, it will also reduce the FFE strength needed in the digital domain, thus

reduce implementation complexity and power at the same time. Table 3.2 shows example ADC resolutions and the corresponding ADC components for different links with different BER specifications and good/bad channels. A BER of 10^{-9} is used as the approximate cut-off for low or high error rate in this case. B_{eye} directly affects BER performance and 0.6 bits are added to the 2 or 3 bits base considering ± 0.5 LSB DNL. $B_{channel} = 0.6$ bits is a reasonable number for a good channel and 1.6 bits for a worse channel. Given the desired resolutions, possible ADC types are also shown in the table.

PAM	BER spec	Channel	B_{PAM}	B_{eye}	$B_{channel}$	Round(B)	ADC type
2	$< 10^{-9}$	Good	0	3.6	0.6	4	Flash
		Bad	0	3.6	1.6	5	Flash
	$> 10^{-9}$	Good	0	2.6	0.6	3	Flash
		Bad	0	2.6	1.6	4	Flash
4	$< 10^{-9}$	Good	1.6	3.6	0.6	6	Flash/SAR
		Bad	1.6	3.6	1.6	7	SAR
	$> 10^{-9}$	Good	1.6	2.6	0.6	5	Flash
		Bad	1.6	2.6	1.6	6	Flash/SAR
8	$< 10^{-9}$	Good	2.8	3.6	0.6	7	SAR
		Bad	2.8	3.6	1.6	8	SAR
	$> 10^{-9}$	Good	2.8	2.6	0.6	6	Flash/SAR
		Bad	2.8	2.6	1.6	7	SAR

Table 3.2: Example ADC resolutions for different link scenarios

To summarize, ADC-based links have enabled the usage of fully adaptive RX FFE in the digital domain, which could perform just as well as TX FFE. However, due to RX front-end's nonlinearity and RX FFE's implementation concerns, TX FFEs could still be effective in conditioning the RX input signal to relax these requirements. We have discussed the various impacts of equalizers before and after the ADC. From a first-order ADC resolution requirement equation, pre-equalization before the ADC is of vital importance since it reduces the channel's PMR similar to TX FFE (without the peak power constraint) and lessens FFE noise boosting because of reduced FFE

coefficients. Figure 3.13 shows the final architecture and what each block's role is from our newly gained perspectives. The major function of the pre-equalizers is not to fully equalize the channel, but to shape the equivalent channel to have a well-controlled PMR to relax both the ADC and DSP design specifications. The next part of this work will focus on power and area efficient circuit implementations of such critical pre-equalizers, system performance and trade-offs.

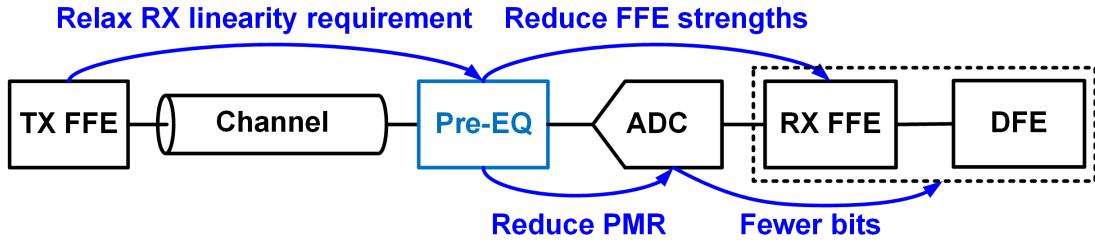


Figure 3.13: Each block's role in an ADC-based link with pre-equalization

Chapter 4

Inverter-based Pre-ADC Equalizers

In the previous chapter, we have established the importance of pre-ADC equalizers in ADC-based links. Recent works such as [19, 26, 27] started to put more emphasis on the front-end by adding equalization before the ADC. However, there was no further analysis to show the tradeoff between pre-equalization and ADC resolution. Furthermore, conventional current-mode-logic (CML) based CTLEs (shown in Figure 4.1(a)) are still the preferred choice in such systems, but they have become increasingly challenging to build as data rate continues to scale. CML-based CTLEs use source degenerated differential pair to achieve high frequency peaking, but the stringent bandwidth requirements made passive inductor/T-coil loads a must. Even though transistor technology scaling brought faster devices, CML-based CTLEs are not benefiting from the smaller devices because of the area overhead from the load elements. As a result, it has become very difficult for CML-based CTLEs to be power and area efficient at this speed.

On the other hand, inverter-based filters (shown in Figure 4.1(b)) have shown potential for area reduction while maintaining low power. By using inverters, such

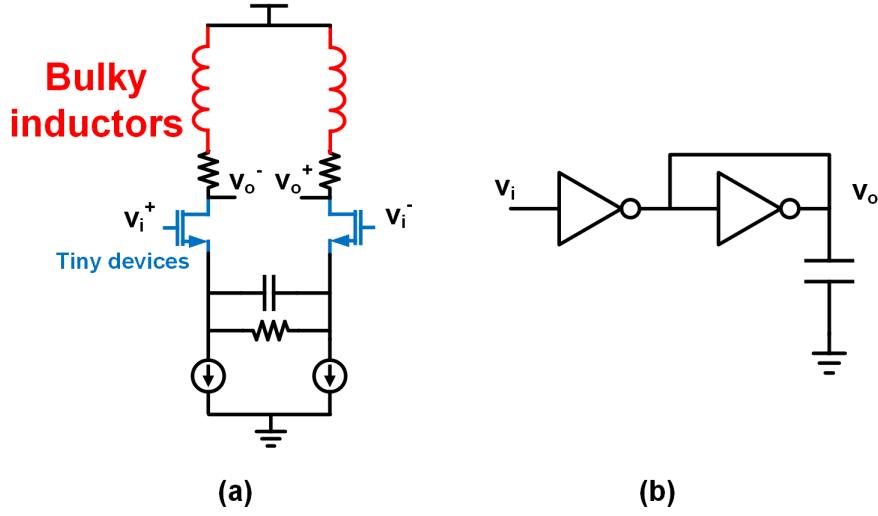


Figure 4.1: (a) CML-based CTLE and (b) inverter-based filter

circuits take full advantage of technology scaling similar to digital circuits. Inverter-based equalizers such as [28] already serve as examples for the area and power efficient nature of this new class of circuitry. In this chapter, we demonstrate two inverter-based CTLEs fully embedded in two transceiver testbeds instead of being standalone equalizers. Analog building blocks by using inverters are discussed in Section 4.1. Sections 4.2 and 4.3 present two different inverter-based CTLEs in 16nm FinFET CMOS for a PAM2 and PAM4 applications. Inverter biasing and linearity performance will also be addressed. Full transceiver performance will be presented as well.

4.1 Inverter as analog elements

4.1.1 Inverter transconductor and diode load

Inverters are typically associated with digital design, and their analog characteristics need to be understood thoroughly first before building inverter-based CTLEs. Inverters as analog elements were extensively studied in [25, 29]. Figure 4.2 shows the two

most basic inverter building blocks. When an inverter is biased appropriately such that both PMOS and NMOS devices are in saturation, it behaves as a transconductor (G_m) cell [29] as depicted in Figure 4.2(a). In advanced processes such as 16nm

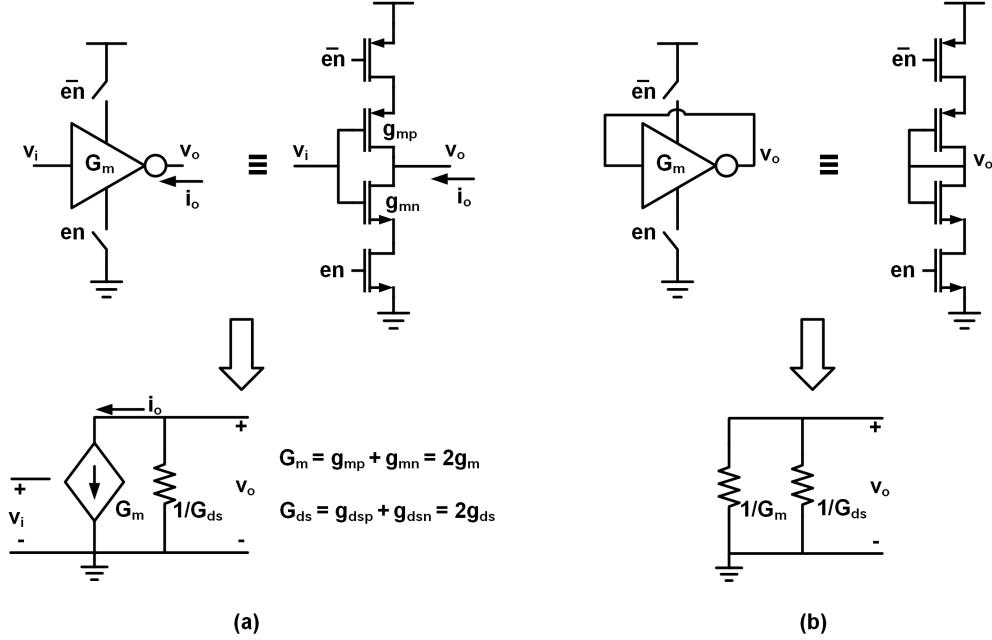


Figure 4.2: (a) Inverter transconductor (b) Inverter resistive load

FinFET CMOS, NMOS and PMOS devices have nearly the same mobility, and thus using PMOS doesn't incur a speed penalty, leading to designs that are truly limited by the technology's peak f_T . This also allows a fully symmetric design and layout for each inverter unit. Enable switches are also added and we assume the on-resistances of the enable switches are small enough to ignore their effects for subsequent analysis. The total transconductance of an inverter G_m is then the sum of PMOS and NMOS' respective g_m , which simply doubles a single device's g_m assuming symmetry. Each device will also have a small signal output conductance, g_{ds} , which limits the intrinsic gain of the transistor. However, g_m/g_{ds} for FinFET devices are usually >10 , which means that g_{ds} can be ignored for most design. A diode connected inverter in Figure 4.2(b) behaves as a self-biased $1/g_m$ resistive load. When the NMOS and

PMOS devices drive strengths are identical, the nominal output voltage is half way between supply and ground. This natural bias point puts all devices in saturation, thus ensuring sufficient output swing. Again, g_{ds} of the load is negligible compared to g_m .

Figure 4.3 shows various configurations and analog blocks that can be realized with these two basic elements. By tuning the active g_m element digitally through the enable switches, a unity gain buffer becomes a programmable gain amplifier. Any pole can be implemented by loading a buffer with a capacitor and its location is determined by G_m/C . Summation is achieved by adding the g_m cells' current outputs.

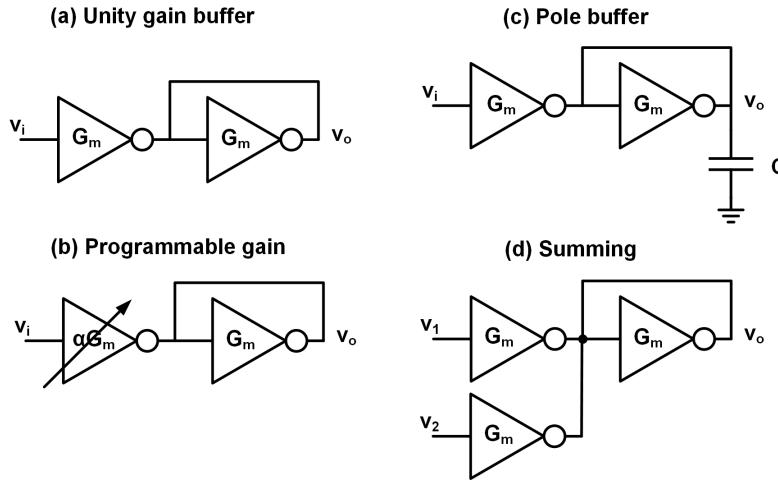


Figure 4.3: Various inverter configurations for different analog blocks

4.1.2 Inverter active inductor

Inverter-based active inductors, shown in Figure 4.4, are used to avoid more area-consuming passive inductors. Appendix D derives the small-signal impedance of this load to be

$$Z_L = \frac{1}{G_m} \frac{1 + sRC_{gs}}{1 + sC_{gs}/G_m} = \frac{1}{G_m} \frac{1 + s/\omega_z}{1 + s/\omega_T} \quad (4.1)$$

where ω_T is the transit frequency of an inverter, given by G_m/C_{gs} . The load introduces a zero ω_z determined by feedback resistance R and inverter total gate capacitance C_{gs} . For frequency range $\omega \ll \omega_T$ this impedance can be simplified to

$$Z_L \approx \frac{1}{G_m}(1 + s/\omega_z) = \frac{1}{G_m} + s \frac{R}{\omega_T} \quad (4.2)$$

which is a resistance of $1/G_m$ and an inductance R/ω_T in series, also shown in Figure 4.4.

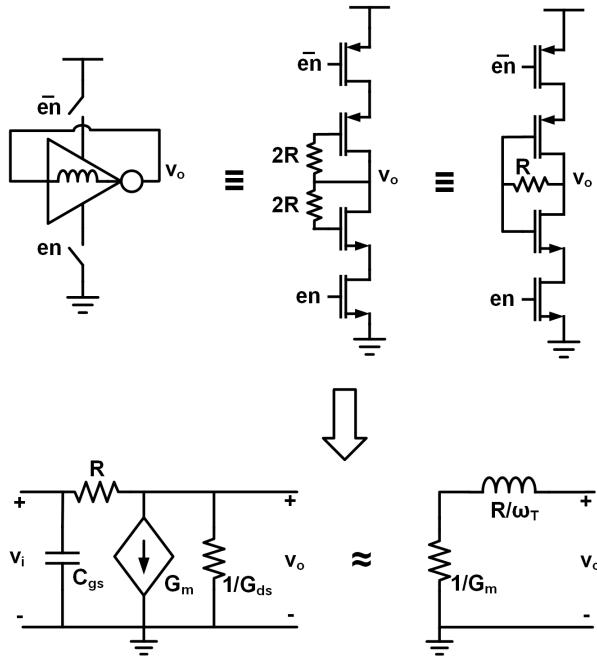


Figure 4.4: Inverter active inductor

The feedback resistor essentially isolates the inverter gate capacitance from the critical output node compared to a diode connected inverter. For the special case $R = 1/G_m$, the pole and zero in (4.1) cancels out and the load impedance becomes a very high bandwidth resistor of value $1/G_m$. Figure 4.5 shows the comparison between two buffers using diode and active inductor load. Since a diode connection shorts an inverter's gate and drain node, the buffer is self-loaded with its own C_{gs} in

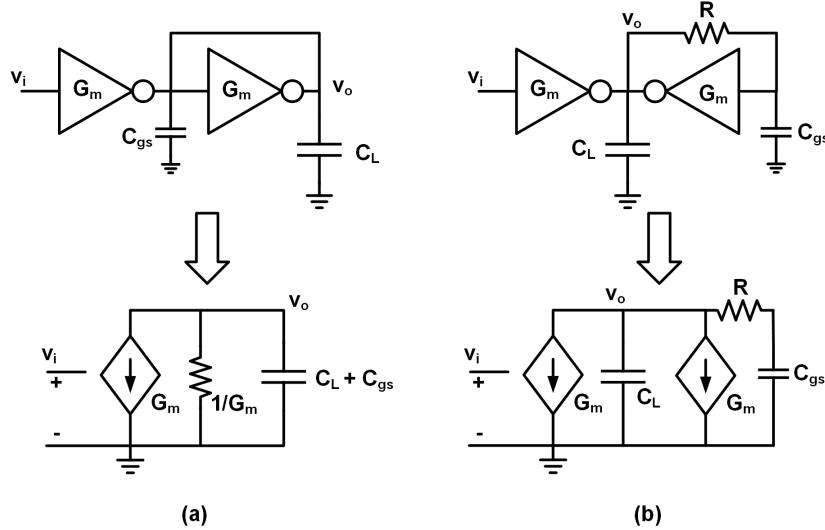


Figure 4.5: Unity gain buffer with (a)diode connected load and (b)active inductor load

addition to the explicit load C_L . Thus, the diode load buffer has a transfer function of

$$G_{diode}(s) = \frac{1}{1 + s \frac{C_L + C_{gs}}{G_m}} \quad (4.3)$$

When $C_L = C_{gs}$ (i.e. fan-out of 1 buffering), the transfer function above can be reduced to

$$G_{diode}(s) = \frac{1}{1 + 2s/\omega_T} \quad (4.4)$$

On the other hand, the gain transfer function of an active inductor load buffer (derived in Appendix E) is

$$G_{ind}(s) = \frac{1 + sRC_{gs}}{1 + s\frac{C_L + C_{gs}}{G_m} + s^2\frac{RC_L C_{gs}}{G_m}} \quad (4.5)$$

It is important to note that C_{gs} effectively disappears when $R = 1/G_m$, and for the

case $C_L = C_{gs}$, $G_{ind}(s)$ becomes

$$\begin{aligned}
 G_{ind}(s) &= \frac{1 + s \frac{C_{gs}}{G_m}}{1 + s \frac{C_L + C_{gs}}{G_m} + s^2 \frac{C_L C_{gs}}{G_m^2}} \\
 &= \frac{1 + s \frac{C_{gs}}{G_m}}{(1 + s \frac{C_{gs}}{G_m})^2} \\
 &= \frac{1}{1 + s/\omega_T}
 \end{aligned} \tag{4.6}$$

We observe that using an active inductor for the given parameters doubles buffer bandwidth. In reality, the bandwidth extension is less than 2x due to the presence of drain capacitance and other layout parasitics, but still can be compensated for by increasing feedback resistance R .

However, the noise aspect of active inductor loads need to addressed to ensure the penalty is a reasonable trade-off for the increased bandwidth. Similar to the gain transfer functions analysis, we start by analyzing the diode connected buffer noise shown in Figure 4.6(a). Each inverter's noise power spectral density (PSD) is given

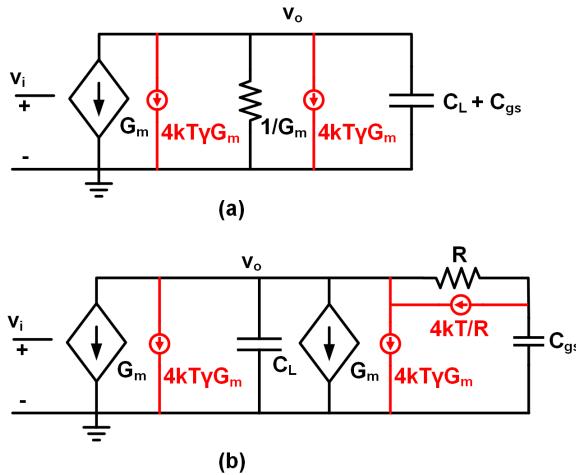


Figure 4.6: Unity gain buffer noise models with (a)diode connected load and (b)active inductor load

by $4kT\gamma G_m$ in which γ is the devices' noise factor (assuming symmetry and equal gammas for PMOS and NMOS devices). For a fan-out of one diode loaded buffer, the output voltage noise PSD is

$$\overline{\frac{v_{n,diode}^2}{\Delta f}} = 8kT\gamma G_m \left| \frac{1}{G_m} \frac{1}{1 + 2s/\omega_T} \right|^2 \quad (4.7)$$

A reasonable noise performance metric in link applications is to look at the output's total integrated noise since the subsequent ADC samples the incoming signal. For our buffer of interest, the total integrated noise variance is then

$$\begin{aligned} \overline{v_{n,diode}^2} &= \int_0^\infty 8kT\gamma G_m \left| \frac{1}{G_m} \frac{1}{1 + j2\pi f \times 2/\omega_T} \right|^2 df \\ &= \frac{8kT\gamma}{4 \times 2C_{gs}} \\ &= \gamma \frac{kT}{C_{gs}} \end{aligned} \quad (4.8)$$

The full expression for a active inductor loaded buffer's output noise PSD and total integrated noise are derived in Appendix F as follows

$$\begin{aligned} \overline{\frac{v_{n,ind}^2}{\Delta f}} &= 8kT\gamma G_m \left| \frac{1}{G_m} \frac{1 + sRC_{gs}}{1 + s\frac{C_L + C_{gs}}{G_m} + s^2\frac{RC_L C_{gs}}{G_m}} \right|^2 \\ &\quad + \frac{4kT}{R} \left| R \frac{1 + sC_{gs}/G_m}{1 + s\frac{C_L + C_{gs}}{G_m} + s^2\frac{RC_L C_{gs}}{G_m}} \right|^2 \end{aligned} \quad (4.9)$$

$$\overline{v_{n,ind}^2} = 2kT\gamma \frac{1 + G_m RC_{gs}/C_L}{C_L + C_{gs}} + kT \frac{G_m R + C_{gs}/C_L}{C_L + C_{gs}} \quad (4.10)$$

Again, if $G_m R = 1$ and $C_L = C_{gs}$, the expression above simplifies to

$$\overline{v_{n,ind}^2} = (2\gamma + 1) \frac{kT}{C_{gs}} \quad (4.11)$$

The ratio of the two cases then becomes

$$\frac{\overline{v_{n,ind}^2}}{\overline{v_{n,diode}^2}} = \frac{2\gamma + 1}{\gamma} = 2 + \frac{1}{\gamma} \quad (4.12)$$

By using active inductors, the bandwidth has already increased by two fold, thus noise variance intrinsically doubles. The extra noise penalty comes from the resistor's noise, and the percentage increase depends on device noise gamma. For modern finFET technologies, γ is approximately 2, which means the example buffer with active inductor load will have 1.58x more RMS noise (compared to 1.41x, which is 1dB less) for twice the bandwidth. This amount of extra noise is manageable in typical link applications where speed and area are the bottlenecks. Nevertheless, the choice of R will then be a trade-off between bandwidth and noise requirement.

4.1.3 Inverter linearity

Even though previous chapters showed that a better conditioned input PDF can alleviate linearity requirements, the pre-equalizer still needs acceptable linearity since it is the first block on the RX side. The inverter-based unity gain buffer's nonlinearity was also studied in [25], but discussions were limited. Here, we present a simple analysis to gain insight into whether inverters can provide enough linearity for link applications, especially for PAM4. For subsequent discussions, we will focus mainly on static nonlinearity.

Qualitatively, inverter-based circuits already have first-order linearization due to its ratiometric design nature. Having PMOS and NMOS devices with equal mobility

means any nonlinear behavior in the transconductance element can be canceled by the nonlinear load element. Since there is no constant current bias, inverter-based circuits have class-AB operation in which there is no slew rate limit for better large signal linearity. It can be shown that for any symmetric inverter design with high output impedance (see Appendix G), the DC relation between a unity gain buffer's large signal output voltage and small signal input is

$$V_o = \frac{V_{DD}}{2} - v_i \quad (4.13)$$

independent of the I-V characteristic of the transistors as long as they are saturated. V_{DD} is the inverter's supply and the operating point is exactly $V_{DD}/2$ by symmetry. In other words, the inverter buffer is perfectly linear under these assumptions when devices are in saturation. Therefore, to understand its linearity limit, we need to find the maximum allow input such that a device gets push out of saturation.

Figure 4.7 shows an inverter buffer's large signal biasing point. The highlighted NMOS device's V_{DS} is the same as the output voltage, $V_{DD}/2 - v_i$. The NMOS' overdrive voltage, also V_{dsat} , is given by $V_{DD}/2 + v_i - V_T$. In order to make sure the

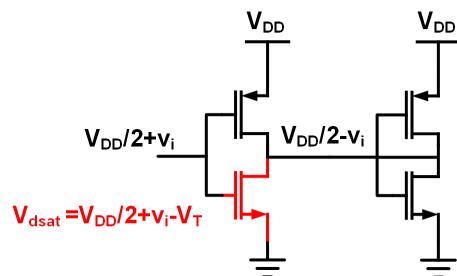


Figure 4.7: Unity gain buffer large signal bias point

device stays in saturation, the $V_{DS} > V_{dsat}$ condition has to be met, with which we

obtain

$$\begin{aligned} \frac{V_{DD}}{2} - v_i &> \frac{V_{DD}}{2} + v_i - V_T \\ v_i &< \frac{V_T}{2} \end{aligned} \quad (4.14)$$

The resultant expression gives us a rough idea how much input an inverter buffer can tolerate before becoming nonlinear. For example, $V_T \approx 0.4V$ for standard V_T (SVT) devices, which means a single-ended inverter buffer can handle about $\pm 0.2V$ of small signal input. A pseudo-differential buffer with two identical half circuits can then handle a maximum allowed differential input of $\pm 0.4V$. In reality, this number will be slightly lower due to device's finite output impedance. Nevertheless, such a voltage range is acceptable for most PAM4 link applications.

4.1.4 Inverter biasing voltage

We established that the V_T of the devices determine inverter-based circuits' linearity, which implies that it also determines the necessary supply voltage V_{DD} . If square law devices are used for analysis again, the required supply V_{DD} is related to each device's transconductance efficiency g_m/I_D by the following

$$\frac{g_m}{I_D} = \frac{2}{V_{ov}} = \frac{2}{V_{DD}/2 - V_T} \quad (4.15)$$

$$V_{DD} = 2 \left(V_T + \frac{2I_D}{g_m} \right) \quad (4.16)$$

Therefore, the V_{DD} necessary to appropriately bias the inverters can be directly calculated through these design equations. For example, if the target g_m/I_D is 10S/A for SVT devices in a PAM4 application, then the required V_{DD} should be around 1.2V. In a PAM2 link where linearity requirement isn't as stringent, only 0.7V is needed

for V_{DD} if ultra-low V_T (ULVT) devices are used ($V_T = 0.15V$).

4.1.5 Inverter cell layout

Layout has become a crucial aspect of modern high-speed circuit design. For the current target data rates, layout parasitics play a significant role in the circuits' final performance. Inverter-based circuits naturally yield layouts with small areas, and the use of active inductors further mitigates the parasitics challenges associated with passive inductors. Nevertheless, it is still important to create unit inverter cell layouts that are low in parasitics and are reliable.

Figure 4.8 shows example diagrams for unit inverter cells. A half-strength inverter can be built by stacking devices in series. By adding dummy transistors, the layout resembles a standard cell style and enables source/drain region sharing. Dummies can also help reduce self heating effects, which is a well-known concern in FinFET technologies [30]. Depending on the thermal and speed requirements, two 2x cell layouts can be realized. The type 1 layout uses two 1x cells and have dummies facing outwards. The type 2 layout gets rid of the dummies and reduces the width of the cell so that the output node travels smaller distance on the top metal layers for higher speed.

Example usages of different layout styles are shown in Figure 4.9. When there are more dummies in a larger inverter cell layout, the self-heating effect gets mitigated, but more parasitics are incurred. Both layout styles are used in the CTLE/AFE presented in this work to achieve a good speed and reliability trade-off. In addition, more dummies can be added to extend the source/drain region once all active inverters are abutted in order to reduce other layout effects such as well proximity and systematic mismatches.

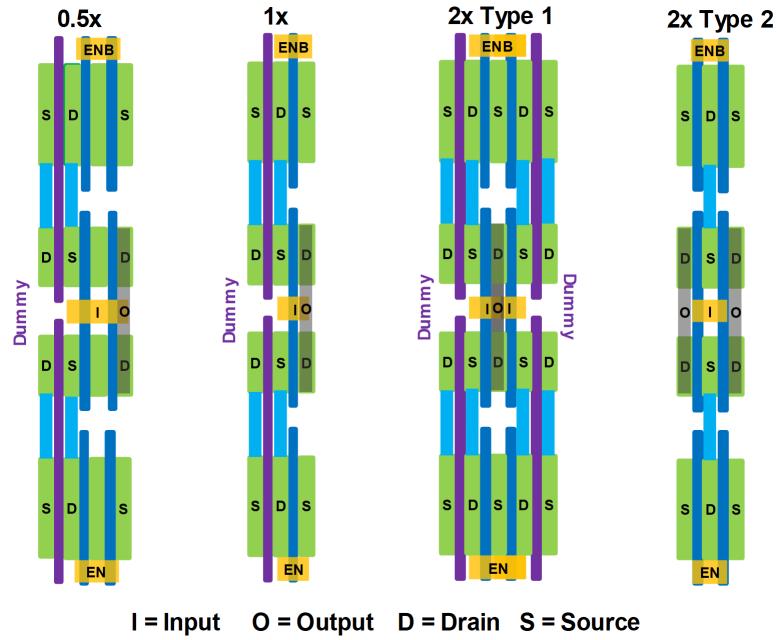


Figure 4.8: Example layout diagrams for unit inverter cells

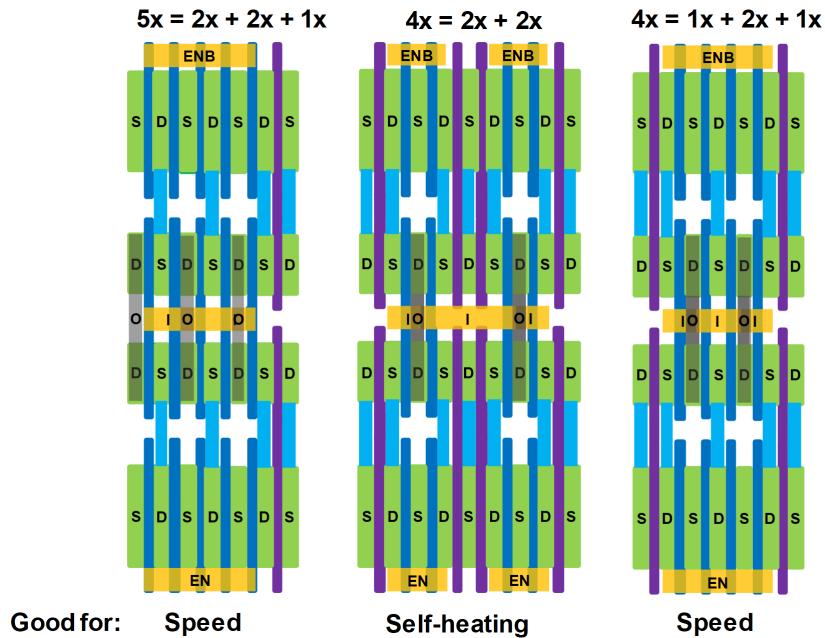


Figure 4.9: Example layout styles for abutted inverter cells

4.2 Inverter-based CTLE for PAM2 application

Before building a full AFE for a PAM4 ADC-based link, this work first investigated feasibility and performance of a simple inverter-based CTLE through a short reach application space. Recent standards, such as CEI-56G-XSR-NRZ, drive the demand for short-reach, high speed and high density wireline interfaces for die-to-die and chip-to-chip links with short PCB traces (see Figure 4.10). As a result of the short trace, impedance discontinuities have a relatively small impact on the channel, as seen by the smooth S21 roll-off and pulse response. This type of channel can be equalized effectively using only one CTLE at the receiver side. Despite these relaxed requirements and low channel loss (<10dB), it is still challenging to implement the CTLE due to the high bandwidth requirements. An area and power efficient single-stage inverter-based CTLE was designed to address these needs.

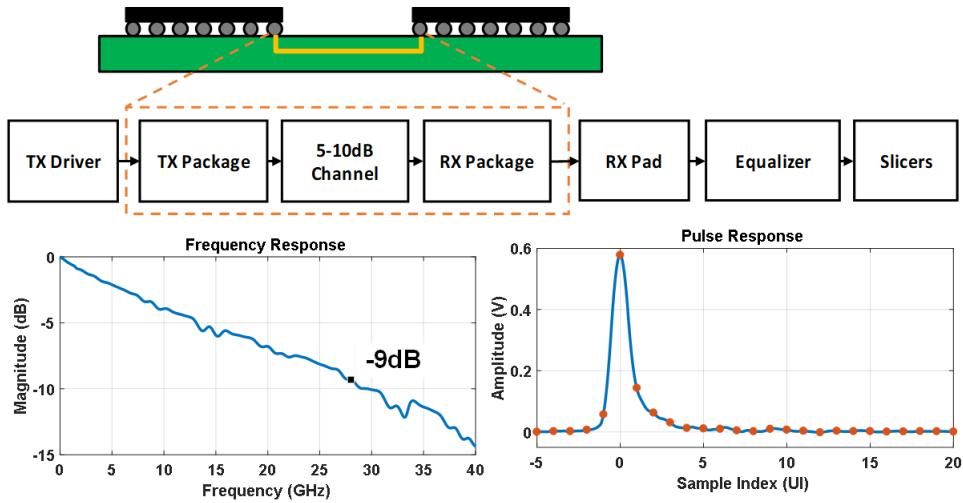


Figure 4.10: Short-reach link application block diagram with channel responses

4.2.1 Additive two-path CTLE

Conventional CML-style CTLEs achieve the desired frequency response with source degeneration, but having a source node network in an inverter would affect its bias point. Therefore, additive two-path CTLEs [31] are a better option for inverter-based circuits. As shown in Figure 4.11(a), an additive CTLE uses a feedforward coupling capacitor to sum the currents from both a low and high gain path. Thus, there is no wasted power in generating high frequency peaking in this topology, which suits a PAM2 short reach application well.

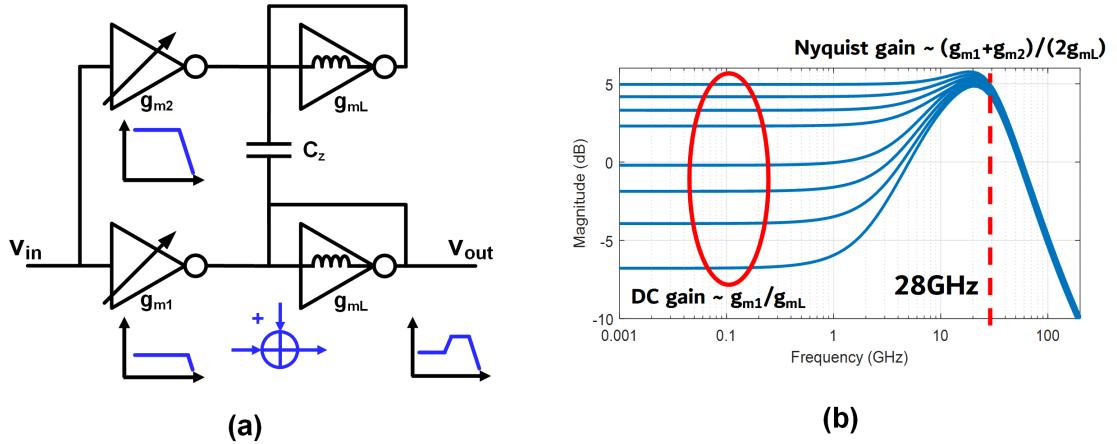


Figure 4.11: (a) Single-ended CTLE schematic and (b) simulated frequency responses

Fig. 4.11(b) shows the simulated frequency responses of our design. The low-frequency gain is determined by the bottom path inverter ratio, g_{m1}/g_{mL} . The high-frequency gain is approximately the ratio of total active and load transconductances, $(g_{m1} + g_{m2})/(2g_{mL})$. The coupling capacitor C_z is implemented with a fingered MOM device. Active inductors are used in both low- and high-frequency paths for bandwidth extension. Both g_{m1} and g_{m2} are tunable and their sum is kept constant, which results in de-emphasis in the CTLE's transfer function and provides peaking in a power efficient manner. The $g_{m1/2}$ and g_{mL} ratios are also tuned to compensate for gain reduction due to finite output resistance of the inverters.

The CTLE's transfer function can be written as

$$\frac{v_{out}}{v_{in}} = -\frac{g_{m1}}{g_{mL}} \cdot \frac{1 + s \left(1 + \frac{g_{m2}}{g_{m1}}\right) \frac{C_z}{g_{mL}}}{1 + s \frac{2C_z}{g_{mL}}} \cdot P(s) \quad (4.17)$$

in which $P(s)$ contains the bandwidth-extending zero and pole from active inductors, as well as parasitic poles determined by load g_m , drain parasitics (C_{dd}), the subsequent slicers input gate capacitance (C_{gg}), and any wiring capacitance. $P(s)$ is an important term since it determines the peaking strength and bandwidth, which approximately scales with the ratio $g_m/(C_{gg}+C_{dd})$. This also corresponds to the inverter small-signal unity gain frequency ω_u . Therefore, a biasing approach that attempts to stabilize this ratio is needed.

4.2.2 Replica ring oscillator based biasing

Though inverter-based circuits can have relatively stable voltage gain due to their ratiometric nature, their frequency response, including parasitic pole location, is determined by the absolute values of the transconductances and capacitances in the circuit, which are heavily dependent on PVT conditions.

To address this issue, this work employs a replica biasing technique using a ring oscillator, which is widely used for process monitoring. Traditional constant- g_m biasing circuits (such as [32]) consume static current and typically, device sizes and power must be scaled up to reduce the impact of random mismatch. On the other hand, a ring oscillator consumes only dynamic power and this power is nearly independent of the number of stages. Thus, it is possible to use a large number of stages to minimize random variations in oscillation frequency.

However, the relationship between a ring oscillator's oscillation frequency f_{osc} and an inverter's unity gain frequency f_u needs to be understood first. For a ring oscillator

with inverters of equal PMOS and NMOS strengths, the oscillation frequency is

$$f_{osc} = \frac{1}{2N \cdot t_p} \quad (4.18)$$

where N is the number of stages and t_p is the average inverter propagation delay. Due to symmetry in the inverters, there is no significant difference between rising and falling edge delays. Assuming that the gate delay is dominated by slewing and that the transistors obey the square-law, we can express the ring oscillator inverter delay as

$$t_p = \frac{V_{DD}}{2} \frac{C_{gg} + C_{dd}}{\frac{W}{2L} \mu C_{ox} (V_{DD} - V_T)^2} \quad (4.19)$$

For the inverters in the analog signal path, the gate bias voltages are at $V_{DD}/2$ and thus their transconductance is

$$g_m = \frac{W}{L} \mu C_{ox} \left(\frac{V_{DD}}{2} - V_T \right) \quad (4.20)$$

We note that V_{DD} directly affects the analog inverter's g_m , and thus also the ratio $\omega_u = g_m/(C_{gg} + C_{dd})$. Now, expressing (4.18) with (4.19) and (4.20), we obtain the following expression for oscillation frequency in terms of g_m

$$f_{osc} = \frac{1}{2N} \frac{(V_{DD} - V_T)^2}{V_{DD}(V_{DD}/2 - V_T)} \frac{g_m}{C_{gg} + C_{dd}} = \frac{\pi}{N} \alpha f_u \quad (4.21)$$

In this expression, α is a function of V_{DD} . However, when $V_{DD} \gg V_T$, α approaches 2 and f_{osc} becomes directly proportional to f_u . In our design, ULVT devices ($V_T=150\text{mV}$) are used for the PAM2 application and the nominal V_{DD} is 700mV for a $g_m/I_D \approx 10\text{S/A}$. Therefore, α is roughly constant as shown in Figure 4.12(a) by plotting the simulated α vs. V_{DD} , demonstrating small variations (within $\pm 5\%$) in the relevant V_{DD} range. As a result, tuning V_{DD} for constant f_{osc} can be exploited to

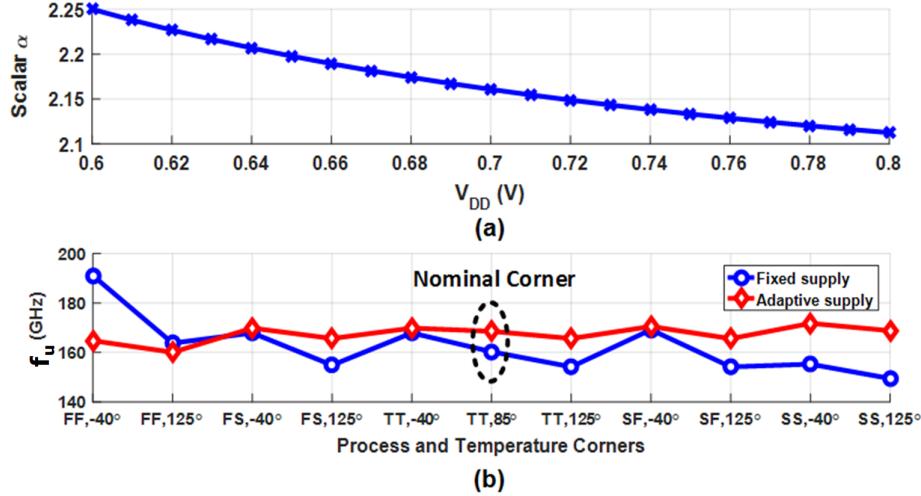


Figure 4.12: (a) Simulation of parameter α vs. V_{DD} and (b) simulated inverter f_u at different process and temperature corners

stabilize f_u across corners. Figure 4.12(b) illustrates this by plotting the simulated inverter f_u for different process and temperature corners. For fixed V_{DD} , we see large f_u variations across corners. However, when V_{DD} is computed by an adaptive loop that maintains constant f_{osc} , the f_u variations become small. This translates to a relatively fixed parasitic pole frequency of $P(s)$, thus maintaining high bandwidth for the CTLE. Note that this tuning mechanism does not perfectly stabilize the (less critical) g_m/C_z zero location, but it still helps in counteracting some g_m variations.

4.2.3 Transceiver testbed and measurements

Using the transceiver in [33] as a testbed in a 16nm FinFET process, we fabricated the proposed CTLE and ring oscillator based LDO biasing scheme to interface with the existing receiver without too much modifications. There are other channels with conventional CML-based CTLEs on the same die for performance comparisons.

Figure 4.13 shows the overall implementation of the system along with the adaptive supply loop. The CTLE core has pseudo-differential paths, which provides some

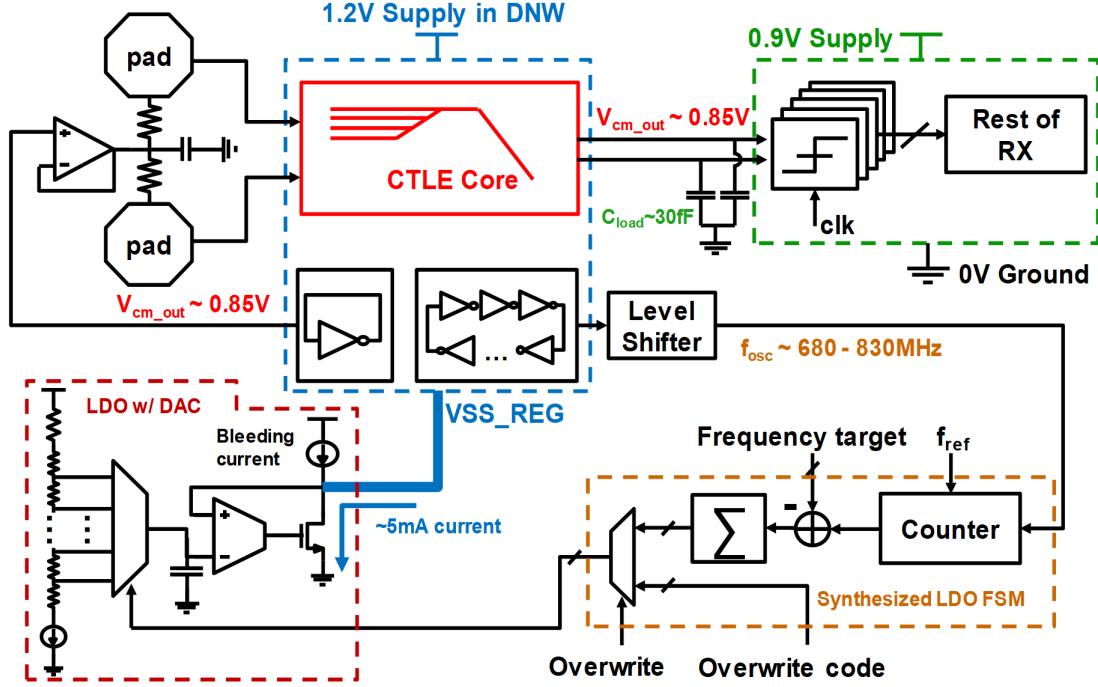


Figure 4.13: PAM2 short reach receiver testbed block diagram

supply and common-mode noise rejection. In order to interface with the succeeding NMOS input slicers, a ground regulation scheme is chosen to achieve a higher output common mode. The nominal output common mode is set to 0.85V. As a result, the core devices all sit in a deep N-well (DNW) to allow ground regulation. The total capacitive load for the CTLE is approximately 30fF (including the input capacitance of five slicers and wiring parasitics). A replica diode-connected inverter is used as the common-mode reference for input termination. The ring oscillator's output is used as the clock for a finite state machine (FSM) that controls the LDO's output voltage as the ground for the core and reference circuits. The FSM uses an available 100 MHz clock reference to tune the LDO voltage such that the ring oscillator clock achieves a programmable frequency target (nominally 740MHz, corresponding to 5ps inverter delay) with hysteresis. The frequency target is set externally for optimal CTLE bandwidth, power and performance. The CTLE's replica-bias block (including reference

diode, ring oscillator, LDO and FSM) can be shared by multiple transceiver channels, thus amortizing the area and power cost of the biasing circuits.

The CTLE is tested as the only means of equalization in the transceiver testbed operating at 56 Gb/s with PAM2 modulation. As shown in Figure 4.14, the inverter-based CTLE has similar performance as the CML-based CTLE when tested with the same channel, achieving 31% UI horizontal opening at $\text{BER} < 10^{-12}$ for a channel with 8dB loss at 28GHz. Measured bathtub curves for different LDO modes and

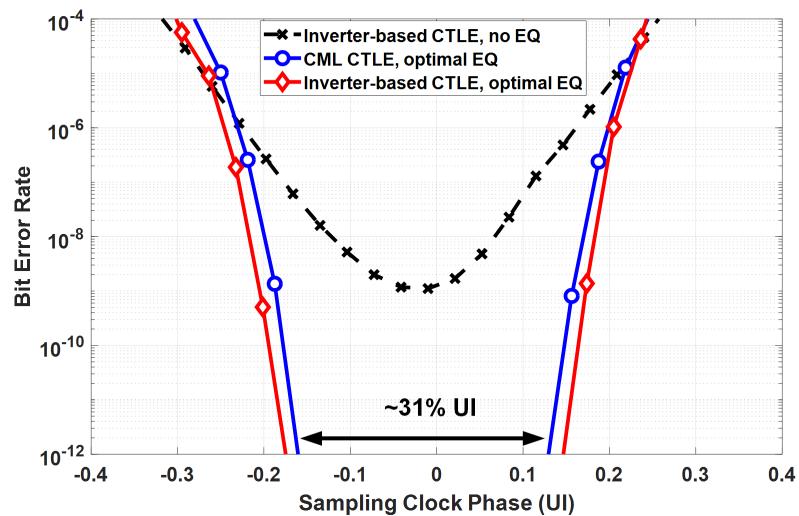


Figure 4.14: Bathtub curve comparison between CML-based and inverter-based CTLEs under nominal conditions

various temperatures are shown in Figure 4.15, demonstrating the effectiveness of the employed biasing scheme. Figure 4.16 shows the bathtub comparison of CML-based and adaptive supply inverter-based CTLE for extreme temperatures, in which inverter-based CTLE performs at least as well.

When the LDO is in adaptive mode, a larger eye width is achieved at higher temperatures as shown in Figure 4.17(a). To further validate the function of the replica bias scheme, the regulated ground voltage is plotted against varying temperature for different LDO modes in Figure 4.17(b). In overwrite mode, a fixed LDO code is

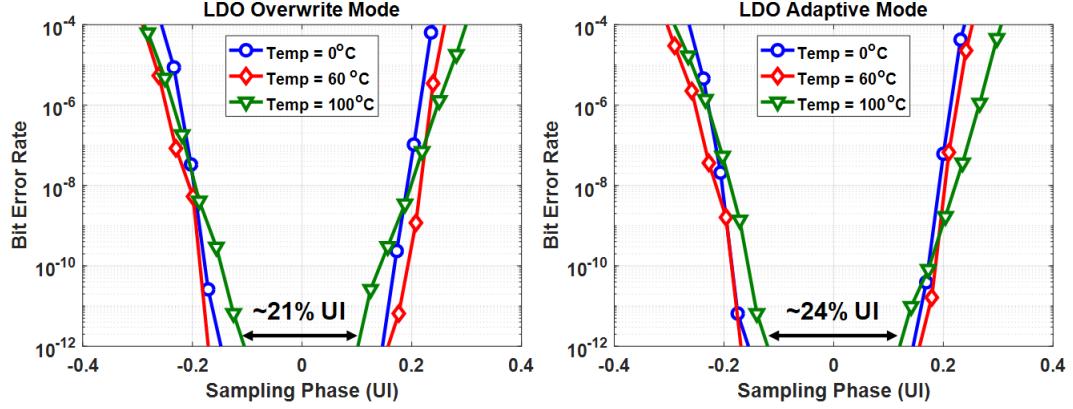


Figure 4.15: Bathtub curve comparison between overwrite and auto LDO modes for different temperatures

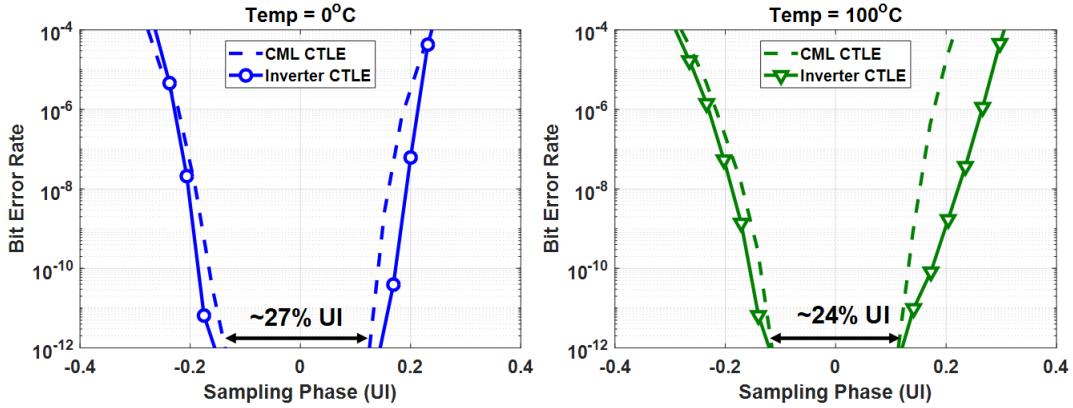


Figure 4.16: Bathtub curve comparison between CML-based and inverter-based CTLEs under extreme temperatures

applied and the ground voltage increases due to its bias resistors temperature coefficient. In auto mode, the LDO code is adapted and the ground voltage decreases as expected to maintain oscillation frequency at higher temperature.

As indicated in Table 4.1, the CTLE core consumes 6 mW (at room temperature), and measures only $20 \mu\text{m} \times 15 \mu\text{m}$ (see Figure 4.18), which is 13x smaller than the CML-based CTLE core on the same test chip. The inverter-based CTLE achieves similar performance with no extra power. Compared to previous work with

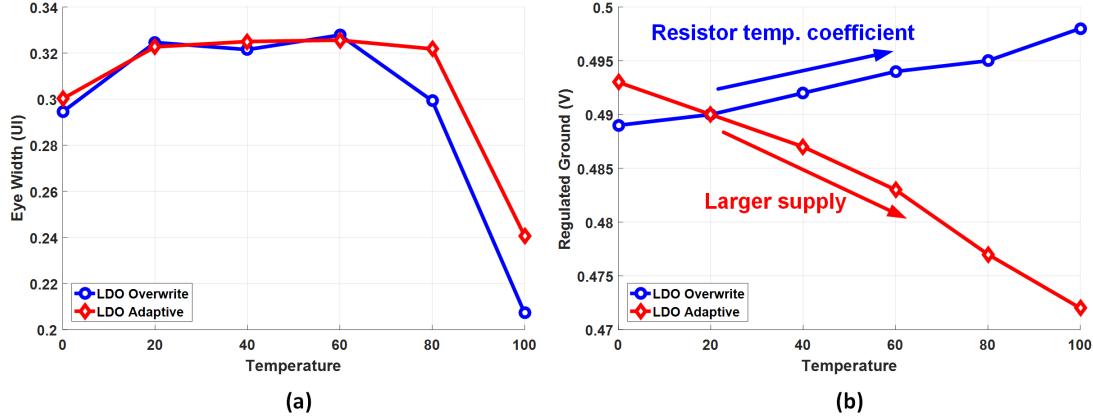


Figure 4.17: (a) Eye widths vs. temperature and (b) regulated ground voltage vs. temperature in different LDO modes

Reference	[34]	[19]	[33]	This work
CTLE type	2-stage CML	CML	CML	Inverter
Modulation	PAM2	PAM4	PAM2	PAM2
Nyquist frequency	6.25 GHz	14 GHz	28 GHz	28 GHz
Process	32 nm SOI	16 nm FinFET	16 nm FinFET	16 nm FinFET
Supply	1.1 V	1.2 V	1.2 V	1.2 V + Ground LDO
Max peaking (DC/Nyq. gain)	-6 dB/12 dB	0 dB/7 dB	-6 dB/6 dB	-6 dB/6 dB
Channel loss	27 dB ^a	31 dB ^a	8 dB	8 dB
Timing margin	50% @1E-12	N/A	>24% @1E-12 (100 °C)	>24% @1E-12 (100 °C)
Core power	5.25 mW ^b	8.4 mW	6 mW	6 mW
Power/freq.	0.84 mW/GHz	0.6 mW/GHz	0.21 mW/GHz	0.21 mW/GHz
Core area	Not reported	125 μm × 40 μm	80 μm × 50 μm	20 μm × 15 μm

^aOther means of equalizations are also used

^bEstimated from power breakdown chart

Table 4.1: Comparison table for inverter-based additive CTLE

similar peaking range, the inverter-based CTLE shows significant improvements in both power and area. The robustness of the biasing approach was proven through measured temperature sweeps.

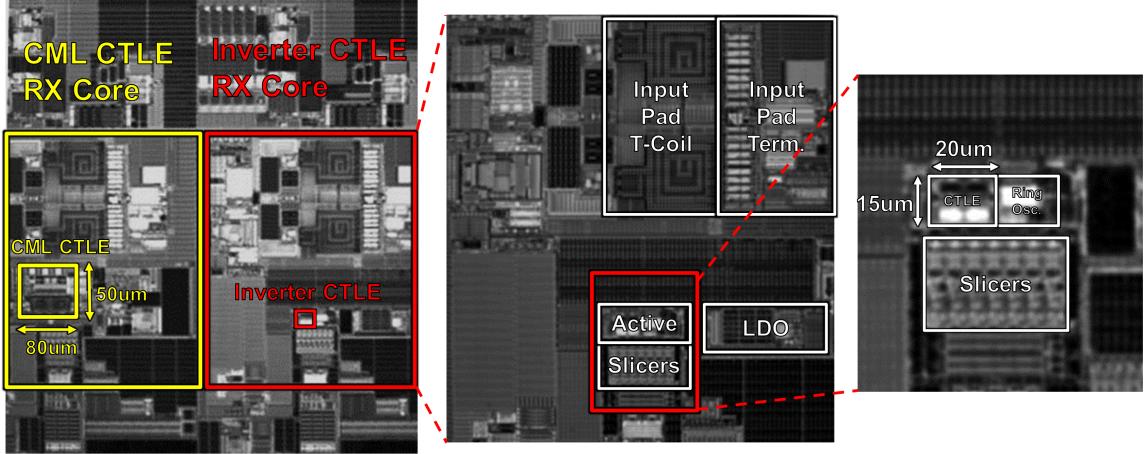


Figure 4.18: Chip photos

Even though we have demonstrated the feasibility and good performance of an inverter-based CTLE for short reach PAM2 applications, the system and circuit requirements for a PAM4 ADC-based link will be quite different. We will expand on this work to show the design, implementation and verification of a complete AFE for a PAM4 transceiver in the next section.

4.3 Inverter-based AFE for PAM4 application

Our proof-of-concept work [28, 35], has illustrated the advantages of inverter-based circuits that extract higher speeds for a given technology and can reject process variations through ratiometric design and replica-based tuning. If we were to extend such designs toward a PAM4 system with more challenging requirements, the specific CTLE topology and interface with ADC needs to be revisited. Similar to the previous chip, we are using another full PAM4 56Gbps ADC-based transceiver [36] as a testbed to validate the performance of proposed inverter-based AFE.

4.3.1 CTLE topologies

The previous section presented an additive CTLE in a PAM2 link context. However, this topology has its limitations when used in a PAM4 receiver due to the linearity and tunability requirements. We introduce a subtractive CTLE topology here, which is more suited for PAM4 applications.

Figure 4.19 shows the half circuit schematic of these two topologies and their advantages and disadvantages. For a fair comparison, the peaking tuning scheme needs to be agreed upon. Different from the previous PAM2 application, it is important to have high frequency gain to restore attenuated main cursor strength for high loss channels in PAM4 scenarios. Therefore, we will compare the two topologies such that the DC gains stay at 0dB and actual high frequency gain boost is realized.

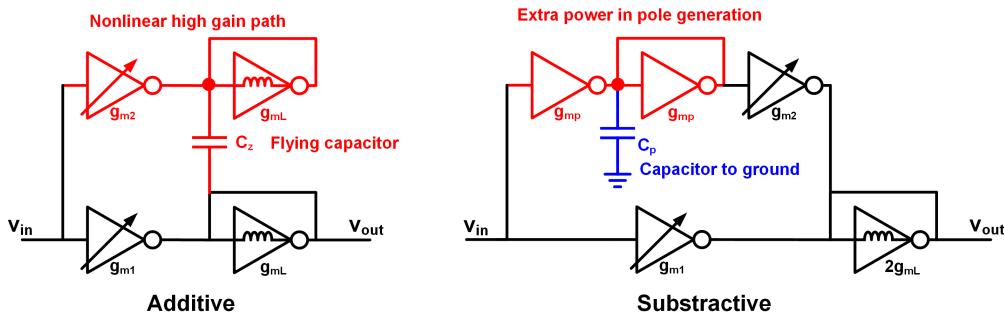


Figure 4.19: Pros and cons of additive and subtractive CTLEs

The subtractive CTLE creates peaking by subtracting a pole generated by the low pass filter path (g_{mp}/C_p). The subtractive CTLE wastes power by throwing away signal in addition to the extra power in the pole buffer. To make a fair power comparison between the two CTLEs, the high frequency output impedance of these two CTLEs are both $1/(2g_{mL})$ to have the same driving capability. Table 4.2 shows the required g_m 's normalized to g_{mL} when using either CTLE for 0dB and 6dB Nyquist gain. It is easy to see that the subtractive CTLE will consume more power since more g_m elements are required to achieve the same frequency response as additive

CTLE	Additive		Subtractive	
Nyquist gain	0 dB	6 dB	0 dB	6 dB
g_{m1}	1	1	2	4
g_{m2}	1	3	0	2
g_{mL}	2	2	2	2
g_{mp}	0	0	ϵ	ϵ
Total	4	6	$4+\epsilon$	$8+\epsilon$

Table 4.2: Additive and subtractive CTLE required g_m comparison

CTLE. Nevertheless, subtractive CTLE offers distinctive advantages in tunability and linearity over additive CTLE. For example, in order to vary the zero location in the frequency response, both capacitors (C_z and C_p) need to be programmable. Since C_p is a capacitor to ground, it is much easier to tune and parasitics will not affect the main high speed signal path as compared to the flying capacitor C_z . C_p can also be a MOSCAP, which has much higher capacitance density.

The more important aspect is the subtractive CTLE's linearity advantage. As we see in Table 4.2, the high gain path in an additive CTLE for 6 dB boost needs to have a gain of 3, given by g_{m2}/g_{mL} . On the other hand, the high gain path in a subtractive CTLE only needs a 2x gain, given by $g_{m1}/(2g_{mL})$. The broadband input signal is much more likely to saturate the additive CTLE's high gain path, while a subtractive CTLE's devices are protected from being pushed into triode due to the subtractive nature. As a result, we will use the subtractive CTLE for PAM4 applications despite the power penalty.

4.3.2 Transceiver and receiver AFE architecture

The transceiver architecture is shown in Figure 4.20. Utilizing the same transmitter, configurable SAR ADC bank and digital signal processing blocks (equalization, adaptation and calibration) as in [36], this work replaces the CML-style AFE with inverter-based circuits. The AFE block contains inverter-based hybrid CTLEs and

programmable gain amplifiers (PGA), as well as the ADC first-rank track-and-hold (T/H) driver stages, which are also implemented with inverters. There are 32 interleaved SAR ADCs whose resolution is reconfigurable from 3 to 7 bits, which we will use to study the trade-offs between ADC resolution and pre-equalizer strengths later using silicon data.

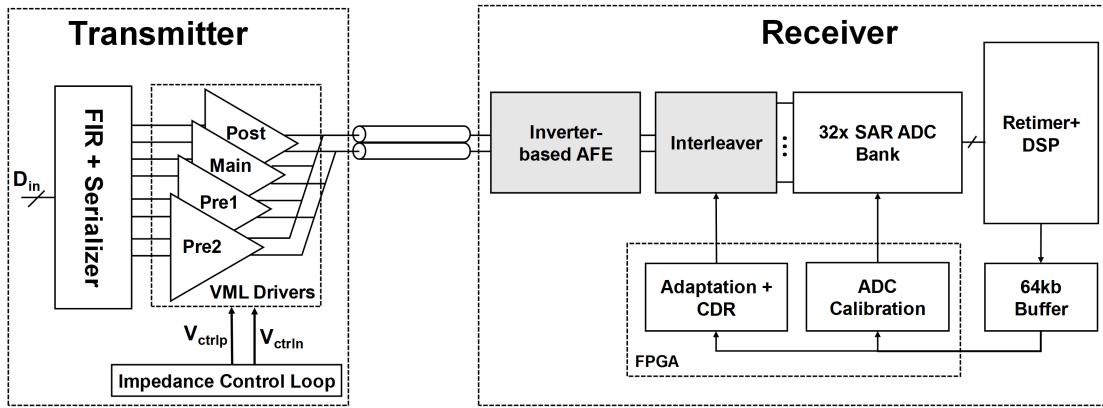


Figure 4.20: PAM4 ADC-based transceiver architecture

Figure 4.21 shows the block diagram of the receiver AFE. Hybrid CTLEs that combine both low- and high frequency peaking into a single stage are employed. Low frequency peaking is very effectively in canceling the long tail ISI portion of a high-loss channel's pulse response, which serves the purpose of reducing channel's PMR well. A larger gain-bandwidth (GBW) product is achieved by dividing the equalizer into two identical CTLE stages. Fewer parasitic poles are created due to fewer number of CTLEs. We have concluded that the input PDF has to be conditioned as early as possible to relax any subsequent circuit block's requirements. Therefore, the first stage CTLE code will be maximally tuned first before the second-stage CTLE dials up its strength. An offset correction node is added at the output of second CTLE stage. The offset correction voltage (V_{offset}) is generated from a resistor ladder DAC with common-mode (CM) feedback. The CM reference is generated by a diode-connected replica inverter. PGAs are used after the CTLEs to optimally clip the ADC. Similar

to the CTLE stages, two PGA stages are used for higher overall GBW. The second stage PGAs also act as the first-rank T/H buffers in the ADC. The inverter-based AFE uses a 1.2V supply to ensure sufficient linearity while devices reliability is not compromised with a 0.6V CM voltage. Unit inverters cells under the 1.2V domain use SVT devices for good linearity as discussed before and for the desired $g_m/I_D = 10\text{S}/\text{A}$ biasing condition. The ADC core uses the same 1.2V analog supply and a 0.9V digital supply.

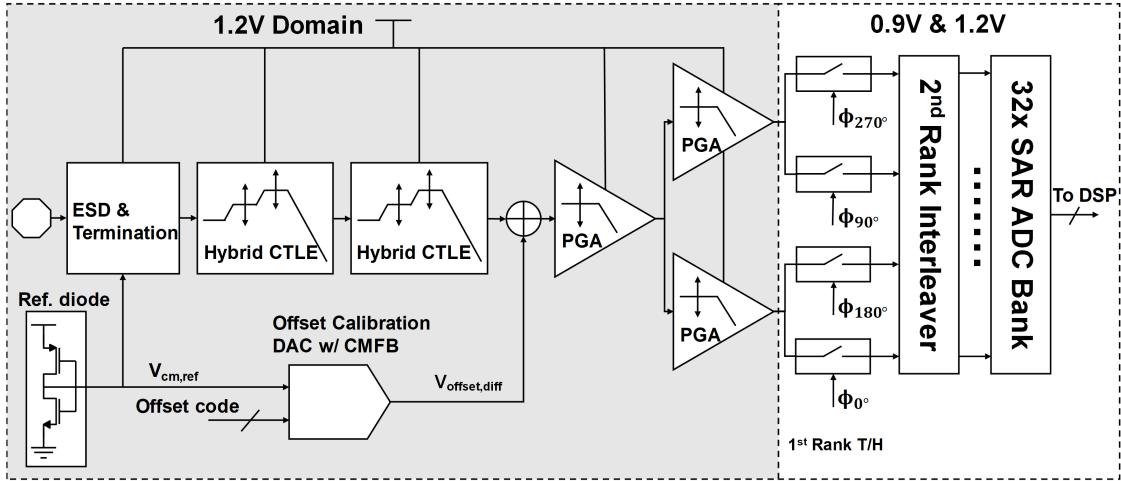


Figure 4.21: PAM4 ADC-based receiver AFE block diagram

4.3.3 Inverter-based hybrid CTLE and ADC T/H

The hybrid CTLEs single-ended schematic is shown in Figure 4.22(a). Both the low- and high-frequency paths are isolated from the all-pass path with transconductors $g_{m,ap}$, $g_{m,hf}$, and $g_{m,lf}$, which also act as current summing elements. The second CTLE stage additionally has an offset correction transconductance ($g_{m,cal}$) at the current summing node. The CTLE's output impedance stays relatively constant and is determined by the active inductor (g_{mL}), which further boosts the bandwidth. Low- and high-frequency poles are formed by the g_{mp} 's and their respective programmable

MOS capacitor banks, C_{lf} and C_{hf} . C_{lf} is implemented with five C_{hf} banks connected in parallel. The equalization strengths are tuned by changing the current summing g_m 's values. The same amount of transconductance (α, β) is added to or subtracted from the all-pass path when LF and HF paths. This ensures that the DC gain stays roughly at 0dB. However, a small amount of de-emphasis is still desired to increase the amount of peaking, especially for large high frequency boost settings. Therefore, the pole generation buffers' gain (g_{mp1}/g_{mp} and g_{mp2}/g_{mp}) are intentionally designed to be slightly larger than unity. The CTLE transfer function can be expressed as

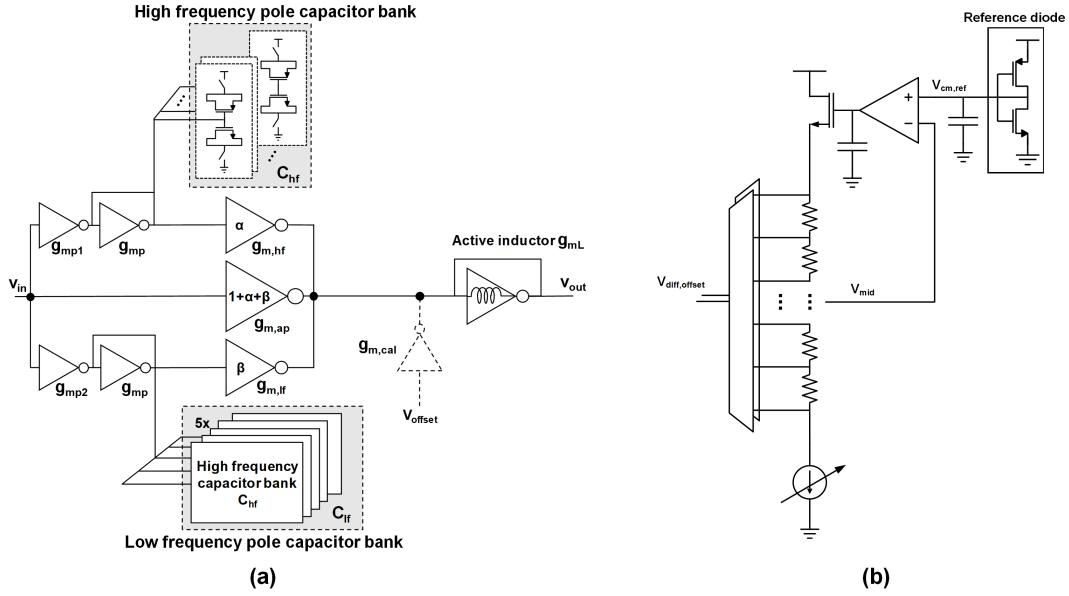


Figure 4.22: (a) Half circuit schematic for hybrid CTLE and (b) schematic for offset calibration DAC with common mode feedback

$$\frac{v_o}{v_i} = - \left(\frac{g_{m,ap}}{g_{mL}} - \frac{g_{m,hf}}{g_{mL}} \frac{g_{mp1}/g_{mp}}{1 + sC_{hf}/g_{mp}} - \frac{g_{m,if}}{g_{mL}} \frac{g_{mp2}/g_{mp}}{1 + sC_{lf}/g_{mp}} \right) \quad (4.22)$$

Figure 4.22(b) shows the schematic for the offset calibration DAC that generates the calibration voltages. It is a current resistor ladder with a common-mode feedback such that the pseudo-differential voltages at the output centered around the desired

inverter diode voltage. The calibration voltage conversion gain is given by $g_{m,cal}/g_{mL}$, which in this case is about 1/4. Monte carlo simulations show that the offset variations is about 10mV_{rms} , thus the DAC voltage range is designed to be $\pm 150\text{mV}$ with 5-bit tuning to cover more than 3σ variations. Each code step translates to 2.3mV at the CTLE output.

The post-layout simulated CTLEs frequency responses are shown in Figure 4.23, illustrating both de-emphasis and high frequency gain. Active inductors give the CTLEs enough bandwidth such that the peaking happens at the Nyquist frequency (14 GHz). The total 2-stage HF peaking range is 12dB with 5-bit tuning, and 6dB

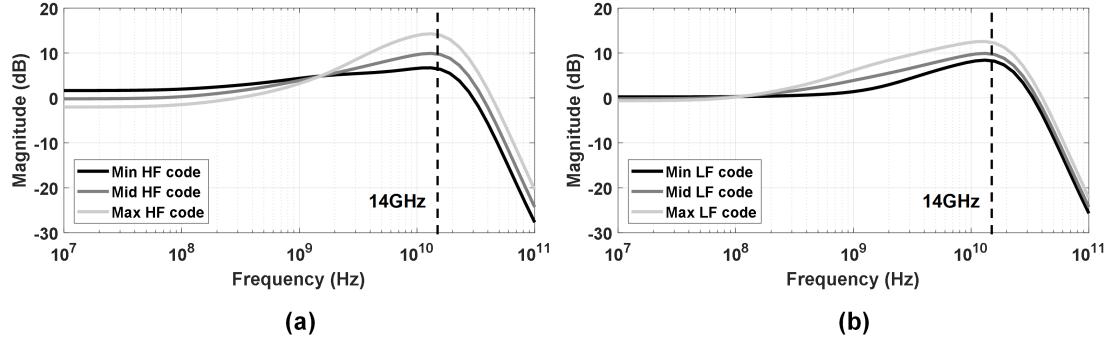


Figure 4.23: Simulated two-stage CTLE frequency response for (a) mid LF code and different HF codes, and (b) mid HF code and different LF codes

with 4-bit tuning for LF peaking. All tunable g_m 's are binary coded, with the MSB determining which stage is active (i.e., MSB = 0 means tuning the first stage, and MSB = 1 means the second stage). Besides, there is interaction between the HF and LF paths seen in (4.22). Nevertheless, the adaptation engine will still find the desired settings given the step sizes are small enough and the ranges are large enough.

Figure 4.24 shows the schematic of the PGAs and T/H circuits that interface with the ADCs. The PGAs are implemented using programmable transconductors loaded by active inductors, which help boost the bandwidth to 25GHz, which is also sufficient bandwidth to satisfy the settling requirement when used as the 1st rank T/H

buffers. The T/H switches are CMOS switches with cross-coupled dummies to cancel capacitive hold-mode feedthrough. Explicit hold capacitors are eliminated to reduce the driver load (the parasitics satisfy the kT/C noise requirements). The CMOS switches are intentionally skewed in size so that charge injection and clock feedthrough provide the necessary output CM voltage drop to optimally bias the subsequent source followers without sacrificing linearity. The 2nd rank T/H bootstrapped switches and following SAR slices are the same design as in [36].

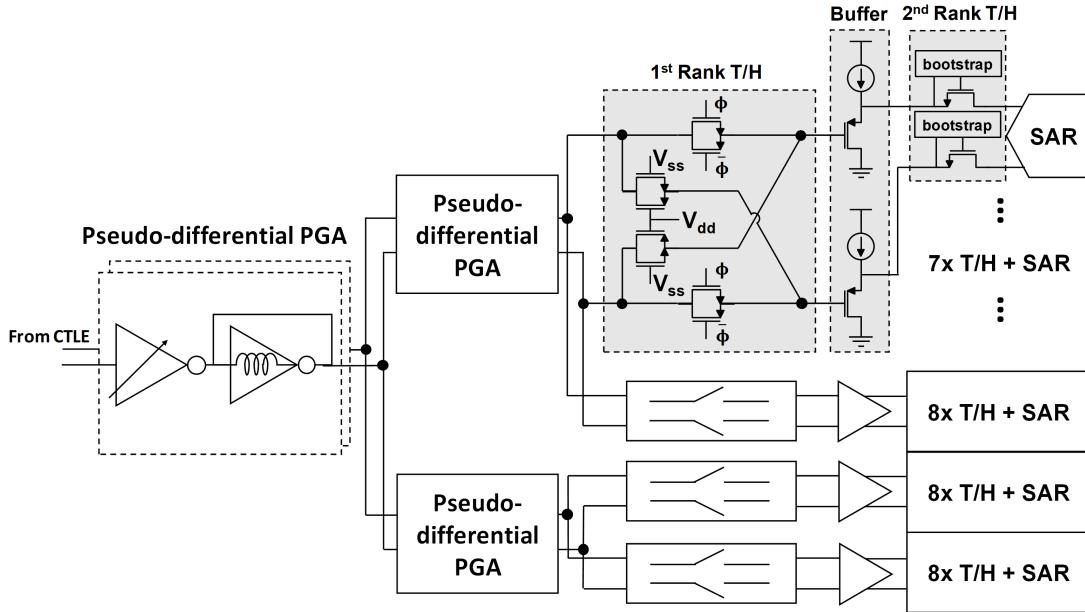


Figure 4.24: Schematics of PGA and ADC interface circuits

4.3.4 AFE and transceiver measurements

The transceiver is fabricated in 16nm FinFET CMOS and tested with a very short reach (VSR) and long reach (LR) channels (10 dB and 35 dB loss respectively at 14 GHz). Figure 4.25(a) shows the 7b ADC output scans for both channels and eye diagrams after DSP equalizations. No TX equalization is used for the VSR channel, which demonstrates sufficient equalization capability and linearity performance of the

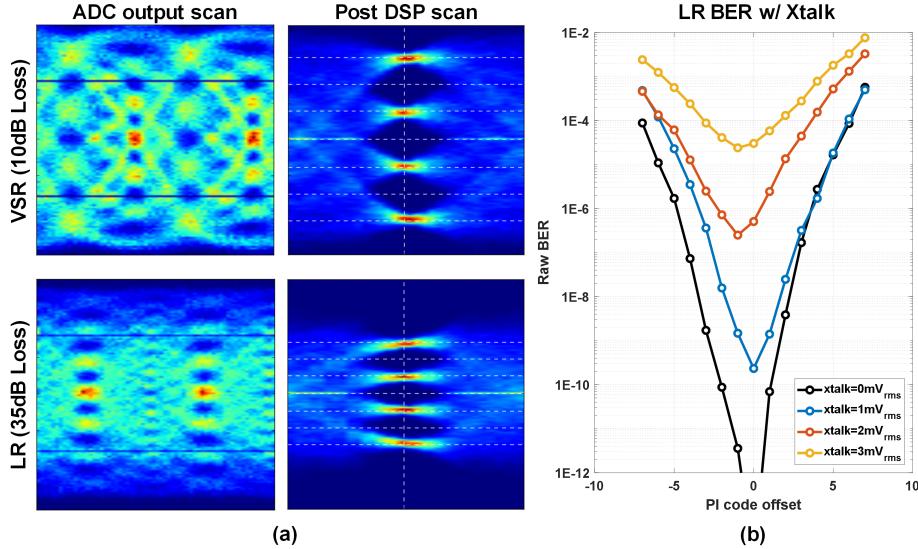


Figure 4.25: (a) ADC output scans and post DSP equalization eye scans of VSR and LR channels without crosstalk and (b) bathtub curves for LR channel under different crosstalk levels.

RX AFE. The bathtub curves for LR channel in Figure 4.25(b) show $< 10^{-12}$ BER without crosstalk and $< 10^{-6}$ BER with 2mV_{rms} crosstalk. This work is compared to [36] and the AFE specifications and transceiver performance are summarized in Table 4.3. The inverter-based CTLEs and PGAs occupy only $50\mu\text{m} \times 85\mu\text{m}$ (see Figure 4.26), and the overall AFE consumes 165mW total. The CTLE power is reduced because of fewer stages, and the ADC power decreases due to the inverter PGA T/H buffers.

With the validated functionality and performance of the designed transceiver, we proceed to test results that demonstrate other important aspects of inverter-based circuits and ADC-based transceivers. The transceiver's TX has a 3-tap FFE, with one pre-cursor and one post-cursor tap each. The receiver DSP has a 14-tap FFE (3 pre-cursor taps and 10 post-cursor taps) and a 1-tap DFE. The first post-cursor tap in the FFE is always set to zero so that the DFE handles it completely. The adaptation engine uses baud-rate Mueller-Muller clock data recovery (CDR) [37] to sample the

Reference	[36]	This work
Process	16nm FinFET	16nm FinFET
Power supplies	0.9V, 1.2V, 1.8V	0.9V, 1.2V, 1.8V
CTLE core area	$120 \mu\text{m} \times 190 \mu\text{m}$	$50 \mu\text{m} \times 85 \mu\text{m}$
CTLE + ADC power	40 mW + 146 mW	34 mW + 131 mW *
Channel loss @14 GHz	32 dB	35 dB**
Data rate	56 Gb/s	56 Gb/s
BER (2mV _{rms} xtalk)	$< 10^{-6}$	$< 10^{-6}$

*Power is code dependent. Settings for LR channel is used

**Different test setup induced additional channel loss

Table 4.3: Comparison table for PAM4 transceivers

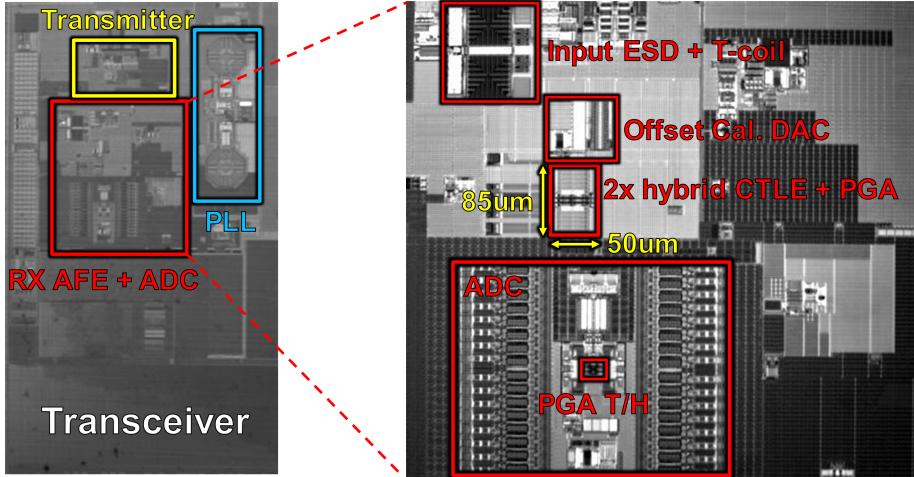


Figure 4.26: PAM4 ADC-based transceiver die photo

incoming signal after a certain amount of pre-equalization. The pre-equalizer is set such that the first pre- and post-cursors are $<1/3$ of the main cursor. The digital equalizers are adapted using least-mean-square (LMS) algorithm.

This architecture results in flexible and effective equalization on the receiver end. One way to demonstrate this is by sweeping TX settings and record transceiver performance. Figure 4.27 shows a contour plot to access system performance for all allowed TX settings. The BER range is roughly within two decades regardless of the TX setting, illustrating the robust nature of ADC-based links in general. It is

interesting to note that the worse BER points happen at minimal TX equalization settings (no RX input signal conditioning) and along the diagonal where sum of pre- and post-cursor settings is large (peak power constrained). This result confirms the system analysis in the previous chapters arguing for an optimal partitioning between TX and RX equalization.

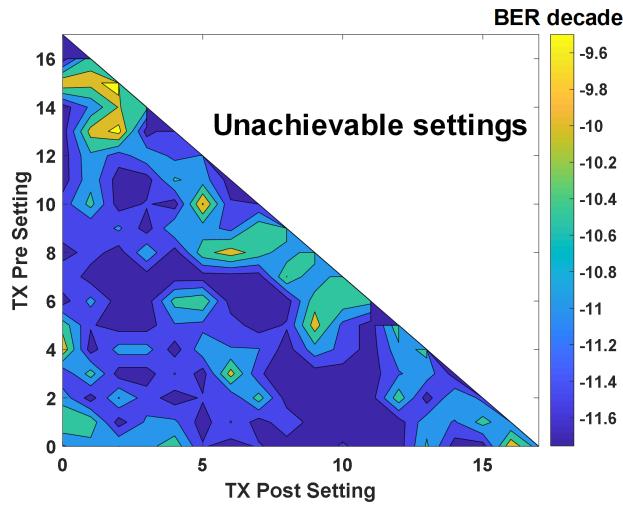


Figure 4.27: Transceiver BER performance for all TX settings

Since there is no adaptive supply biasing loop for this transceiver, system performance variations due to voltage and temperature changes need to be checked. Figure 4.28 and 4.29 show the measured BER and settled CTLE/PGA codes with respect to different voltage and temperature corners when the TX setting is fixed at 3 dB post-cursor boost. It is expected that the BER degrades for extremely high temperature since both the circuit bandwidth reduces and thermal noise increases. However, the system is still able to achieve a $< 10^{-7}$ BER, which meets current standard requirements with reasonable margin. It is more interesting to see how the AFE settings respond to different operating conditions. The high-frequency CTLE code (HF) increases significantly with rising temperature, showing that the system is combating the extra high frequency loss at high temperature. Meanwhile, both the LF and PGA

codes stay relatively constant across voltage and temperature corners. This is due to the ratiometric nature of inverter-based circuits, specifically the absolute gain is quite insensitive to the circuit's environment. For the channels under test, the used PGA codes are consistently lower than 50% of the total available codes, which means that the necessary gain range could be reduced for future iterations and further cut down circuit power and complexity.

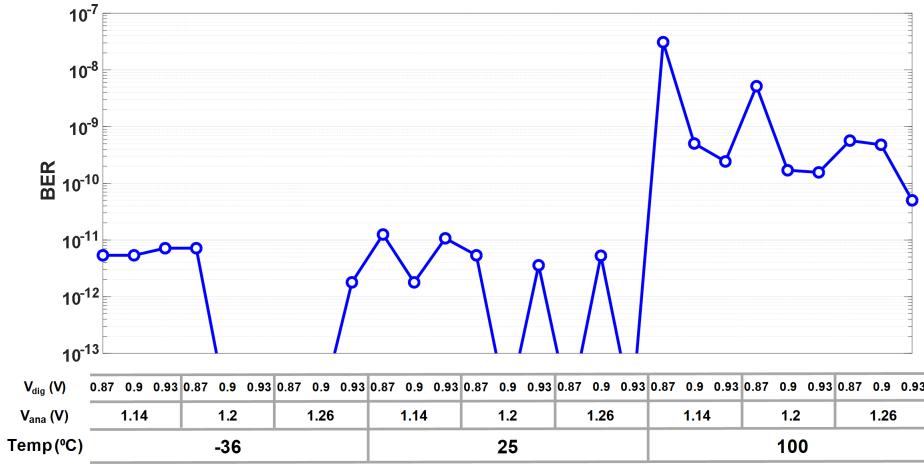


Figure 4.28: Transceiver BER for different voltage and temperature corners

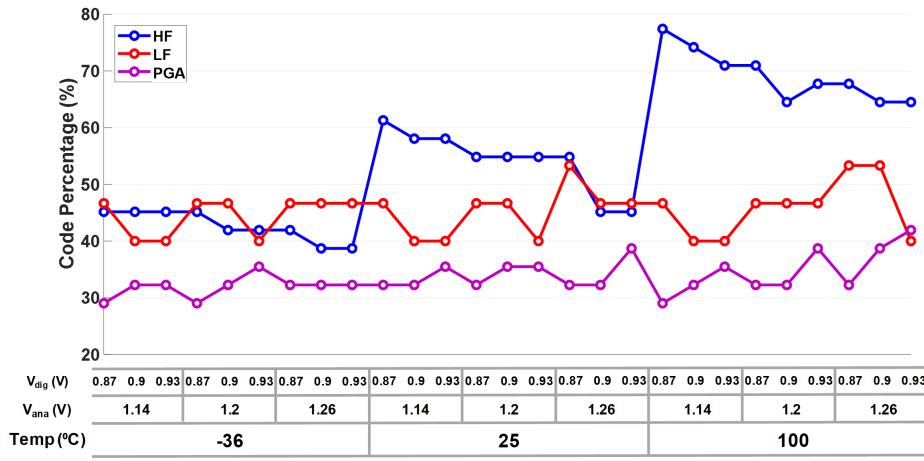


Figure 4.29: CTLE/PGA codes for different voltage and temperature corners

In this chapter, we have demonstrated two area- and power-efficient inverter-based equalizers for next-generation transceivers. The PAM4 ADC-based transceiver is particularly important for showing the effectiveness and robustness of such inverter-based AFEs. We will continue to use this chip to further study the system trade-offs with silicon data next to gain more insights and validate our system analysis.

Chapter 5

System Identification of ADC-based Links

Using sinusoidal input test signals has historically been the preferred choice due to its simplicity and natural extension to the frequency domain. However, the fact that a real link system sees a broadband signal and the final performance metric BER does not directly relate to the frequency domain means some other analysis and debug methods are needed. More importantly, such a method should give us a comprehensive view of the limiting error sources in the probability domain as this work has emphasized.

In this chapter, a system identification (SID) methodology is used to extract different circuit nonidealities. Silicon data from the previously described PAM4 transceiver will be used. SID can not only extract the equivalent model of the system, but the learned errors can also be used for BER estimation. Finally, this chapter closes the loop by using SID to show the trade-off between pre-equalization and ADC resolution and correlate with PMR improvements, verifying the importance of pre-equalization and further motivating an optimal pre-equalizer and ADC trade-off.

5.1 System identification

5.1.1 Working principles

System identification in essence is to use a least mean squares (LMS) adaptation algorithm [38] to extract the linear portion of any system and separate out any random and uncorrelated error sources given a known input signal. The block diagram for a SID engine in MATLAB is shown in Figure 5.1. PRBS signals are used to test the

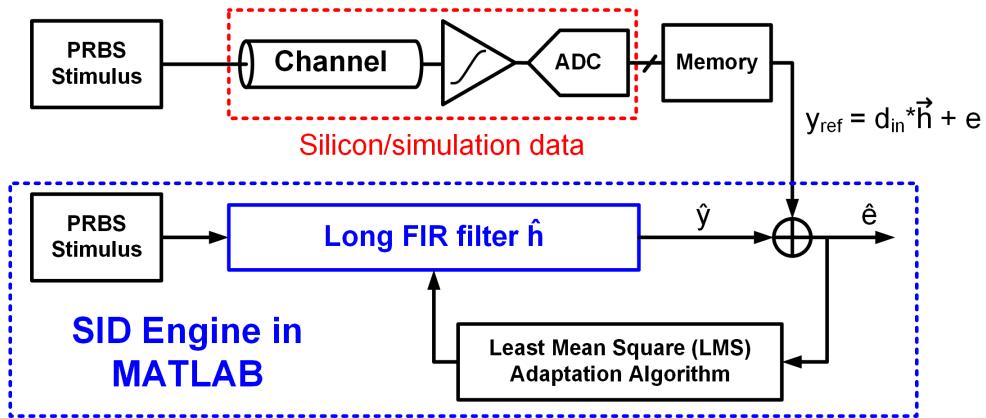


Figure 5.1: Block diagram for SID engine

ADC-based link, and such a system can be modeled as an equivalent linear filter \vec{h} with any circuit noise and errors e . The filter \vec{h} includes the physical communication channel, CTLE, any finite bandwidth effects in the ADC, etc. The error samples e include random thermal noise, circuit nonlinearity, ADC quantization, etc.

The SID engine is a statistical tool that estimates \vec{h} with a long FIR filter \hat{h} in software given the known test input d_{in} and reference output samples y_{ref} (saved on chip in memory). After the LMS algorithm converges, we are left with a linear estimate of the reference signal \hat{y} and residual errors \hat{e} , which are also estimates of the circuit nonidealities on silicon. Therefore, the SID estimates, \hat{h} , \hat{y} , and \hat{e} contain valuable information regarding the various error contributions in the system. For

example, \hat{h} models the discrete time channel pulse response up to the ADC. One can already visually check if there is ISI outside of FFE/DFE range and calculate PMR. The relationship between \hat{e} and \hat{y} contains information on what the dominant nonideality is, which we will investigate further in the following sections.

5.1.2 SID data collection and pre-processing

The presented PAM4 transceiver is used as a testbench to collect data for SID collection. First, the data collection and evaluation process is presented below:

1. For this discussion, only results for the LR (35dB loss) channel are presented.
2. There is 8KByte of on-chip memory, which means it can store 8192 ADC output samples (7 bits per sample). Therefore, we use PRBS13 data as input to the link since each PRBS13 sequence can fill the memory completely. This is a good trade-off between randomness in the data and processing complexity.
3. The transceiver is set to fully adaptive mode, and we let the system converge to its final point in terms of CTLE settings, FFE/DFE coefficients, and sampling phase.
4. After the system settles, all equalizer settings are frozen and the phase interpolator setting is swept for data collection at different sampling phase.
5. For each sampling phase, 32 memory snapshots are taken (which also means 32 runs of PRBS13 data. One example snapshot is shown in Figure 5.2). This gives us about 262 thousand samples to work with, which should have enough statistics on both deterministic and random errors.
6. Due to the nature of on-chip memory collection, the starting sample of each snapshot is not synced to a fixed bit position in a PRBS13 sequence. Therefore,

some post-processing is needed to rotate each snapshot to align with a reference PRBS13 sequence. Since there are only 8191 data samples in a PRBS13, the last sample in the snapshot is thrown away. This allows snapshot stitching so that a long periodic reference signal is obtained.

7. For the next examples, a 300-tap FIR filter (30 pre-cursors) is used for channel estimation and the LMS adaptation parameters are set appropriately.

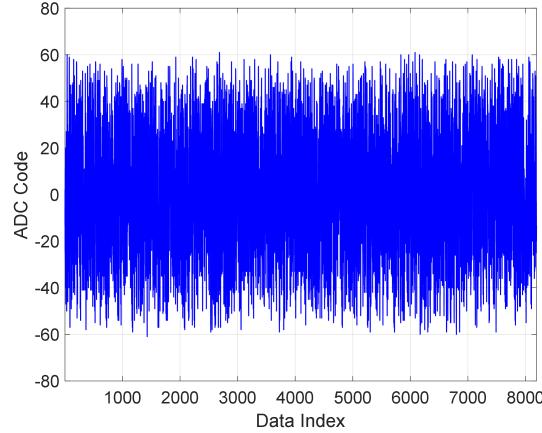


Figure 5.2: Example snapshot of PRBS13 ADC output stored in memory

5.1.3 Example SID outputs

Figure 5.3 shows example outputs of a SID run for data at the optimal sampling phase. The estimated channel \hat{h} is plotted in Figure 5.3(a) and truncated to 100 taps. Again, this is the equivalent channel pulse response up to the ADC, and several important features can also be seen. The main cursor value is >30 LSBs (7-bit ADC). The pulse is sampled at almost equal pre- and post-cursor strength (1/3 of main cursor), which is the desired criterion for Mueller-Muller CDR. The pre-equalizer performs well in cancelling ISI beyond 10 post-cursors (sampling index >40) while giving a

slight undershoot. There are small far out ISIs near 30+ post-cursor taps, which is due to reflections in the physical channel. The PMR of this channel is about 2.

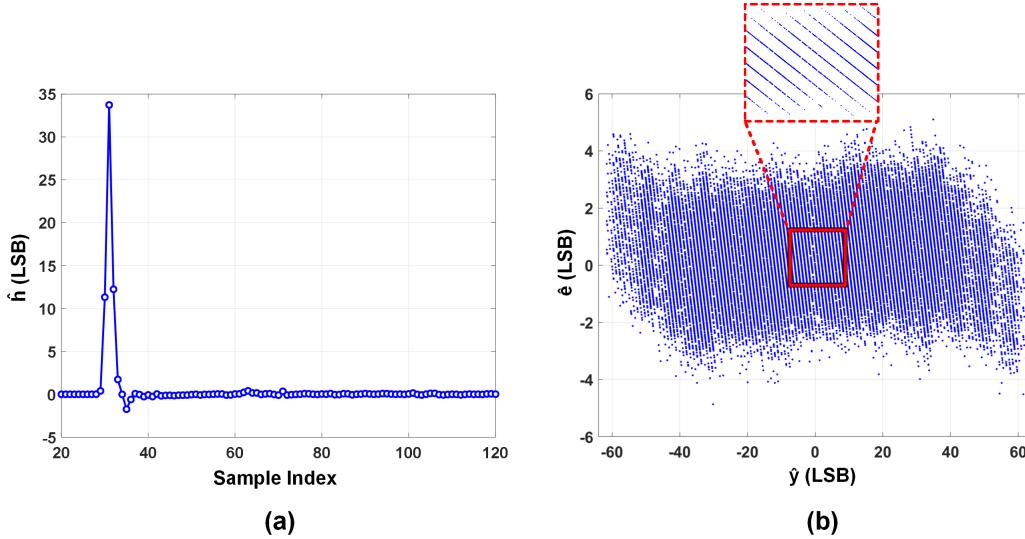


Figure 5.3: (a) Estimated channel and (b) estimated error correlation with estimated linear output

Figure 5.2(b) plots estimated error against the SID linear output. Interestingly, this plot is particularly useful in isolating different error sources in the system. For example, zooming into the center of the error scatter plot shows the familiar quantization sawtooth shape. There is also a noticeable static nonlinearity, which gives larger absolute errors for larger \hat{y} . This can also be interpreted as an input dependent mean shift in errors due to nonlinearity. This plot shows that the average error due to nonlinearity is merely ± 2 LSBs (3% w.r.t. FSR) for large output values, demonstrating the reasonable linearity of inverter-based AFE. Thermal noise is also captured statistically in this plot, which gives the occasional large errors (>5 LSBs).

More information regarding the pulse response can be gained when running SID on neighboring sampling phases. Figure 5.4 shows the pulse responses of sampling phases $\pm 8\%$ UI away from the optimal phase. It is interesting to note the symmetry of these sampled pulse responses when sampling phase moves left and right, indicating

a symmetry in the continuous time pulse response as well. The main cursor value doesn't decrease significantly, meaning the optimal phase samples at the peak of the pulse. Both the pre- and post-cursors changes significantly (by $\sim 20\%$), implying a relatively sharp slope in the pulse response. This further validates the high bandwidth performance of the inverter-based equalizer.

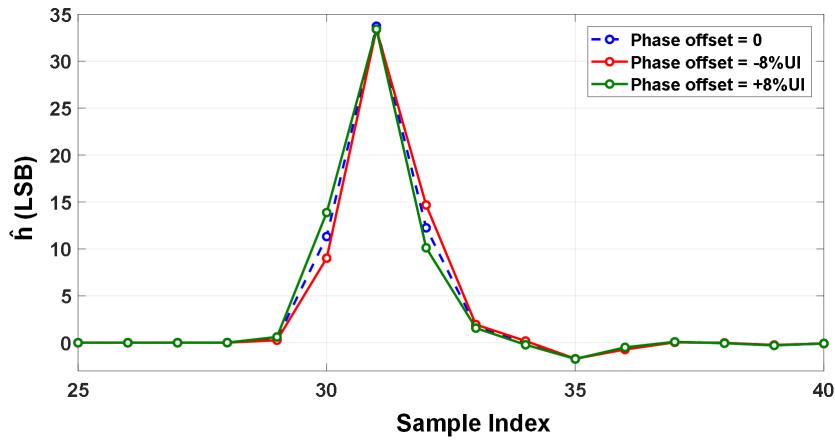


Figure 5.4: SID pulse responses at $\pm 8\%$ UI phase offsets

5.1.4 SID with FFE

One of the main advantages of SID for ADC-based links is the capability of mirroring DSP equalizers in software and checking its functionality. In addition to being a valuable debug tool, more sophisticated equalization techniques can be studied in software first with the real front end silicon data. Figure 5.5 shows how an FFE could be embedded in the SID engine. The result of such a SID run will be output and error estimates of the equalized channel after the FFE. Since an FFE is considered perfectly linear in the digital domain, it should not incur any extra errors other than filtering the incoming circuit noise, whose effects the SID will capture.

Figure 5.6 shows the SID output on the same data set as in Figure 5.3 with the

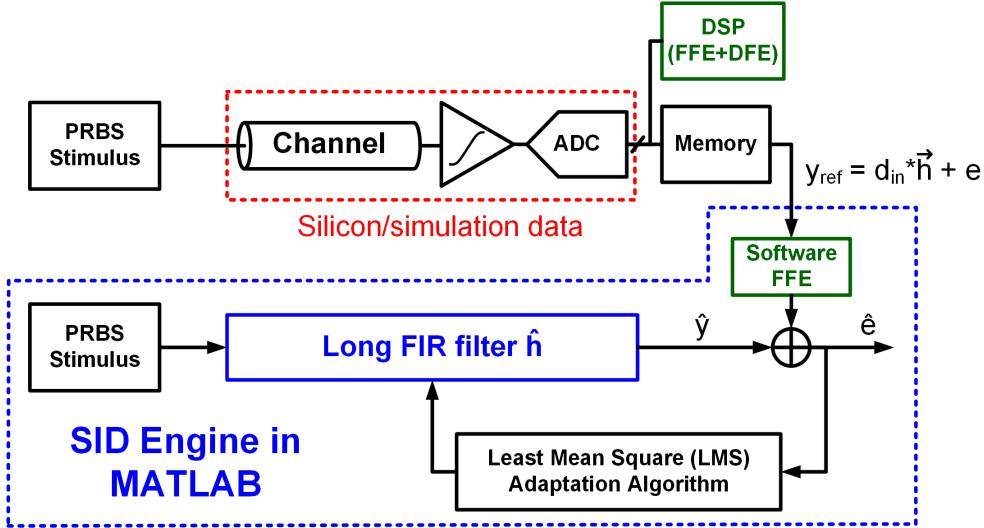


Figure 5.5: Block diagram for SID engine with FFE

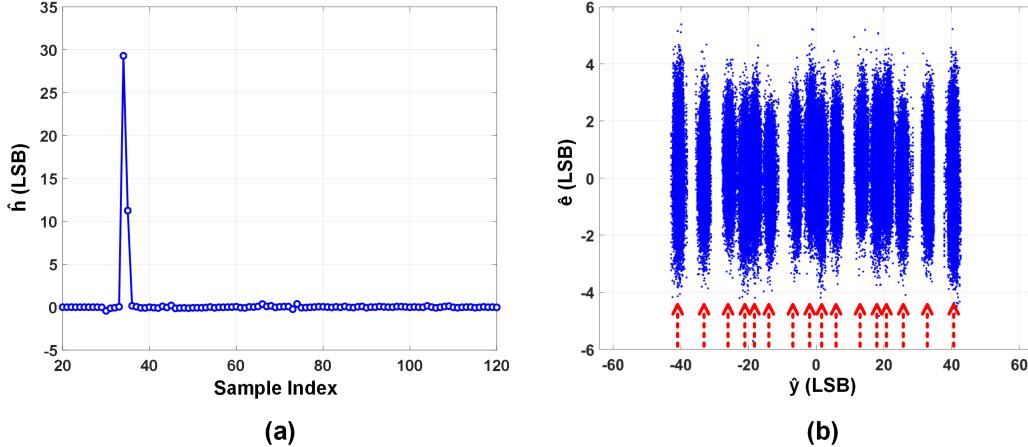


Figure 5.6: (a) Estimated channel and (b) estimated error correlation with estimated linear output with FFE

FFE added. The FFE coefficients are copied into software from on-chip DSP. The reference signal to the SID engine is then filtered by the FFE first before running. It is important to note that the SID engine still converges and provides the channel estimate with FFE incorporated. The estimated \hat{h} in Figure 5.6(a) only has main cursor and one post cursor left, which will be canceled by the DFE. This means that

the settled FFE coefficients are very effective in canceling rest of the ISI. The error scatter plot in Figure 5.6(b) also becomes very different. There are 16 clustered \hat{y} "slices", which are determined by all the combinations of main and post-cursor data (i.e., $4^2 = 16$). The error spread along the \hat{y} axis is due to residual ISI while the spread in \hat{e} is from the circuit noise filtered by FFE.

With software equalizers embedded in the SID engine, we can also analyze the output data further and estimate the BER performance statistically. BER estimation and prediction methods will be presented in the next section.

5.2 BER estimation using SID

Statistical link models have been studied before and show significant in BER estimation accuracy compared to simple Gaussian model [39]. The underlying principle is similar to the statistical framework for ADC-based links' circuit nonidealities. With the help of SID, we are able to separate circuit impairments from residual ISI in the channels while incorporating any DSP equalizer. In this section, we present a statistical BER estimation methodology using the SID outputs so that realistic circuit nonidealities are considered as well.

5.2.1 Residual ISI PDF

Here we briefly review the residual ISI PDF computation method proposed in [40]. In essence, the proposed method numerically calculates the residual ISI PDF by convolving individual PDFs due to each cursor. As shown in Figure 5.7, each cursor of strength s_i corresponds to a PDF with delta functions whose locations are given by $[-s_i, -s_i/3, s_i/3, s_i]$ in a PAM4 system. When enough such PDFs convolve together, the final PDF resembles a Gaussian distribution again (central limit theorem), but is still bounded by the L1-norm of residual ISIs.

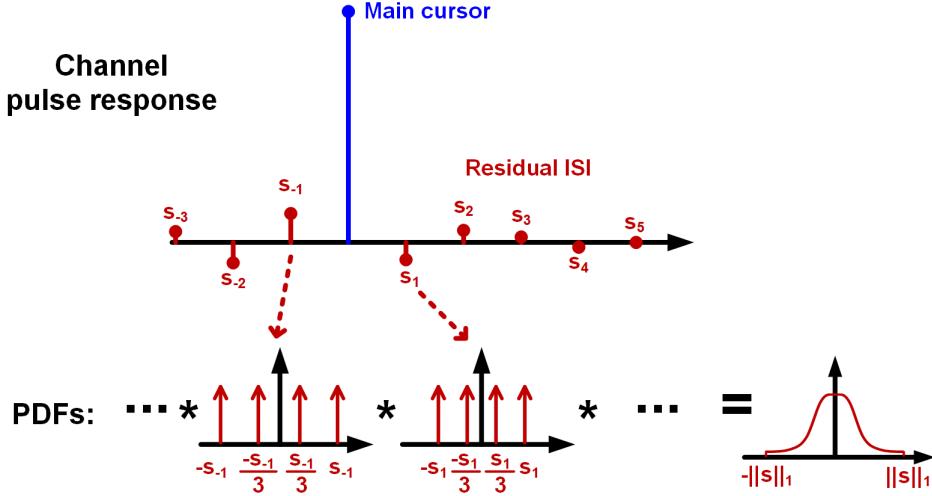


Figure 5.7: Numerical calculation of residual ISI PDF by convolution

Since SID directly outputs a estimated channel pulse response, residual ISI cursors are all known after equalization. The function of the DFE can be simply modeled as subtracting a certain value from desired cursor locations. For example, the digital 1-tap DFE of coefficient $h_{1,DFE}$ is subtracted from the first post-cursor value of \hat{h} , thus the residual ISI at the first post-cursor location is $\hat{h}_1 - h_{1,DFE}$. Therefore, the channel estimate of the SID engine allows us to explicitly calculate a residual ISI PDF as one component of the BER estimation.

5.2.2 Conditional PDFs for circuit errors

The second component of the BER estimation comes from the SID's error output \hat{e} . Different from residual ISI whose PDF needs to be numerically calculated, the circuit error PDFs can be estimated from the histogram of the \hat{e} samples. However, previous chapters have discussed the data dependent nature of various error sources in a PAM4 link, especially nonlinearity. As a result, we need to make sure each error sample is aligned with the correct data to generate conditional PDFs.

Figure 5.8 shows example conditional PDFs generated from a data set when the ADC is clipped. As expected, SID is able to capture clipping events, which are manifested as quantization error function extensions as discussed before. In the PDF domain, we see that these clipping events show up as the long tails for $D = \pm 1$ PDFs, while not affecting $D = \pm 1/3$ PDFs at all. Now, since we have already established how to obtain the conditional circuit error PDFs from SID output errors, the BER estimation methods will be presented next.

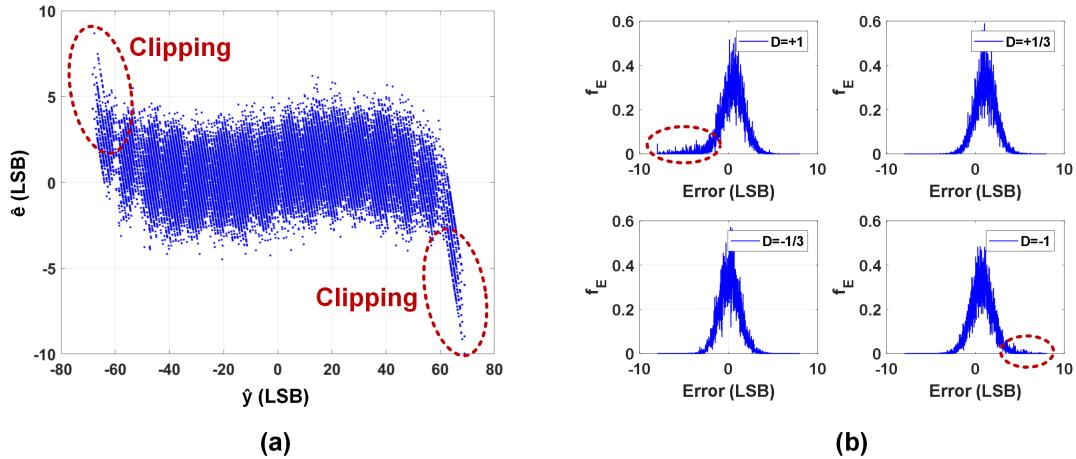


Figure 5.8: Clipped ADC SID outputs. (a) Error scatter plot and (b) conditional PDFs from error histograms

5.2.3 BER estimation method and results

The complete SID and BER estimation flow is visually summarized in Figure 5.9 and outlined below

1. Record the converged FFE, DFE coefficients from DSP and collect ADC output snapshots from silicon. Run SID with FFE in the loop to obtain the equivalent channel and filtered error estimates at the output of FFE.

2. Subtract DFE coefficient from the first post-cursor in channel estimate. Extract the residual ISI vector (by excluding the main cursor from the channel estimate after DFE), and numerically calculate residual ISI PDF.
3. Extract the four circuit error conditional PDFs from the SID output error samples.
4. Convolve the residual ISI PDF with each of the conditional PDFs to obtain four final conditional error PDFs, which will be used for tail integration.
5. An interpolation PDF (e.g., Gaussian with a small σ) could be applied to smooth out the PDFs.
6. Finally, BER is estimated by integrating the tails of each error PDF up to $\pm \hat{h}/3$.

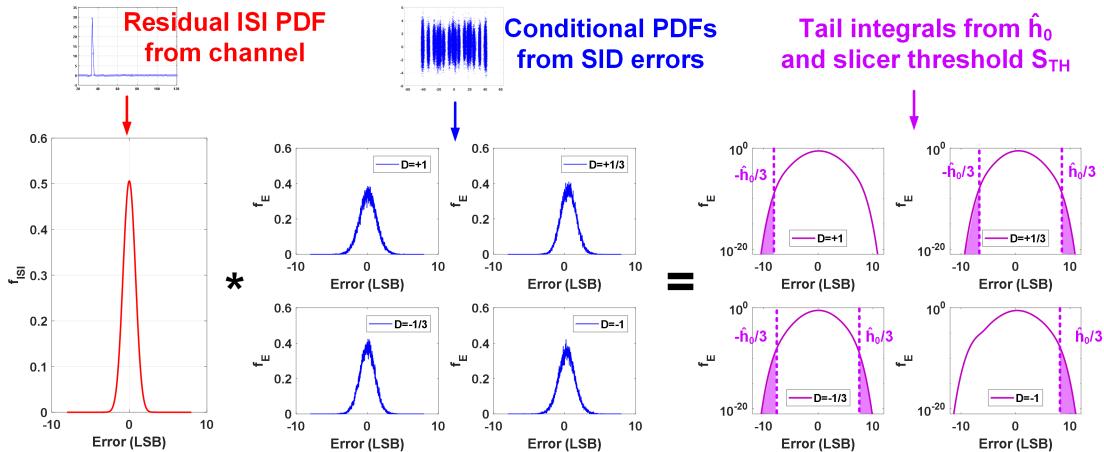


Figure 5.9: BER estimation flow from SID results

In order to compare BER estimation accuracy, both BER when using PRBS13 and PRBS31 are measured. Besides, the $Q(x)$ BER estimation curve is also calculated by using $1/(PAM - 1) \times \hat{h}_0/\sigma$ as the SNR. The noise standard deviation σ is directly calculated from residual ISI and SID error power. Figure 5.10 shows the measured BER without crosstalk and estimated BER curves using different methods with 7b

ADC data. The measured BER curves seem to be bounded by the two estimation

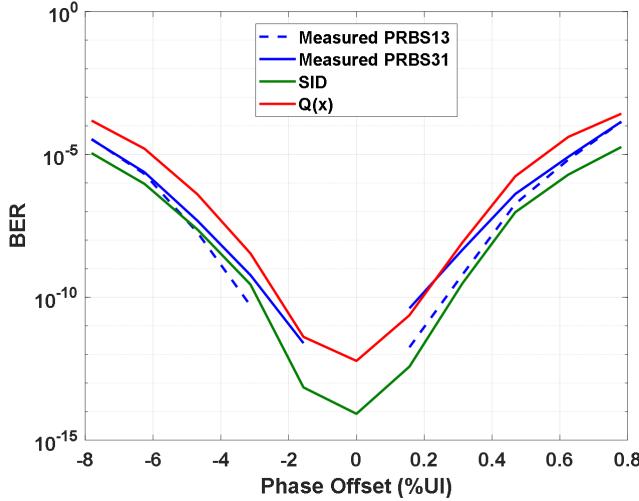


Figure 5.10: BER estimation comparisons with measured BER with 7b data

methods. SID tends to be an underestimation due to the limited statistics collected with finite number of data, and the $Q(x)$ method overestimates due to the wrong assumption that every noise source is Gaussian distributed. However, one might argue that the amount of overestimation in the Gaussian method is not significant in this case. This is mainly due to the fact that quantization error is not the dominant source when using a relatively high resolution ADC (7 bits), in which case other noises can be reasonably modeled as Gaussian noise. Larger discrepancies will be seen for lower resolution ADCs.

5.2.4 BER prediction for lower resolution ADCs

In Section 3.2, we have discussed how the quantization error PDF propagates through an FFE. Therefore, an extra quantization PDF can be numerically calculated if ADC resolution were lowered. The transceiver has reconfigurable ADCs whose resolution can be programmed by truncating the LSBs. For example, in 5 bit mode, the ADC outputs' last two LSBs are masked and set always to zero. Such truncation has two

major effects: 1. there will be a global offset due to the intrinsic floor operation in truncation. When the DSP settings are frozen, this offset will not be canceled. This offset will also propagate through the FFE as a DC value, and scaled by the sum of all FFE coefficients; 2. the new quantization error PDF becomes a wider uniform box and filtered after FFE according to the mechanism described before. Combining these two facts, new error PDFs can be devised to predict BER when using lower resolution ADCs. Figure 5.11 summarizes how the lower resolution quantization PDF passes through the FFE and is used for BER prediction

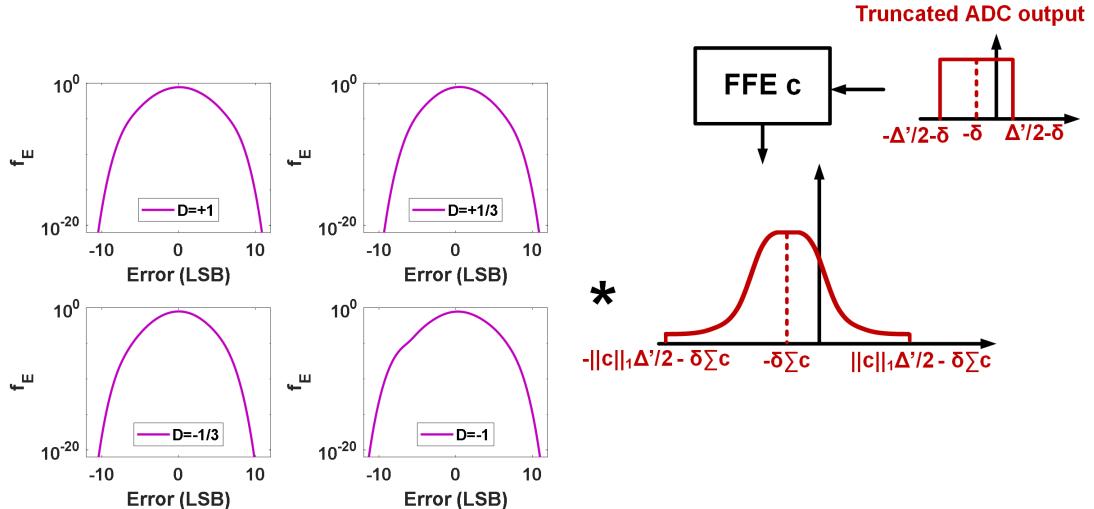


Figure 5.11: Truncated ADC output quantization PDF propagation through FFE

Figure 5.12 plots again the estimated and measured BER curves but for 6b and 5b ADC samples. We see that the SID estimated BER has much higher prediction accuracy compared to $Q(x)$ method. There is a consistent one decade of BER overestimation when Gaussian noise is assumed. This also shows the validity of quantization uniform PDF model. The difference between the two methods narrow again for 5b ADC since the BER performance degraded, but the overestimation trend of $Q(x)$ method is still there. SID BER estimation overall serves as a better method, especially for the moderate ADC resolution case. From these plots we can conclude

that 6b ADCs are good candidates for PAM4 links due to their moderate degradation from the 7b case, echoing the theoretical requirements derived in Section 3.2 as well.

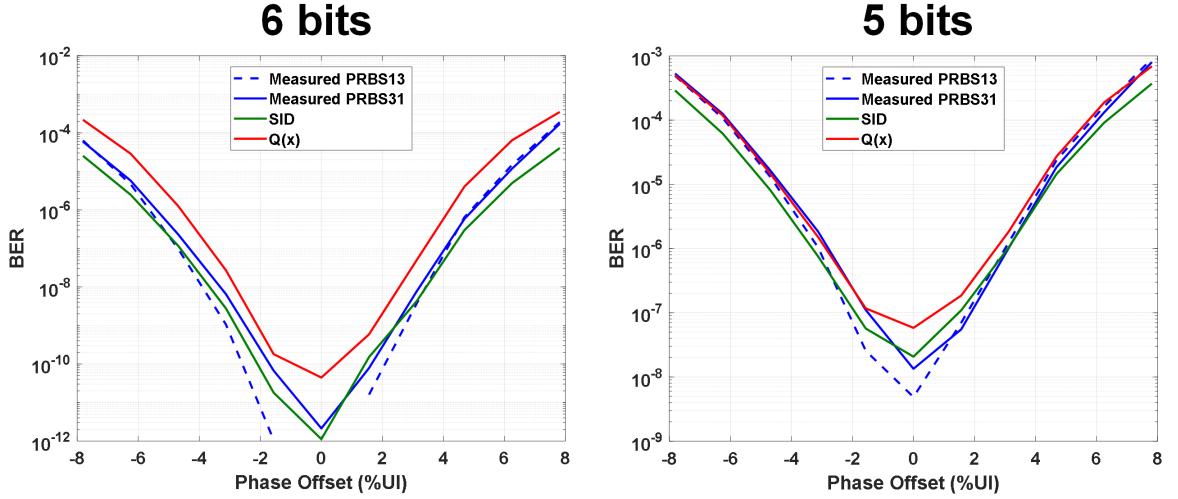


Figure 5.12: BER estimation comparisons with measured BER with 6b and 5b data

5.3 Pre-equalization and ADC resolution trade-off

One of the main points of this work is realizing and analyzing the trade-off between pre-ADC equalization and ADC's resolution. We have shown that PMR serves as an important metric and directly affects ADC requirements. SID is a valuable tool in extracting the channel information, which can be used to calculate channel PMR exactly. In this section, an experiment is presented to validate and quantify this trade-off, and implications will be discussed.

The previous sections illustrated BER performance when the equalizer settings are fixed and ADC resolution is changed. Here, we use a different setup in which both pre-equalizer's settings and ADC resolution are changed and allow the DSP to adapt. Since the 7b ADC drastically outperforms 5b ADC, the only fair comparison

to make in order to visualize this trade-off is by first finding the pre-equalizer settings such that the 7b ADC's BER curve is similar to that of the 5b ADC in Figure 5.11. This provides us with a BER performance baseline, and the 10^{-8} BER point is also a reasonable target specification. Figure 5.13 shows the BER curves for the three ADC resolutions when pre-equalizer setting is set to 10% of maximum high frequency peaking and 50% of maximum low frequency peaking. As expected, the 5b ADC performance is much worse even when DSP adapts.

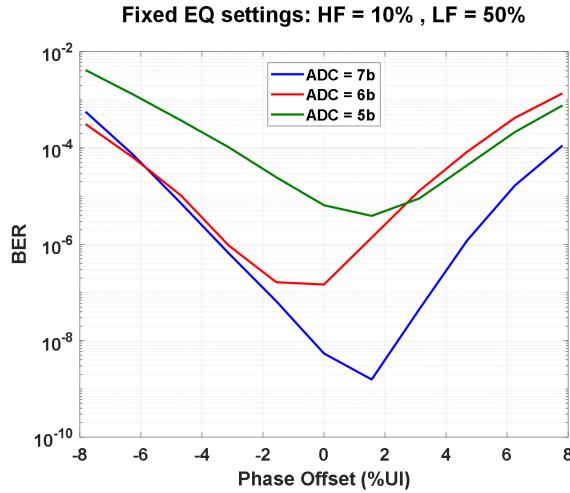


Figure 5.13: BER curves for different ADC resolutions at fixed pre-equalizer settings

The next step is to increase the pre-equalizer strength for lower resolution ADCs and see how the BER curves respond. Figure 5.14(a) shows the BER curves when larger HF peaking is applied for lower resolution ADCs. By pre-equalizing more in front of the ADC, a 5b ADC system almost has the same performance as a 7b ADC system without too much pre-equalization. Again, this trade-off exists because of the PMR reduction of the equivalent channel before the ADC. Figure 5.14(b) plots the SID estimated channels for the three ADC resolutions at optimal sampling phase. We see that the PMRs only improved slightly (about 14% from 7b to 6b and 36% from 7b to 5b) to reduce the number bits for similar performance. Instead of requiring a strict

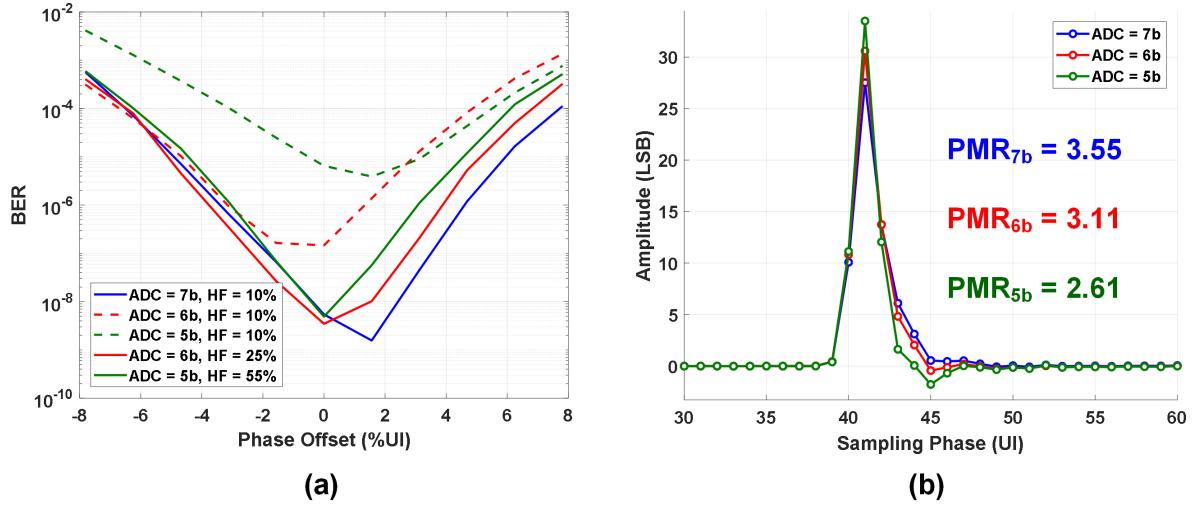


Figure 5.14: (a) BER curves for different ADC resolutions at different pre-equalizer settings and (b) SID estimated channels and PMRs

2x improvement in PMR, the presence of other noise and higher BER specifications allow the ADC resolution to be relaxed with less amount of PMR reduction. This implies that the effect of pre-equalization is even more pronounced in a realistic system and the trade-off between pre-equalizers and ADC should definitely be exploited in future ADC-based links. The measured raw BER for these settings also meets most backplane standard specifications, which means that a PAM4 link using effective pre-equalizers and a 5b ADC could be a viable solution. This widens the playing field in terms of AFE and ADC circuit designs since flash ADCs become attractive at 5 bits as more challenges arise for even higher data rate.

Chapter 6

Conclusions and Future Work

6.1 Summary and Conclusions

As ADC-based data links become more popular for next-generation serial systems with $>56\text{Gbps}$ data rate and multi-level modulation, a more fundamental understanding of the ADC's and equalizers' effects in a link context is needed. It is no longer enough to design a wireline link on a block basis without properly analyzing the overall system impact, since such a design paradigm could lead to overdesign.

After brief overview of various link architectures and challenges in Chapter 1, we first presented a statistical framework in Chapter 2 for understanding the ADC non-idealities' effects on system performance. Each ADC error source, including quantization and nonlinearity could be modeled as an independent noise source with unique data dependent PDFs. Under this framework, we concluded that conventional BER estimation methods assuming Gaussian distributions will result in overdesign in terms of ADC resolution requirements. Analysis with conditional PDFs should become a standard practice due to the data-dependent nature of nonlinearity, especially in a PAM4 system. This also motivated a closer investigation into the receiver input PDFs' effects on linearity performance. In Chapter 3, this work studied the pros and

cons of different equalizer locations along the signal chain. With the insights developed from the statistical framework, the optimal partitioning of equalization power at different positions is discussed. More specifically, the importance of pre-equalization before the ADC is highlighted. By deriving a first order ADC resolution requirement equation, PMR as a metric is emphasized and discussed extensively with respect to relaxing ADC requirements.

Chapter 4 focused on the circuit implementations of such pre-equalizers. Conventional CML-based CTLEs face challenges in power and area efficiency due to their limited technology scaling and passive inductor area overhead. This work presented two inverter-based CTLEs fully embedded in a 56Gbps short-reach PAM2 and 56Gbps flexible-reach PAM4 transceivers. The performance of each transceiver is verified under different voltage and temperature conditions, demonstrating robustness of inverter-based AFEs in addition to the area and power benefits. The good linearity and crosstalk performance for the PAM4 ADC-based link were also highlighted.

The PAM4 ADC-based transceiver is used for further SID analysis in Chapter 5. Statistical BER estimation methods using SID output data is introduced and compared to the conventional Q-function approach. The trade-off between pre-equalizer and ADC resolution is finally validated with silicon data, demonstrating reasonable BER performance for <6b ADCs given current standards. Therefore, this work has shown the feasibility of an ADC-based link using efficient inverter-based pre-equalizers and low resolution ADCs.

6.2 Future Work

One of the hardest future challenges is to implement an area- and energy- efficient ADC-based link with 112Gbps data rate. Several advances need to be made to achieve this goal.

From the circuit side, the inverter-based AFE's bandwidth needs to immediately double if PAM4 signaling is still used. One possible solution is to bring passive inductors back and use them in parallel with the active inductor loads used in this work. The passive inductor area overhead will be mitigated due to the partial bandwidth boost from the active inductors, but the design requires more attention to details, such as parasitics sensitivity. In addition, more research effort can be put into finding better inverter-based equalizers that are more effective for reducing channel PMRs. CTLEs only deal with post-cursors, and pre-cursor canceling equalizers should be evaluated for next generation links. Different PVT stabilization techniques can also be investigated to push the performance of inverter-based circuits further.

From a system perspective, 112Gbps links will pose many new problems since errors due to timing and ADC metastability become quite significant. Therefore, these error sources need to be incorporated into the statistical framework and analyzed in the PDF domain. The nature of timing and metastability errors will heavily depend on the ADC topology. For example, massively time-interleaved SAR ADCs will have more difficulties with metastability and clocking compared to flash ADCs. Cross-domain equalization can also be studied as an alternative solution to reduce PMR. For example, instead of using LF CTLEs to cancel long tail ISI, analog IIR and floating tap DFEs with data from the digital domain can also cancel long tail and reflection ISIs. More sophisticated digital processing, adaptation and calibration algorithms can be developed to take advantage of process scaling.

Finally, new link architectures and application spaces should also be explored for ADC-based links. As copper interfaces approach their physical bandwidth limits, full-duplex ADC-based wireline links could be a viable solution to double the aggregate bandwidth with manageable cost [41, 42]. The use of ADC-based links over other communication channel media, such as optical fibers [27] and dielectric waveguides [43], also becomes a subject worth pursuing.

Appendix A

Pseudo-Independent Quantization Noise Proof

This work is based on the assumption that for most link systems, quantization error can be treated as an independent noise source. Here we use the sampling theory for quantization in [44] to prove the soundness of this assumption. We will also discuss how this relates to traditional ADC dithering theory such as [45].

Figure A.1 shows the model for this proof. The channel output x with added noise $\overline{v_n^2}$ is quantized to y . The problem can be formulated in the PDF domain as finding the error PDF $f_E(e)$ for random variable $E = Y - X$ given input PDF $f_X(x)$, a Gaussian distributed thermal noise source and quantization. To find $f_E(e)$, we use the marginal PDF or joint PDF of E and X as follows

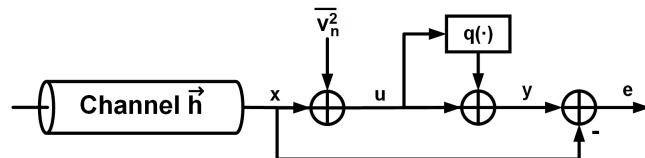


Figure A.1: Model for quantization's pseudo-independence

$$f_E(e) = \int_{-\infty}^{+\infty} f_X(x) f_{E|X}(e|x) dx \quad (\text{A.1})$$

Since the relationship between Y , E and X is deterministic, the above equation is also equivalent to

$$f_E(e) = \int_{-\infty}^{+\infty} f_X(x) f_{Y|X}(x + e|x) dx \quad (\text{A.2})$$

Therefore, we need to find the conditional PDF $f_{Y|X}$ and this can be accomplished using the quantization sampling theory in [44]. Figure A.2 illustrates quantization as interval sampling of a PDF. To obtain $f_{Y|X}(y|x)$, the quantizer converts a continuous

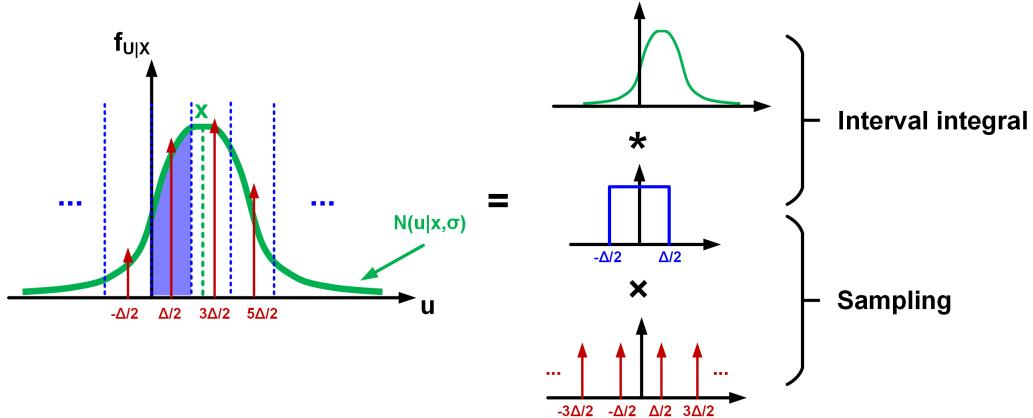


Figure A.2: Quantization as PDF interval sampling

input PDF $f_{U|X}(u|x)$ (a Gaussian distribution with mean at value x) into a PDF with dirac delta impulses $\delta(x)$ (continuous representation of a PMF). Each delta impulse's height is determined by the interval integral of the input PDF within the $\pm\Delta/2$ bound. Thus, $f_{Y|X}(y|x)$ can be written as

$$f_{Y|X}(y|x) = (\Delta f_Q * f_V)(y - x) \cdot \sum_{k=-\infty}^{+\infty} \delta\left(y - k\Delta + \frac{\Delta}{2}\right) \quad (\text{A.3})$$

f_Q is the uniform PDF $\Pi(x|\Delta)$ presented in Chapter 2, so that Δf_Q becomes a interval

integral when convolved. f_V is a Gaussian function as discussed before. By plugging A.3 into A.2, we obtain the following

$$\begin{aligned}
f_E(e) &= \int_{-\infty}^{+\infty} f_X(x) f_{Y|X}(x + e|x) dx \\
&= \int_{-\infty}^{+\infty} f_X(x) \cdot (\Delta f_Q * f_V)(x + e - x) \cdot \sum_{k=-\infty}^{+\infty} \delta\left(x + e - k\Delta - \frac{\Delta}{2}\right) dx \\
&= (f_Q * f_V)(e) \int_{-\infty}^{+\infty} f_X(x) \cdot \Delta \sum_{k=-\infty}^{+\infty} \delta\left(x + e - k\Delta - \frac{\Delta}{2}\right) dx \\
&= (f_Q * f_V)(e) \sum_{k=-\infty}^{+\infty} \Delta \cdot \int_{-\infty}^{+\infty} f_X(x) \delta\left(x + e - k\Delta - \frac{\Delta}{2}\right) dx \\
&= (f_Q * f_V)(e) \sum_{k=-\infty}^{+\infty} \Delta \cdot f_X\left(k\Delta + \frac{\Delta}{2} - e\right)
\end{aligned} \tag{A.4}$$

This is an interesting result since the total error PDF $f_E(e)$ is almost the convolution of f_Q and f_V . The summation term is the extra term that brings in the dependent nature of the total error. However, when certain conditions are met, this expression can be simplified to $f_Q * f_V$.

1. When Δ approaches zero (infinitely high resolution ADC), the summation term asymptotically become an integral

$$\lim_{\Delta \rightarrow 0} \sum_{k=-\infty}^{+\infty} \Delta \cdot f_X\left(k\Delta + \frac{\Delta}{2} - e\right) = \int_{-\infty}^{+\infty} f_X\left(x + \frac{\Delta}{2} - e\right) dx = 1 \tag{A.5}$$

This means that the summation term itself approaches value 1, therefore $f_E(e) = (f_Q * f_V)(e)$, which means quantization becomes an independent noise source.

2. The summation term is effectively a Riemann sum approximation of the area under f_X , given e as an offset in the argument. Therefore, the condition when Δ approaches zeros results in the area under the PDF f_X . The other case when

this Riemann sum approximates the total area under f_X well (regardless of the value of e) is for f_X to be relatively flat within any quantization bin. This is also known as local uniformity. Local uniformity can be achieved by reducing quantization bin size (thus the same as 1.) or quantizing an input PDF that has a wide spread or close to uniform distribution. In a link-context, the input PDF contains a large spread due to channel ISI, and the use of PAM4 further turns the input PDF into more of a uniform shape. Therefore, we conclude again that independent quantization noise is a safe assumption. A similar proof of quantization's "whiteness" with local uniformity is also presented in [11, 46].

The survey in [45] also examined the independent nature of quantization noise in the presence of dither. The conclusion was that for certain dither statistics, quantization can be treated as an independent noise source. More specifically, the dither signal's characteristic function, defined as the Fourier transform of its PDF, needs to be zero at integer multiples of $2\pi/\Delta$ (Schuchman's Condition). This echos our own proof since such an input PDF will exhibit uniform like shape or a wide spread compared to the quantization bin size. We will not go into details regarding the proof in [45], but we highlight that basically channel ISI serves as the "dither" signal in an ADC-based link to make quantization an pseudo-independent noise source.

Appendix B

Nonlinearity PDF

The solution to finding the PDF $f_Y(y)$ of random variable Y , given the PDF of random variable X to be $f_X(x)$, and $Y = -cX^3$ starts with finding the CDF of Y , $F_Y(y)$ and its relationship with $F_X(x)$.

$$\begin{aligned}
 F_Y(y) &= P(Y \leq y) \\
 &= P(-cX^3 \leq y) \\
 &= P\left(X \geq \sqrt[3]{-\frac{y}{c}}\right) \\
 F_Y(y) &= 1 - F_X\left(\sqrt[3]{-\frac{y}{c}}\right)
 \end{aligned} \tag{B.1}$$

Any random variable's PDF is the derivative of its CDF, thus we can now find $f_Y(y)$ by differentiating the equation above

$$\begin{aligned}
 f_Y(y) &= \frac{d}{dy}F_Y(y) = \frac{d}{dy}\left(1 - F_X\left(\sqrt[3]{-\frac{y}{c}}\right)\right) \\
 &= -f_X\left(\sqrt[3]{-\frac{y}{c}}\right) \cdot \frac{d}{dy}\left(\sqrt[3]{-\frac{y}{c}}\right) \\
 f_Y(y) &= \frac{1}{3c}\left(-\frac{y}{c}\right)^{-2/3}f_X\left(\sqrt[3]{-\frac{y}{c}}\right)
 \end{aligned} \tag{B.2}$$

Appendix C

ADC Resolution Requirement

The reason for a more accurate derivation is to capture the edge case when the eyes are wide open and the ADC threshold levels essentially just become decision slicers. When only 1 bit is needed per eye and channel effects don't add any bits, the first-order equation gives 2.6 bits instead of the theoretical 2 bits for PAM4. The main cause for this discrepancy lies in the multiplication of number of eyes when solving for the number of quantizer levels. For example, if L ADC levels are needed within an eye, it is not exactly $3L$ levels for 3 eyes. Two of the levels will overlap, therefore only $3N - 2$ levels are required. To generalize this, the following equation describes the number of levels L required given the PAM scheme and B_{eye} .

$$L = (\text{PAM} - 1) \cdot 2^{B_{eye}} - (\text{PAM} - 2) = (\text{PAM} - 1)(2^{B_{eye}} - 1) + 1 \quad (\text{C.1})$$

Accounting for the channel and FFE degradation as described before, the required ADC resolution B becomes

$$\begin{aligned} B &= \log_2 L + B_{channel} \\ &= \log_2 ((\text{PAM} - 1)(2^{B_{eye}} - 1) + 1) + B_{channel} \end{aligned} \quad (\text{C.2})$$

For the extreme case when the channel is ideal (i.e. $B_{eye} = 1$ and $B_{channel} = 0$), the expression above reduces to \log_2 PAM, which is the correct resolution when the ADC is used as decision slicers. For the purpose of most realistic applications, the first-order equation provides reasonable approximation to ADC's required resolution.

Appendix D

Inverter-based Active Inductor Impedance

Assuming the inverter's output impedance is large enough to be ignored, the current i_t flowing into the active inductor with a test voltage source v_t (shown in Figure D.1) is given by

$$i_t = G_m v_i + \frac{v_t - v_i}{R} = \frac{v_t}{R} + \frac{G_m R - 1}{R} v_i \quad (\text{D.1})$$

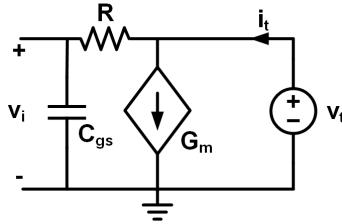


Figure D.1: Test voltage to find active inductor impedance

v_i is determined by v_t low passed by the feedback resistor R and inverter gate capacitance C_{gs} .

$$v_i = v_t \frac{1}{1 + sRC_{gs}} \quad (\text{D.2})$$

Substituting v_i gives the following expression for i_t and v_t

$$\begin{aligned}
 i_t &= \frac{v_t}{R} + \frac{G_m R - 1}{R} \frac{1}{1 + sRC_{gs}} v_t \\
 &= \frac{1 + sRC_{gs} + G_m R - 1}{R(1 + sRC_{gs})} v_t \\
 &= \frac{G_m R + sRC_{gs}}{R(1 + sRC_{gs})} v_t \\
 &= G_m \frac{1 + sC_{gs}/G_m}{1 + sRC_{gs}} v_t
 \end{aligned} \tag{D.3}$$

Therefore, the impedance of the active inductor Z_L is derived to be

$$Z_L = \frac{v_t}{i_t} = \frac{1}{G_m} \frac{1 + sRC_{gs}}{1 + sC_{gs}/G_m} \tag{D.4}$$

Appendix E

Inverter-based Active Inductor Buffer

The transfer function of a unity gain buffer with active inductor load can be derived by finding the effective load impedance first. The total load impedance is the parallel combination of active inductor impedance (derived in Appendix D) and load capacitance C_L .

$$\begin{aligned} Z_{tot} &= \frac{1}{sC_L} // Z_L \\ &= \frac{1}{sC_L + G_m \frac{1+sC_{gs}/G_m}{1+sRC_{gs}}} \\ &= \frac{1 + sRC_{gs}}{sC_L(1 + sRC_{gs}) + G_m(1 + sC_{gs}/G_m)} \\ &= \frac{1}{G_m} \frac{1 + sRC_{gs}}{1 + s(C_{gs} + C_L)/G_m + s^2 RC_L C_{gs}/G_m} \end{aligned} \tag{E.1}$$

The transfer function then simply becomes

$$G_{ind}(s) = G_m Z_{tot} = \frac{1 + sRC_{gs}}{1 + s \frac{C_{gs} + C_L}{G_m} + s^2 \frac{RC_L C_{gs}}{G_m}} \tag{E.2}$$

Appendix F

Inverter-based Active Inductor Buffer Noise

The noise PSD for an active inductor unity buffer can be found by superposition. First, let's consider the device noise from both the active and load transconductors. Since they are in parallel in the small-signal model, their noise contribution can be summed together. The noise current is converted to output noise voltage through the total output impedance found in Appendix E.

$$\frac{\sqrt{v_o^2}}{\sqrt{i_{n,g_m}^2}} = Z_{tot} = \frac{1}{G_m} \frac{1 + sRC_{gs}}{1 + s(C_{gs} + C_L)/G_m + s^2RC_LC_{gs}/G_m} \quad (\text{F.1})$$

The transfer function from the feedback resistor's noise current to output voltage can be derived from the simplified small signal noise model in Figure F.1. v_e can be found by superposition

$$v_e = \frac{1}{1 + sRC_{gs}} v_o - \frac{R}{1 + sRC_{gs}} \sqrt{i_{n,R}^2} \quad (\text{F.2})$$

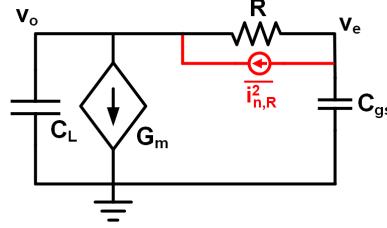


Figure F.1: Small signal model for resistor noise

Then v_o can be derived from the following node equation

$$\begin{aligned}
 sC_Lv_o + G_mv_e + \frac{v_o - v_e}{R} - \sqrt{i_{n,R}^2} &= 0 \\
 \frac{1 + sRC_L}{R}v_o + \frac{G_mR - 1}{R}v_e - \sqrt{i_{n,R}^2} &= 0 \\
 \frac{1 + sRC_L}{R}v_o + \frac{G_mR - 1}{R} \left(\frac{1}{1 + sRC_{gs}}v_o - \frac{R}{1 + sRC_{gs}}\sqrt{i_{n,R}^2} \right) - \sqrt{i_{n,R}^2} &= 0
 \end{aligned} \tag{F.3}$$

Rearranging and combining terms gives the relationship between v_o and $\sqrt{i_{n,R}^2}$ to be

$$\frac{G_m + s(C_{gs} + C_L) + s^2 RC_{gs} C_L}{1 + sRC_{gs}} v_o = \frac{G_m R + sRC_{gs}}{1 + sRC_{gs}} \sqrt{i_{n,R}^2} \tag{F.4}$$

Therefore, the resistor noise contribution to the final output noise is

$$\frac{\sqrt{v_o^2}}{\sqrt{i_{n,R}^2}} = R \frac{1 + sC_{gs}/G_m}{1 + s \frac{C_{gs} + C_L}{G_m} + s^2 \frac{RC_{gs} C_L}{G_m}} \tag{F.5}$$

Finally, by superposition again, we combine the noise contributions from both the

devices and resistor and find the total output noise PSD to be

$$\begin{aligned} \frac{\overline{v_o^2}}{\Delta f} = & 8kT\gamma G_m \left| \frac{1}{G_m} \frac{1 + sRC_{gs}}{1 + s\frac{C_L+C_{gs}}{G_m} + s^2\frac{RC_LC_{gs}}{G_m}} \right|^2 \\ & + \frac{4kT}{R} \left| R \frac{1 + sC_{gs}/G_m}{1 + s\frac{C_L+C_{gs}}{G_m} + s^2\frac{RC_LC_{gs}}{G_m}} \right|^2 \end{aligned} \quad (\text{F.6})$$

in which $8kT\gamma G_m$ is noise current from both the two G_m 's, and $4kT/R$ is the noise current from the resistor. Both noise transfer functions take the form

$$\frac{1 + \frac{s}{\omega_z}}{1 + \frac{s}{\omega_n Q} + \frac{s^2}{\omega_n^2}} \quad (\text{F.7})$$

and there is a closed form equation for such noise PSD integrals [47]

$$\int_0^{+\infty} \left| \frac{1 + \frac{s}{\omega_z}}{1 + \frac{s}{\omega_n Q} + \frac{s^2}{\omega_n^2}} \right|^2 df = \frac{\omega_n Q}{4} \left(1 + \frac{\omega_n^2}{\omega_z^2} \right) \quad (\text{F.8})$$

Using this closed form integral, we will be able to find the total integrated noise for our unity gain buffer as follows

$$\begin{aligned} \overline{v_o^2} = & \frac{8kT\gamma}{G_m} \int_0^{+\infty} \left| \frac{1 + sRC_{gs}}{1 + s\frac{C_L+C_{gs}}{G_m} + s^2\frac{RC_LC_{gs}}{G_m}} \right|^2 df \\ & + 4kTR \int_0^{+\infty} \left| R \frac{1 + sC_{gs}/G_m}{1 + s\frac{C_L+C_{gs}}{G_m} + s^2\frac{RC_LC_{gs}}{G_m}} \right|^2 df \\ = & \frac{2kT\gamma}{G_m} \frac{G_m}{C_L + C_{gs}} \left(1 + \frac{G_m RC_{gs}}{C_L} \right) + kTR \frac{G_m}{C_L + C_{gs}} \left(1 + \frac{C_{gs}}{G_m RC_L} \right) \\ = & 2kT\gamma \frac{1 + G_m RC_{gs}/C_L}{C_L + C_{gs}} + kT \frac{G_m R + C_{gs}/C_L}{C_L + C_{gs}} \end{aligned} \quad (\text{F.9})$$

Appendix G

Inverter-based Buffer DC Characteristics

Figure G.1 shows an inverter buffer with each device's DC current annotated. To find the large signal relationship between V_i and V_o assuming all devices stay in saturation, the total PMOS current need to be equal to the total NMOS current.

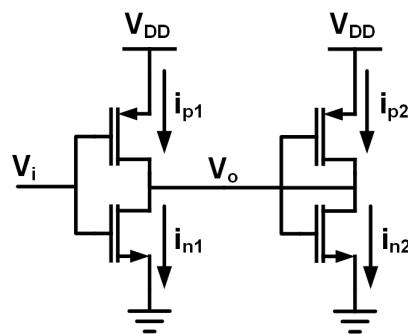


Figure G.1: DC currents in an inverter buffer

$$i_{p1} + i_{p2} = i_{n1} + i_{n2} \quad (\text{G.1})$$

G.0.1 Square law device

If we use square law devices, we can arrive at the following equation

$$\begin{aligned} \mu_p C_{ox} \frac{W_p}{L_p} & \left((V_{DD} - V_i - V_{Tp})^2 (1 + \lambda(V_{DD} - V_o)) + (V_{DD} - V_o - V_{Tp})^2 (1 + \lambda(V_{DD} - V_o)) \right) \\ & = \mu_n C_{ox} \frac{W_n}{L_n} \left((V_i - V_{Tn})^2 (1 + \lambda V_o) + (V_o - V_{Tn})^2 (1 + \lambda V_o) \right) \end{aligned} \quad (\text{G.2})$$

In 16nm FinFET CMOS, we can assume the PMOS and NMOS device mobilities and threshold voltages are the same and the inverter buffer is designed to be fully symmetric. Therefore, this equation can be simplified to

$$\begin{aligned} & (V_{DD} - V_i - V_T)^2 (1 + \lambda(V_{DD} - V_o)) + (V_{DD} - V_o - V_T)^2 (1 + \lambda(V_{DD} - V_o)) \\ & = (V_i - V_T)^2 (1 + \lambda V_o) + (V_o - V_T)^2 (1 + \lambda V_o) \end{aligned} \quad (\text{G.3})$$

At first glance, this is a 3rd order equation and finding a closed form solution for V_o in terms of V_i with many parameters (V_{DD} , V_T , λ) might not give too much insight. We make another simplification again by letting $\lambda = 0$, which means that the output impedance of our devices are large enough to ignore V_{ds} effects. Therefore, now we have

$$(V_{DD} - V_i - V_T)^2 + (V_{DD} - V_o - V_T)^2 = (V_i - V_T)^2 + (V_o - V_T)^2 \quad (\text{G.4})$$

By inspection, the solution to this equation is simply

$$V_o = V_{DD} - V_i \quad (\text{G.5})$$

which means that such devices stay perfectly linear if they are in saturation and V_{ds} effects are negligible. In fact, it doesn't matter if these devices obey perfect square

laws, as long as PMOS and NMOS in an inverter stay symmetrical and follow these assumptions, the large signal transfer function of the inverter buffer will be quite linear.

G.0.2 Velocity saturated device

Modern deep sub-micron devices suffer from velocity saturation. Therefore the saturation current of such devices then follows a nearly linear law, which is given by

$$I_{d,sat} = v_{sat}WC_{ox}(V_{gs} - V_T)(1 + \lambda V_{ds}) \quad (\text{G.6})$$

By using this model, (G.3) can be reduced to a second order equation, which we could solve by hand. Assuming symmetry in the inverter again, we obtain the following

$$\begin{aligned} & (V_{DD} - V_i - V_T)(1 + \lambda(V_{DD} - V_o)) + (V_{DD} - V_o - V_T)(1 + \lambda(V_{DD} - V_o)) \\ &= (V_i - V_T)(1 + \lambda V_o) + (V_o - V_T)(1 + \lambda V_o) \end{aligned} \quad (\text{G.7})$$

The solution to this equation is the following

$$V_o = \frac{V_{DD}(2\lambda V_{DD} - 2\lambda V_T + 2)}{3\lambda V_{DD} - 4\lambda V_T + 2} - \frac{\lambda V_{DD} + 2}{3\lambda V_{DD} - 4\lambda V_T + 2} V_i \quad (\text{G.8})$$

It is also interesting to see that V_o and V_i have a perfectly linear relationship except that the slope is a function of V_{DD} , V_T and λ . When $\lambda = 0$, we will arrive at the same conclusion as before and the slope is exactly negative one. V_{DD} and V_T also play a role in this slope, but their affects are much smaller due to the multiplication by λ .

Bibliography

- [1] D. C. Daly, L. C. Fujino, and K. C. Smith. Through the Looking Glass - The 2018 Edition: Trends in Solid-State Circuits from the 65th ISSCC. *IEEE Solid-State Circuits Magazine*, 10(1):30–46, winter 2018. ISSN 1943-0582. doi: 10.1109/MSSC.2017.2771103.
- [2] IEEE Standard for Ethernet Amendment 10: Media Access Control Parameters, Physical Layers, and Management Parameters for 200 Gb/s and 400 Gb/s Operation. *IEEE Std. 802.3bs*, 2017.
- [3] Common Electrical I/O (CEI) - Electrical and Jitter Interoperability agreements for 6G+ bps, 11G+ bps, 25G+ bps I/O and 56G+ bps. *OIF CEI 4.0*, 2017.
- [4] J. F. Bulzacchelli. Equalization for Electrical Links: Current Design Techniques and Future Directions. *IEEE Solid-State Circuits Magazine*, 7(4):23–31, Fall 2015. ISSN 1943-0582. doi: 10.1109/MSSC.2015.2475996.
- [5] C. K. K. Yang and E. H. Chen. ADC-based serial I/O receivers. In *2009 IEEE Custom Integrated Circuits Conference*, pages 323–330, Sept 2009. doi: 10.1109/CICC.2009.5280831.
- [6] J. Kim, E. H. Chen, J. Ren, B. S. Leibowitz, P. Satarzadeh, J. L. Zerbe, and C. K. K. Yang. Equalizer Design and Performance Trade-Offs in ADC-Based

- Serial Links. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 58(9):2096–2107, Sept 2011. ISSN 1549-8328. doi: 10.1109/TCSI.2011.2162465.
- [7] I. S. Reed and G. Solomon. Polynomial Codes Over Certain Finite Fields. *Journal of the Society for Industrial and Applied Mathematics*, 8(2):300–304, 1960. URL <http://dx.doi.org/10.1137/0108018>.
 - [8] M. Harwood, N. Warke, R. Simpson, T. Leslie, A. Amerasekera, S. Batty, D. Colman, E. Carr, V. Gopinathan, S. Hubbins, P. Hunt, A. Joy, P. Khandelwal, B. Killips, T. Krause, S. Lytollis, A. Pickering, M. Saxton, D. Sebastio, G. Swanson, A. Szczepanek, T. Ward, J. Williams, R. Williams, and T. Willwerth. A 12.5Gb/s SerDes in 65nm CMOS Using a Baud-Rate ADC with Digital Receiver Equalization and Clock Recovery. In *2007 IEEE International Solid-State Circuits Conference. Digest of Technical Papers*, pages 436–591, Feb 2007. doi: 10.1109/ISSCC.2007.373481.
 - [9] J. Cao, B. Zhang, U. Singh, D. Cui, A. Vasani, A. Garg, W. Zhang, N. Kocaman, D. Pi, B. Raghavan, H. Pan, I. Fujimori, and A. Momtaz. A 500 mW ADC-Based CMOS AFE With Digital Calibration for 10 Gb/s Serial Links Over KR-Backplane and Multimode Fiber. *IEEE Journal of Solid-State Circuits*, 45(6):1172–1185, June 2010. ISSN 0018-9200. doi: 10.1109/JSSC.2010.2047473.
 - [10] B. Murmann. ADC Performance Survey 1997-2018. URL <https://web.stanford.edu/~murmann/adcsurvey.html>.
 - [11] A. Sripad and D. Snyder. A necessary and sufficient condition for quantization errors to be uniform and white. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 25(5):442–448, Oct 1977. ISSN 0096-3518. doi: 10.1109/TASSP.1977.1162977.

- [12] W. R. Bennett. Spectra of quantized signals. *The Bell System Technical Journal*, 27(3):446–472, July 1948. ISSN 0005-8580. doi: 10.1002/j.1538-7305.1948.tb01340.x.
- [13] Maxim Integrated. Tutorial 1040: Coherent Sampling vs. Window Sampling. URL <https://www.maximintegrated.com/en/app-notes/index.mvp/id/1040>.
- [14] V. Stojanovic and M. Horowitz. Modeling and analysis of high-speed links. In *Proceedings of the IEEE 2003 Custom Integrated Circuits Conference, 2003.*, pages 589–594, Sept 2003. doi: 10.1109/CICC.2003.1249467.
- [15] Leo K. Wong. Analyzing High-Speed Serial Links (Rambus). URL <https://www.design-reuse.com/articles/8740/analyzing-high-speed-serial-links-rambus.html>.
- [16] Yu Chang, Dan Oh, and C. Madden. Jitter modeling in statistical link simulation. In *2008 IEEE International Symposium on Electromagnetic Compatibility*, pages 1–4, Aug 2008. doi: 10.1109/ISEMC.2008.4652155.
- [17] E. H. Chen and C. K. K. Yang. ADC-Based Serial I/O Receivers. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 57(9):2248–2258, Sept 2010. ISSN 1549-8328. doi: 10.1109/TCSI.2010.2071431.
- [18] E. H. Chen, R. Yousry, and C. K. K. Yang. Power Optimized ADC-Based Serial Link Receiver. *IEEE Journal of Solid-State Circuits*, 47(4):938–951, April 2012. ISSN 0018-9200. doi: 10.1109/JSSC.2012.2185356.
- [19] Y. Frans, M. Elzeftawi, H. Hedayati, J. Im, V. Kireev, T. Pham, J. Shin, P. Upadhyaya, Lei Zhou, S. Asuncion, C. Borrelli, G. Zhang, Hongtao Zhang, and K. Chang. A 56Gb/s PAM4 wireline transceiver using a 32-way time-interleaved

- SAR ADC in 16nm FinFET. In *2016 IEEE Symposium on VLSI Circuits (VLSI-Circuits)*, pages 1–2, June 2016. doi: 10.1109/VLSIC.2016.7573474.
- [20] T. O. Dickson, H. A. Ainspan, and M. Meghelli. A 1.8pJ/b 56Gb/s PAM-4 transmitter with fractionally spaced FFE in 14nm CMOS. In *2017 IEEE International Solid-State Circuits Conference (ISSCC)*, pages 118–119, Feb 2017. doi: 10.1109/ISSCC.2017.7870289.
- [21] G. Steffan, E. Depaoli, E. Monaco, N. Sabatino, W. Audoglio, A. A. Rossi, S. Erba, M. Bassi, and A. Mazzanti. 6.4 A 64Gb/s PAM-4 transmitter with 4-Tap FFE and 2.26pJ/b energy efficiency in 28nm CMOS FDSOI. In *2017 IEEE International Solid-State Circuits Conference (ISSCC)*, pages 116–117, Feb 2017. doi: 10.1109/ISSCC.2017.7870288.
- [22] C. Menolfi, M. Braendli, P. A. Francese, T. Morf, A. Cevrero, M. Kossel, L. Kull, D. Luu, I. Ozkaya, and T. Toifl. A 112Gb/S 2.6pJ/b 8-Tap FFE PAM-4 SST TX in 14nm CMOS. In *2018 IEEE International Solid - State Circuits Conference - (ISSCC)*, pages 104–106, Feb 2018. doi: 10.1109/ISSCC.2018.8310205.
- [23] J. Kim, A. Balankutty, R. Dokania, A. Elshazly, H. S. Kim, S. Kundu, S. Weaver, K. Yu, and F. O’Mahony. A 112Gb/s PAM-4 transmitter with 3-Tap FFE in 10nm CMOS. In *2018 IEEE International Solid - State Circuits Conference - (ISSCC)*, pages 102–104, Feb 2018. doi: 10.1109/ISSCC.2018.8310204.
- [24] A. Nazemi, K. Hu, B. Catli, D. Cui, U. Singh, T. He, Z. Huang, B. Zhang, A. Momtaz, and J. Cao. 3.4 A 36Gb/s PAM4 transmitter using an 8b 18GS/S DAC in 28nm CMOS. In *2015 IEEE International Solid-State Circuits Conference - (ISSCC) Digest of Technical Papers*, pages 1–3, Feb 2015. doi: 10.1109/ISSCC.2015.7062924.

- [25] R. Boesch. *Signal preconditioning using feedforward equalizers in ADC-based data links*. PhD thesis, Stanford University, 2014.
- [26] D. Cui, H. Zhang, N. Huang, A. Nazemi, B. Catli, H. G. Rhew, B. Zhang, A. Momtaz, and J. Cao. 3.2 A 320mW 32Gb/s 8b ADC-based PAM-4 analog front-end with programmable gain control and analog peaking in 28nm CMOS. In *2016 IEEE International Solid-State Circuits Conference (ISSCC)*, pages 58–59, Jan 2016. doi: 10.1109/ISSCC.2016.7417905.
- [27] S. Palermo, S. Hoyos, A. Shafik, E. Z. Tabasy, S. Cai, S. Kiran, and K. Lee. CMOS ADC-based receivers for high-speed electrical and optical links. *IEEE Communications Magazine*, 54(10):168–175, October 2016. ISSN 0163-6804. doi: 10.1109/MCOM.2016.7588288.
- [28] R. Boesch, K. Zheng, and B. Murmann. A 0.003 mm² 5.2 mW/tap 20 GBd inductor-less 5-tap analog RX-FFE. In *2016 IEEE Symposium on VLSI Circuits (VLSI-Circuits)*, pages 1–2, June 2016. doi: 10.1109/VLSIC.2016.7573522.
- [29] B. Nauta. A CMOS transconductance-C filter technique for very high frequencies. *IEEE Journal of Solid-State Circuits*, 27(2):142–153, Feb 1992. ISSN 0018-9200. doi: 10.1109/4.127337.
- [30] S. E. Liu, J. S. Wang, Y. R. Lu, D. S. Huang, C. F. Huang, W. H. Hsieh, J. H. Lee, Y. S. Tsai, J. R. Shih, Y. H. Lee, and K. Wu. Self-heating effect in FinFETs and its impact on devices reliability characterization. In *2014 IEEE International Reliability Physics Symposium*, pages 4A.4.1–4A.4.4, June 2014. doi: 10.1109/IRPS.2014.6860642.
- [31] J. F. Bulzacchelli, C. Menolfi, T. J. Beukema, D. W. Storaska, J. Hertle, D. R. Hanson, P. H. Hsieh, S. V. Rylov, D. Furrer, D. Gardellini, A. Prati,

- T. Morf, V. Sharma, R. Kelkar, H. A. Ainspan, W. R. Kelly, L. R. Chieco, G. A. Ritter, J. A. Sorice, J. D. Garlett, R. Callan, M. Brandli, P. Buchmann, M. Kossel, T. Toifl, and D. J. Friedman. A 28-Gb/s 4-Tap FFE/15-Tap DFE Serial Link Transceiver in 32-nm SOI CMOS Technology. *IEEE Journal of Solid-State Circuits*, 47(12):3232–3248, Dec 2012. ISSN 0018-9200. doi: 10.1109/JSSC.2012.2216414.
- [32] R. J. E. Jansen, J. Haanstra, and D. S. Greenpeak. Complementary constant-gm biasing of Nauta-transconductors in low-power gm-C filters to +/-2% accuracy over temperature. In *2012 Proceedings of the ESSCIRC (ESSCIRC)*, pages 466–469, Sept 2012. doi: 10.1109/ESSCIRC.2012.6341356.
- [33] M. Erett, D. Carey, J. Hudner, R. Casey, K. Geary, P. Neto, M. Raj, S. McLeod, H. Zhang, A. Roldan, H. Zhao, P. C. Chiang, H. Zhao, K. Tan, Y. Frans, and K. Chang. A 126mW 56Gb/s NRZ wireline transceiver for synchronous short-reach applications in 16nm FinFET. In *2018 IEEE International Solid - State Circuits Conference - (ISSCC)*, pages 274–276, Feb 2018. doi: 10.1109/ISSCC.2018.8310290.
- [34] T. Toifl, C. Menolfi, M. Ruegg, R. Reutemann, D. Dreps, T. Beukema, A. Prati, D. Gardellini, M. Kossel, P. Buchmann, M. Brandli, P. A. Francese, and T. Morf. A 2.6 mW/Gbps 12.5 Gbps RX With 8-Tap Switched-Capacitor DFE in 32 nm CMOS. *IEEE Journal of Solid-State Circuits*, 47(4):897–910, April 2012. ISSN 0018-9200. doi: 10.1109/JSSC.2012.2185342.
- [35] K. Zheng, Y. Frans, K. Chang, and B. Murmann. A 56 Gb/s 6 mW 300 um² inverter-based CTLE for short-reach PAM2 applications in 16 nm CMOS. In *2018 IEEE Custom Integrated Circuits Conference (CICC)*, pages 1–4, April 2018. doi: 10.1109/CICC.2018.8357076.

- [36] P. Upadhyaya, C. F. Poon, S. W. Lim, J. Cho, A. Roldan, W. Zhang, J. Namkoong, T. Pham, B. Xu, W. Lin, H. Zhang, N. Narang, K. H. Tan, G. Zhang, Y. Frans, and K. Chang. A fully adaptive 19-to-56Gb/s PAM-4 wireline transceiver with a configurable ADC in 16nm FinFET. In *2018 IEEE International Solid - State Circuits Conference - (ISSCC)*, pages 108–110, Feb 2018. doi: 10.1109/ISSCC.2018.8310207.
- [37] K. Mueller and M. Muller. Timing Recovery in Digital Synchronous Data Receivers. *IEEE Transactions on Communications*, 24(5):516–531, May 1976. ISSN 0090-6778. doi: 10.1109/TCOM.1976.1093326.
- [38] B. Widrow and S. Streans. *Least-mean-square adaptive filters*. Prentice-Hall, Inc., 1985.
- [39] V. Stojanovic. *Channel-limited high-speed links: modeling, analysis and design*. PhD thesis, Stanford University, 2005.
- [40] B. K. Casper, M. Haycock, and R. Mooney. An accurate and efficient analysis method for multi-Gb/s chip-to-chip signaling schemes. In *2002 Symposium on VLSI Circuits. Digest of Technical Papers (Cat. No.02CH37302)*, pages 54–57, June 2002. doi: 10.1109/VLSIC.2002.1015043.
- [41] K. Lam, L. Dennison, and W. Dally. Simultaneous bidirectional signaling for IC systems. In *ICCD*, 1990.
- [42] L. Dennison, W. Lee, and W. Dally. High-performance Bidirectional Signaling in VLSI Systems. In *Proceedings of the 1993 Symposium on Research on Integrated Systems*, pages 300–319, Cambridge, MA, USA, 1993. MIT Press. ISBN 0-262-02357-1.

- [43] N. Dolatsha and A. Arbabian. Dielectric waveguide with planar multi-mode excitation for high data-rate chip-to-chip interconnects. In *2013 IEEE International Conference on Ultra-Wideband (ICUWB)*, pages 184–188, Sept 2013. doi: 10.1109/ICUWB.2013.6663845.
- [44] B. Widrow, I. Kollar, and Ming-Chang Liu. Statistical theory of quantization. *IEEE Transactions on Instrumentation and Measurement*, 45(2):353–361, Apr 1996. ISSN 0018-9456. doi: 10.1109/19.492748.
- [45] S. Lipshitz, R. Wannamaker, and J. Vanderkooy. Quantization and Dither: A Theoretical Survey. *J. Audio Eng. Soc*, 40(5):355–375, 1992. URL <http://www.aes.org/e-lib/browse.cfm?elib=7047>.
- [46] H. Viswanathan and R. Zamir. On the whiteness of high-resolution quantization errors. *IEEE Transactions on Information Theory*, 47(5):2029–2038, Jul 2001. ISSN 0018-9448. doi: 10.1109/18.930935.
- [47] A. Dastgheib and B. Murmann. Calculation of total integrated noise in analog circuits. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 55(10):2988–2993, Nov 2008. ISSN 1549-8328. doi: 10.1109/TCSI.2008.923276.
- [48] B. Zhang, A. Nazemi, A. Garg, N. Kocaman, M. R. Ahmadi, M. Khanpour, H. Zhang, J. Cao, and A. Momtaz. A 40 nm CMOS 195 mW/55 mW Dual-Path Receiver AFE for Multi-Standard 8.5 - 11.5 Gb/s Serial Links. *IEEE Journal of Solid-State Circuits*, 50(2):426–439, Feb 2015. ISSN 0018-9200. doi: 10.1109/JSSC.2014.2364032.
- [49] A. Varzaghani, A. Kasapi, D. N. Loizos, S. H. Paik, S. Verma, Sotirios, and S. Sidiropoulos. A 10.3-GS/s, 6-Bit Flash ADC for 10G Ethernet Applications. *IEEE Journal of Solid-State Circuits*, 48(12):3038–3048, Dec 2013. ISSN 0018-9200. doi: 10.1109/JSSC.2013.2279419.

- [50] B. Murmann, K. Zheng, and R. Boesch. Equalization and A/D Conversion for High-Speed Links. In K. Gunnam and M. Vahidfar, editors, *Selected Topics in RF, Analog and Mixed Signal Circuits and Systems*. River Publishers, 2017.
- [51] K. Zheng, Y. Frans, S. Ambatipudi, S. Asuncion, H. Reddy, K. Chang, and B. Murmann. An Inverter-based Analog Front End for a 56 Gb/s PAM4 Wireline Transceiver in 16nm CMOS. In *2018 IEEE Symposium on VLSI Circuits (VLSI-Circuits)*, pages 1–2, June 2018.
- [52] H. Chung and G. Y. Wei. ADC-Based Backplane Receiver Design-Space Exploration. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 22(7):1539–1547, July 2014. ISSN 1063-8210. doi: 10.1109/TVLSI.2013.2275742.
- [53] L. Kull, T. Toifl, M. Schmatz, P. A. Francese, C. Menolfi, M. Braendli, M. Kossel, T. Morf, T. M. Andersen, and Y. Leblebici. A 90GS/s 8b 667mW 64x interleaved SAR ADC in 32nm digital SOI CMOS. In *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, pages 378–379, Feb 2014. doi: 10.1109/ISSCC.2014.6757477.
- [54] L. Kull, D. Luu, C. Menolfi, M. Braendli, P. A. Francese, T. Morf, M. Kossel, A. Cevrero, I. Ozkaya, and T. Toifl. A 24-to-72GS/s 8b time-interleaved SAR ADC with 2.0-to-3.3pJ/conversion and >30dB SNDR at Nyquist in 14nm CMOS FinFET. In *2018 IEEE International Solid - State Circuits Conference - (ISSCC)*, pages 358–360, Feb 2018. doi: 10.1109/ISSCC.2018.8310332.
- [55] S. Rylov, T. Beukema, Z. Toprak-Deniz, T. Toifl, Y. Liu, A. Agrawal, P. Buchmann, A. Rylyakov, M. Beakes, B. Parker, and M. Meghelli. A 25Gb/s ADC-based serial line receiver in 32nm CMOS SOI. In *2016 IEEE International Solid-State Circuits Conference (ISSCC)*, pages 56–57, Jan 2016. doi: 10.1109/ISSCC.2016.7417904.

- [56] Y. Duan and E. Alon. A 6b 46GS/s ADC with >23GHz BW and sparkle-code error correction. In *2015 Symposium on VLSI Circuits (VLSI Circuits)*, pages C162–C163, June 2015. doi: 10.1109/VLSIC.2015.7231250.