

# Energy-Efficient Receiver Design for High-Speed Interconnects

Thesis by  
Kuan-Chang Chen

In Partial Fulfillment of the Requirements for  
the degree of  
Doctor of Philosophy

The Caltech logo, featuring the word "Caltech" in a bold, orange, sans-serif font, centered within a light orange rectangular background.

CALIFORNIA INSTITUTE OF TECHNOLOGY  
Pasadena, California

2022  
Defended July 20, 2021

© 2021

Kuan-Chang Chen  
ORCID: 0000-0003-2968-4656

## ACKNOWLEDGEMENTS

My studies at Caltech have been fueling my aspiration to pursue technological advancement for the greater good. This is made possible by the exemplars of most talented and wonderful researchers and engineers, whom I have been fortunate to work with and learn from. I would like to express my sincere appreciation to them and to others empowering me to tackle the challenges and uncertainties throughout this journey.

First and foremost, my deepest gratitude goes to my advisor, Prof. Azita Emami. It is her superb expertise as well as vision that guides my studies in the fascinating field of high-speed interconnect research. The support from her, sensible in various forms, has been vitally magnificent. It can be her remarkable devotion when I am in need of help. It can be her encouragement to explore innovative and original research ideas, or her words of cheer by telling me “Don’t worry. I’m always behind you.” Over the years, I have come to realize that Azita is one of the few, besides my family, who wholeheartedly wish me every success in all my endeavors. It has therefore been a wonderful privilege for me to learn and to significantly benefit from her advice, inspiring and allowing me to keep aiming at higher goals. For all these, I am immensely grateful to Prof. Azita Emami with highest regards.

Having my upmost respect for David A. Nelson as a circuit architect and designer, I am massively indebted and grateful to David for his guidance and mentorship. His technical feedback has always been right on target, and moreover, it is extraordinarily admirable that his insightful responses to my reports or numerous questions are mostly composed of kind educational purposes to encourage my research efforts. These technical support and enlightenment from him have been pivotal to the growth of my knowledge in chip design and to the success of my chip tape-outs as well, for which I am deeply thankful to David.

It is my great honor to have Prof. Axel Scherer, Prof. Alireza Marandi, and Prof. Ali Hajimiri on my PhD defense/candidacy committee. Not only do I have my tremendous appreciation for their participation in examining my research work, but I also have always looked up to them in light of their dedication and contributions to the future science and engineering.

Looking further back, I am certain that I have to express my enormous gratitude to Prof. Yi-Chang Lu of National Taiwan University (NTU). As my undergraduate advisor, Prof. Lu has constantly amazed me with his brilliance in solving all kinds of problems and his passion for teaching and guiding students. For all these setting the start of my academic pursuit and graduate studies, I am enormously thankful to Prof. Yi-Chang Lu.

In the summer of 2019, I had the splendid pleasure to join Xilinx SerDes Technology Group as an intern. This world-class research and development team granted me highly fruitful and enjoyable experience, all thanks to the talented team members and managers. My profound gratitude goes to Mayank Raj, Yohan Frans, Chuan Xie, Ken Chang, Stanley Chen, Parag Upadhyaya, Jay Im, Didem Turker, and Ping-Chuan Chiang.

I eternally treasure the friendships with the past and current members of the Caltech Mixed-mode Integrated Circuits and Systems (MICS) Lab, Mayank Raj, Manuel Monge, Saman Saeedi, Mahsa Shoaran, Abhinav Agarwal, Arian Hashemi Talkhooncheh (Aryan), Fatemeh Aghlmand (Fatima), Sahil Shah, Benyamin Allahgholizadeh Haghi (Ben), Saransh Sharma, William Wei-Ting Kuo, Minwo Wang, Lin Ma, Shawn Sheng, and Steven Bulfer. I am particularly thankful to Mayank, Manuel, and Abhinav for sharing their valuable experience in chip testing with me. During my second tape-out, I owed a debt of gratitude to Abhinav, Aryan, Fatima, Saransh, William, and Minwo for supporting me until the end.

I hope to extend my great gratitude to Caltech High-Speed/Holistic Integrated Circuits (CHIC) Lab members and alumni for generously sharing testing resources, with huge thanks to the CHIC Lab director, Prof. Ali Hajimiri, and to Aroutin Khachaturian, Reza Fatemi, Matan Gal-Katziri, Amirreza Safaripour, and Behrooz Abiri.

I very much appreciate the support and assistance from Caltech administrative professionals and from the Caltech International Student Programs (ISP) advisors, with special thanks to Michelle Chen, Tanya Owen, Carol Sosnowski, Angie Riley, Kathryn Finigan (Kate), Laura Flower Kim, and Daniel Yoder.



Lastly, my gratitude to my parents, sisters, and aunt can never be sufficient. Their endless and unconditional love is simply the best part of my life and has made the best part of me.

## ABSTRACT

High-speed interconnects are of vital importance to the operation of high-performance computing and communication systems, determining the ultimate bandwidth or data rates at which the information can be exchanged. Optical interconnects and the employment of high-order modulation formats are considered as the solutions to fulfilling the envisioned speed and power efficiency of future interconnects. One common key factor in bringing the success is the availability of energy-efficient receivers with superior sensitivity. To enhance the receiver sensitivity, improvement in the signal-to-noise ratio (SNR) of the front-end circuits, or equalization that mitigates the detrimental inter-symbol interference (ISI) is required. In this dissertation, architectural and circuit-level energy-efficient techniques serving these goals are presented.

First, an avalanche photodetector (APD)-based optical receiver is described, which utilizes non-return-to-zero (NRZ) modulation and is applicable to burst-mode operation. For the purposes of improving the overall optical link energy efficiency as well as the link bandwidth, this optical receiver is designed to achieve high sensitivity and high reconfiguration speed. The high sensitivity is enabled by optimizing the SNR at the front-end through adjusting the APD responsivity via its reverse bias voltage, along with the incorporation of 2-tap feedforward equalization (FFE) and 2-tap decision feedback equalization (DFE) implemented in current-integrating fashion. The high reconfiguration speed is empowered by the proposed integrating dc and amplitude comparators, which eliminate the  $RC$  settling time constraints. The receiver circuits, excluding the APD die, are fabricated in 28-nm CMOS technology. The optical receiver achieves bit-error-rate (BER) better than  $1E-12$  at  $-16$ -dBm optical modulation amplitude (OMA), 2.24-ns reconfiguration time with 5-dB dynamic range, and 1.37-pJ/b energy efficiency at 25 Gb/s.

Second, a 4-level pulse amplitude modulation (PAM4) wireline receiver is described, which incorporates continuous time linear equalizers (CTLEs) and a 2-tap direct DFE dedicated to the compensation for the first and second post-cursor ISI. The direct DFE in a PAM4 receiver (PAM4-DFE) is made possible by the proposed CMOS track-and-regenerate slicer. This

proposed slicer offers rail-to-rail digital feedback signals with significantly improved clock-to-Q delay performance. The reduced slicer delay relaxes the settling time constraint of the summer circuits and allows the stringent DFE timing constraint to be satisfied. With the availability of a direct DFE employing the proposed slicer, inductor-based bandwidth enhancement and loop-unrolling techniques, which can be power/area intensive, are not required. Fabricated in 28-nm CMOS technology, the PAM4 receiver achieves BER better than  $1E-12$  and 1.1-pJ/b energy efficiency at 60 Gb/s, measured over a channel with 8.2-dB loss at Nyquist frequency.

Third, digital neural-network-enhanced FFEs (NN-FFEs) for PAM4 analog-to-digital converter (ADC)-based optical interconnects are described. The proposed NN-FFEs employ a custom learnable piecewise linear (PWL) activation function to tackle the nonlinearities with short memory lengths. In contrast to the conventional Volterra equalizers where multipliers are utilized to generate the nonlinear terms, the proposed NN-FFEs leverage the custom PWL activation function for nonlinear operations and reduce the required number of multipliers, thereby improving the area and power efficiencies. Applications in the optical interconnects based on micro-ring modulators (MRMs) are demonstrated with simulation results of 50-Gb/s and 100-Gb/s links adopting PAM4 signaling. The proposed NN-FFEs and the conventional Volterra equalizers are synthesized with the standard-cell libraries in a commercial 28-nm CMOS technology, and their power consumptions and performance are compared. Better than 37% lower power overhead can be achieved by employing the proposed NN-FFEs, in comparison with the Volterra equalizer that leads to similar improvement in the symbol-error-rate (SER) performance.

## PUBLISHED CONTENT AND CONTRIBUTIONS

K. Chen and A. Emami, "A 25Gb/s APD-based burst-mode optical receiver with 2.24ns reconfiguration time in 28nm CMOS," *2018 IEEE Custom Integrated Circuits Conference (CICC)*, 2018, pp. 1-4, doi: 10.1109/CICC.2018.8357074.

K. -C. C. participated in conceiving the ideas, designed the CMOS chip, performed the experiments, and co-wrote the manuscript.

K. C. Chen and A. Emami, "A 25-Gb/s Avalanche Photodetector-Based Burst-Mode Optical Receiver With 2.24-ns Reconfiguration Time in 28-nm CMOS," in *IEEE Journal of Solid-State Circuits*, vol. 54, no. 6, pp. 1682-1693, June 2019, doi: 10.1109/JSSC.2019.2902471.

K. -C. C. participated in conceiving the ideas, designed the CMOS chip, performed the experiments, and co-wrote the manuscript.

K. Chen, W. W. Kuo and A. Emami, "A 60-Gb/s PAM4 Wireline Receiver with 2-Tap Direct Decision Feedback Equalization Employing Track-and-Regenerate Slicers in 28-nm CMOS," *2020 IEEE Custom Integrated Circuits Conference (CICC)*, 2020, pp. 1-4, doi: 10.1109/CICC48029.2020.9075948.

K. -C. C. participated in conceiving the ideas, designed the CMOS chip, performed the experiments, and co-wrote the manuscript.

K. -C. Chen, W. W. -T. Kuo and A. Emami, "A 60-Gb/s PAM4 Wireline Receiver With 2-Tap Direct Decision Feedback Equalization Employing Track-and-Regenerate Slicers in 28-nm CMOS," in *IEEE Journal of Solid-State Circuits*, vol. 56, no. 3, pp. 750-762, March 2021, doi: 10.1109/JSSC.2020.3025285.

K. -C. C. participated in conceiving the ideas, designed the CMOS chip, performed the experiments, and co-wrote the manuscript.

K. -C. Chen and A. Emami, "Nonlinear Equalization for Optical Interconnects," *accepted to 2021 IEEE Photonics Conference (IPC)*, 2021.

K. -C. C. participated in conceiving the proposed equalizer architecture, performed the simulations, and co-wrote the manuscript.

K. -C. Chen and A. Emami, "Energy-Efficient Neural-Network-Enhanced FFE for PAM4 ADC-Based Optical Interconnects," *to be submitted*.

K. -C. C. participated in conceiving the proposed equalizer architectures, performed the simulations, and co-wrote the manuscript.

## TABLE OF CONTENTS

Acknowledgements .....	iii
Abstract .....	vi
Published Content and Contributions .....	viii
Table of Contents .....	ix
List of Illustrations .....	xi
List of Tables .....	xviii
Chapter I: Introduction .....	1
1.1 Optical Interconnects .....	2
1.2 PAM4 Receivers .....	3
1.3 Organization .....	6
Chapter II: Background .....	10
2.1 Transmitter-Side FFE .....	10
2.2 Transmitter-Side Nonlinear Equalization .....	11
2.3 Receiver-Side CTLE .....	14
2.4 Receiver-Side FFE .....	16
2.5 Receiver-Side DFE .....	19
2.5.1 Direct DFE-FIR .....	19
2.5.2 Loop-Unrolling DFE .....	21
2.5.3 Look-Ahead Multiplexing DFE .....	22
2.5.4 DFE-IIR .....	25
2.6 Receiver-Side Nonlinear Equalization .....	26
Chapter III: APD-Based Burst-Mode Optical Receiver .....	30
3.1 Overview .....	30
3.2 Avalanche Photodetector (APD) .....	32
3.3 APD-Based Optical Receiver Architecture .....	35
3.4 Equalizer Design .....	37
3.5 Burst-Mode Reconfiguration Loops .....	41
3.5.1 Pulse-Triggered State Machine .....	42
3.5.2 Integrating DC Comparator .....	44
3.5.3 Integrating Amplitude Comparator .....	45
3.5.4 Analog Settling Time Reduction .....	50
3.5.5 Simulation Results .....	53
3.6 Experimental Results .....	53
3.7 Summary .....	56
Chapter IV: PAM4 Wireline Receiver with 2-Tap Direct DFE .....	58
4.1 Overview .....	58
4.2 Receiver Architecture .....	62
4.2.1 Overall Architecture .....	62
4.2.2 CTLE .....	62
4.2.3 Summer .....	64

4.2.4 Linearity Characterizations .....	67
4.2.5 CML-to-CMOS Clock Converter .....	67
4.2.6 DCC Circuits .....	68
4.3 Slicer Design .....	71
4.3.1 Slicer Overview .....	71
4.3.2 Prevalent Slicer Topologies.....	72
4.3.3 CMOS Track-and-Regenerate Slicer .....	74
4.3.4 Simulation Results .....	76
4.4 DFE Loops .....	83
4.5 Experimental Results .....	87
4.6 Summary .....	91
Chapter V: Energy-Efficient Neural-Network-Enhanced FFE for PAM4	
ADC-Based Optical Interconnects.....	92
5.1 Overview .....	92
5.2 MRM-Based PAM4 Interconnects .....	94
5.2.1 MRM Nonlinear Distortion and Bandwidth .....	96
5.2.2 Volterra Series Fitting .....	97
5.3 Principle and Noise Analysis.....	99
5.3.1 Overview of Activation Functions .....	99
5.3.2 Custom Learnable PWL Activation Function .....	99
5.3.3 Level-Dependent Noise Analysis.....	101
5.3.4 Numerical Examples and Comparisons .....	104
5.4 Neural-Network-Enhanced FFE.....	107
5.4.1 Architecture .....	107
5.4.2 Extended Noise Analysis Techniques.....	110
5.5 Link Simulations and SER Performance.....	111
5.6 Design Framework Summary and Hardware Synthesis .....	114
5.7 Summary .....	115
Chapter VI: Conclusion .....	117
Bibliography .....	121

## LIST OF ILLUSTRATIONS

<i>Number</i>	<i>Page</i>
1.1	3
1.2	5
1.3	6
2.1	11
2.2	12
2.3	13
2.4	13
2.5	15
2.6	20
2.7	22
2.8	23
2.9	24
2.10	24
2.11	26
2.12	27
2.13	28

3.1	(a) Transmission of distinct data bursts originated from different transmitters to a single optical line terminal (OLT), including a burst-mode optical receiver (BMRX). (b) Simplified timing diagram of burst-mode reconfiguration scheme, in which the dc offset cancellation and amplitude control are the focus of this paper. ....	32
3.2	SNR improvements (in decibels) versus $M$ with a given level of optical input power ( $-16$ dBm). $k = 0.2$ , $R = 0.7$ , and different amounts of input-referred thermal noise are used in the computations. ....	34
3.3	(a) Architecture of the BMRX. (b) Circuit schematic of the VCS. (c) Circuit schematic of the three-stage inverter-based TIA. $R_{F1} = 1.2$ k $\Omega$ and $R_{F2} = 275$ $\Omega$ nominally in this design. (d) Circuit schematic of the single-ended-to-differential amplifier (S2D), with load resistors set to $172$ $\Omega$ in this design. (e) Circuit schematic of the current-steering VGA, with load resistors set to $172$ $\Omega$ in this design. (f) Circuit schematic of the enable/disable control scheme for the LPF loop. ....	36
3.4	(a) Pulse responses at the AFE outputs before applying equalization. (b) Pulse responses at the AFE outputs after applying ideal two-tap FFE. ....	39
3.5	Schematic of the EQ performing double-sampling and two-tap DFE. ....	40
3.6	Block diagram of the burst-mode reconfiguration loops.....	42
3.7	Block diagram of the pulse-triggered state machine. ....	43
3.8	(a) Conventional $RC$ LPF-based dc comparator. (b) Simulation results showing the tradeoff between tracking time and settling behavior. ....	45
3.9	(a) Circuit schematic of the proposed integrating dc comparator. (b) Simulation results showing the operation of the proposed integrating dc comparator, where the dc level of $V_{IN}$ is lower than that of $V_{IP}$ by	



20 mV. (c) Integrating dc comparator differential output voltage versus different clock duty cycles with four distinct dc-level differences. ....	46
3.10 (a) Building block of the proposed integrating amplitude comparator. (b) Simulation results showing the operation of the building block in the proposed integrating amplitude comparator. ....	48
3.11 (a) Circuit schematics of the proposed integrating amplitude comparator. (b) Integrating amplitude comparator differential output voltage versus different clock duty cycles with four distinct amplitude differences. ....	50
3.12 (a) Circuit schematic of the conventional DAC loaded with diode-connected transistors for current mirroring. (b) Circuit schematic of the DAC loaded with low-resistance resistors. The differential output voltage is taken as the differential input of a voltage-to-current (V2I) converter. ....	52
3.13 Simulated AFE outputs in burst-mode reconfiguration.....	54
3.14 Block diagram of the experiment setup. ....	54
3.15 Bathtub curve measured with $-16$ -dBm OMA at 25 Gb/s.....	55
3.16 Waterfall plot with fixed EQ setting at 25 Gb/s, with PRBS-7.....	56
3.17 Waterfall plot with fixed EQ setting at 25 Gb/s, with PRBS-31.....	56
3.18 (a) Micrograph of the core circuitry, including the pad wire-bonded to the APD (APDIN), AFE, quarter-rate EQ (EQ), integrating dc comparator (Int. dc Comp.), integrating amplitude comparator (Int. Amp. Comp.). (b) Power consumption breakdown of the receiver data path at 25 Gb/s.....	57
4.1 Hardware implementation of PAM4-DFE in half-rate designs. Only the even data-path is shown for clarity, where $TH_H$ , $TH_0$ , and $TH_L$ are the three distinct threshold levels, and $h_1$ corresponds to the first post-cursor ISI. (a) Direct 1-tap PAM4-DFE. (b) Loop-unrolling 1-tap PAM4-DFE. ....	61

4.2	Timing constraint for a direct DFE design for $N$ -th post-cursor ISI compensation.....	61
4.3	Overall architecture of the PAM4 receiver.....	63
4.4	(a) Schematic of the source-degenerated CTLE. (b) Simulated frequency response of the CTLE (single stage) with different settings of $V_{CAP}$ .....	63
4.5	Architecture and performance of the summer for 2-tap DFE. ....	65
4.6	(a) Schematic of the common-mode restoration circuits. (b) Simulated performance of the common-mode restoration circuits, showing the deviation from the target common mode with and without the common-mode restoration circuits. ....	66
4.7	(a) Nomenclature for PAM4 eye diagrams and the definition for PAM4 EL. (b) Simulated linearity performance of the summer. (c) Simulated linearity performance of the CTLE. ....	68
4.8	(a) Schematic of the CML-to-CMOS clock converter. (b) Simulated minimum required input peak-to-peak amplitude with different input clock frequencies for the CML-to-CMOS clock converter.....	69
4.9	(a) Schematic of the DCC circuits. (b) Simulated performance of DCC with 15-GHz clock signals.....	70
4.10	Prevalent slicer topologies. (a) StrongArm slicer. (b) CML slicer. ...	73
4.11	Simulated waveforms showing the typical operations including the reset, sample, and regenerate phases of the StrongArm slicer. ....	73
4.12	Proposed CMOS track-and-regenerate slicer. (a) Overall circuit schematic. (b) Proposed slicer in track mode. (c) Propose slicer in regenerate mode. ....	75
4.13	Simulations and comparisons of the large-signal performance between the reset-and-regenerate StrongArm and the proposed CMOS track-and-regenerate slicer. (a) Input signals to the slicers. (b) Optimal clock signals and the resulting output waveforms of the slicers with 900-mV supply. (c) Optimal clock signals and the	

	resulting output waveforms of the slicers with 850-mV supply. (d) Faster reaction to strong symbols with the proposed slicer.....	77
4.14	(a) Slicer input signals, and the simulated slicer output swings with distinct input swings at 30 GBaud/s. (b) Slicer input signals, and the simulated slicer input sensitivity at different baud rates. ....	79
4.15	Simulated output swing and clock-to-Q delay performance of the StrongArm (SA) and the proposed track-and-regenerate slicer (T/R) for typical strong symbols and weak symbols. (a) Output swing versus input common-mode level, $V_{DD} = 0.9$ V. (b) Output swing versus supply voltage, $V_{CM} = 0.75$ V. (c) Clock-to-Q delay versus input common-mode level, $V_{DD} = 0.9$ V. (d) Clock-to-Q delay versus supply voltage, $V_{CM} = 0.75$ V. ....	81
4.16	Simulated results from the PSS and PNOISE simulations at 30-GBaud/s operation; the clock frequency is 15 GHz. (a) Clock signal and the proposed slicer's differential output signal from PSS simulation. (b) Simulated differential output integrated noise of the proposed slicer from PNOISE simulations in time domain. (c) Resultant differential output SNR of the proposed slicer.....	82
4.17	(a) Simulated SSF of the proposed track-and-regenerate slicer at 30 GBaud/s. (b) Simulated ISF of the proposed track-and-regenerate slicer at 30 GBaud/s. ....	85
4.18	Timing diagram for the first tap DFE in a half-rate design, and the simulated numbers for the timing constraints at 30 GBaud/s. ....	86
4.19	Pulse responses with considerable post-cursor ISI. (a) Simulated normalized pulse response at the input of the summer. (b) Simulated normalized pulse response at the output of the summer.....	87
4.20	Simulated differential output of the summer with distinct DFE settings. (a) First-tap DFE and second-tap DFE are both disabled. (b) First-tap DFE is disabled, while the second-tap DFE is enabled. (c)	

First-tap DFE is enabled, while the second-tap DFE is disabled. (d)	
First-tap DFE and the second-tap DFE are both enabled.....	88
4.21 Block diagram of the experiment setup.....	89
4.22 (a) Measured 30-GBaud/s pulse response at the input of the receiver chip. (b) Measured (single-ended) 60-Gb/s PAM4 data eyes at the input of the receiver chip.....	89
4.23 (a) Measured bathtub curves at 60-Gb/s PAM4, with DFE loops disabled/enabled. (b) Measured eye contour color map at 60-Gb/s PAM4 after equalization.....	90
4.24 (a) Chip micrograph with key building blocks highlighted. (b) Measured receiver data-path power consumption at 60-Gb/s PAM4. .....	90
5.1 Description of an output sample ( $y[n]$ ) with functions of input ( $x$ ) samples lying within certain time-spans.....	93
5.2 System overview of MRM-based PAM4 optical interconnects.....	96
5.3 (a) PAM4 eye-diagram and signal constellations simulated with MRM of $Q = 7820$ at 50 Gb/s. (b) PAM4 eye-diagram and signal constellations simulated with MRM of $Q = 15640$ at 50 Gb/s.....	97
5.4 (a) Volterra series fitting example at 50 Gb/s. (b) Volterra series fitting example at 100 Gb/s.....	99
5.5 (a) Conventional linear FFE. (b) Feedforward neuron (one-layer neural network).....	100
5.6 Approximations of nonlinear functions using superpositions of PWL functions.....	100
5.7 Level-dependent noise CDFs associated with nonlinear equalization. (a) 2nd-order Volterra. (b) FReLU.....	106
5.8 Discrepancy between transient-simulation-based CDF and statistical CDF.....	108
5.9 Discrepancy between analytical CDF and numerical-integral CDF.....	108

5.10 (a) Volterra equalizer of memory length 2 with 5-tap FFE. (b) Custom neural-network-enhanced 5-tap FFE.....	109
5.11 SER simulations with distinct MRM and equalizer designs. (a) At 50-Gb/s PAM4. (b) At 100-Gb/s PAM4. ....	113
5.12 Design framework summary.....	114

## LIST OF TABLES

<i>Number</i>	<i>Page</i>
1.1 Link power budget for the laser diode power consumption for given receiver sensitivity and link loss, using a commercial laser diode.....	3
3.1 Performance summary and comparisons of optical receivers.....	57
4.1 Slicer comparisons (the proposed CMOS track-and-regenerate slicer vs. the conventional StrongArm slicer). .....	82
4.2 Performance summary and comparison of wireline receivers. ....	91
5.1 Results of nonlinear equalization examples. ....	106
5.2 Hardware and power overhead comparisons (the proposed NN-FFE vs. the conventional VT-FFE). .....	114

*Chapter 1*

## INTRODUCTION

The notion of *high-speed* evolves with time, reflecting the ever-growing data traffic that connects and benefits our daily lives. With the continually emerging internet applications which unceasingly incite the growth of the numbers of users and connected devices, it is observed that the volume of the data traffic has been increasing exponentially. As the momentum for fast-growing internet connections continues to thrive, it is forecasted the speed performance of various networks will advance more than two-fold from 2018 to 2023 [1]. In addition, the advent and progressive developments of both artificial intelligence (AI) and the fifth generation (5G) communication technologies also necessitate high-speed interconnects serving as the backbone to support fast data communication within the computers and infrastructures. In light of all these technological pursuits pointing to an era of big data, the evolution of high-speed interconnects allows exchanging data with higher speed and lower power, thereby shaping the future of high-performance computing and communication systems.

In response to the demand for interconnects of higher speed, efforts have been made to innovate the per-pin data rate of the interconnects. In every three to four years, the speed has approximately doubled for almost all I/O standards [2]. However, on the way towards higher data rates, electrical interconnects suffer from high channel losses that increase with the modulation frequency and/or transmission distance. In consequence, improvement or even preservation of the energy efficiency with electrical interconnects becomes prohibitively difficult to achieve at high data rates. More specifically, a channel with 30-dB more loss corresponds to about 10 times more power consumption per bit [3]. By contrast, optical interconnects have shown the favorable superiority in that little modulation-frequency-dependent loss is introduced by the fibers. Accordingly, optical interconnects possess promising potentials to fulfill the envisioned power, data rate, and reach requirements [4]. Meanwhile, for a given bandwidth limitation, it is feasible to increase the data rate with the

utilization of high-order modulation formats thanks to the augmented spectral efficiency. In particular, 4-level pulse amplitude modulation (PAM4) is an appealing option, since its Nyquist frequency is only half of that of non-return-to-zero (NRZ) modulation, at the expense of moderately reduced signal swings relative to other higher-order modulation formats with even more signal levels.

Optical interconnects and PAM4 format, which promise lower channel loss and higher spectral efficiency, have fueled the evolution of high-speed interconnects. The pivotal design considerations enabling energy-efficient high-speed interconnects leveraging optics and/or PAM4 are presented in the following.

### **1.1 Optical Interconnects**

A high-level illustration of an optical interconnect is shown in Fig. 1.1. This optical interconnect consists of a continuous-wave laser source, an optical modulator driven by an electrical driver, an optical channel, and an optical receiver formed by the combination of a photodetector (PD) and electrical receiver circuits. Although it is known that optical interconnects suffer much less channel loss compared to the copper wires at high modulation frequencies, the power loss in the optical link itself can be highly considerable, especially in the cases where long fibers and a large number of connectors, couplers, or splitters are included. To investigate the deciding factors in the overall link power consumption, Table 1.1 is presented as a representative example. In Table 1.1, the first and second columns respectively assume realistic receiver sensitivity and link loss numbers, and the resultant required laser output power specifications are shown in the third column. The corresponding laser diode power consumptions, based on the characteristics of a commercialized laser diode, is given in the rightmost column of Table 1.1.

Since the laser source such as the aforementioned laser diode generally consumes a major portion of the total link power, as can be observed in Table 1.1, the receiver sensitivity plays a critical role in affecting the overall power consumption, especially when high power loss is inevitable in the link. In other words, a high-sensitivity optical receiver significantly



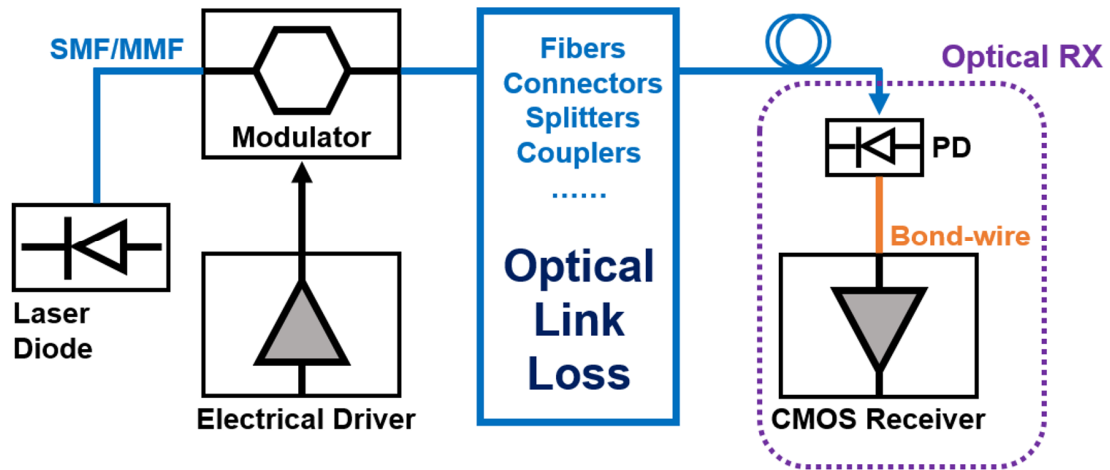


Fig. 1.1. Illustration of an optical link/interconnect.

Receiver Sensitivity	Link Loss	Laser Output Power	Laser Diode Power Consumption
-5 dBm	15 dB	$\geq 10$ mW	$\geq 66$ mW
-10 dBm	15 dB	$\geq 3.16$ mW	$\geq 45$ mW
Receiver Sensitivity	Link Loss	Laser Output Power	Laser Diode Power Consumption
-5 dBm	25 dB	$\geq 100$ mW	$\geq 480$ mW
-10 dBm	25 dB	$\geq 31.6$ mW	$\geq 150$ mW

Table 1.1. Link power budget for the laser diode power consumption for given receiver sensitivity and link loss, using a commercial laser diode.

benefits the overall link energy efficiency, and therefore sensitivity optimization should be one focus of an optical receiver design.

## 1.2 PAM4 Receivers

In contrast to the NRZ systems where each symbol contains only one bit of information (i.e., bit 0 or bit 1), PAM4 signaling utilizes four distinct levels with each corresponding to two

bits of information (i.e., 00, 01, 10, and 11 symbols). Consequently, in comparison with NRZ modulation, PAM4 signaling allows twice more data bits to be transmitted/received for a given symbol rate, or baud rate, thereby doubling the data rate. This attractive benefit from PAM4 signaling can also be understood with the power spectral density (PSD) plots in the frequency domain. As displayed in Fig. 1.2(a), compared to NRZ modulation, PAM4 signaling improves the spectral efficiency and halves the Nyquist frequency. Therefore, when PAM4 signaling is adopted to replace the more conventional NRZ modulation, the bandwidth requirements are relaxed for the channel and the transceiver circuits as well.

However, it can be challenging to design the PAM4 transceivers by virtue of the multi-level signaling. In particular, the reduced eye-height at the receiver side urges the decision circuits to be designed with improved sensitivity. As illustrated in Fig. 1.2(b), with a fixed signal swing  $V_{sw}$ , the eye-height in PAM4 systems is nominally only one third of that in NRZ systems, in the absence of nonlinearity. Moreover, this smaller eye-height implies the necessity of effective ISI suppression in that any residual ISI can further compromise the eye-opening and hence deteriorate the bit-error-rate (BER) performance. It has been a popular option to include a decision feedback equalizer (DFE) in a PAM4 receiver, because a DFE is capable of compensating the post-cursor ISI without the undesirable noise or crosstalk enhancement. Nonetheless, the inclusion of a PAM4-DFE may lead to more demanding specifications for the decision circuits (i.e., slicers), in view of the timing constraints associated with the DFE. On top of that, in designing the slicers to achieve the speed and sensitivity requirements, it is important to note that the power and area consumptions of each slicer are of great concern, since at least three slicers are required to accommodate the three distinct thresholds.

The foregoing suggests two major design considerations. For one thing, with the availability of high-sensitivity slicers, it becomes possible to correspondingly relax the specification of the PAM4 transmitter swing. For the other thing, if the speed performance of the slicers permits, the advantages of incorporating a DFE in a PAM4 receiver can be realized. In other

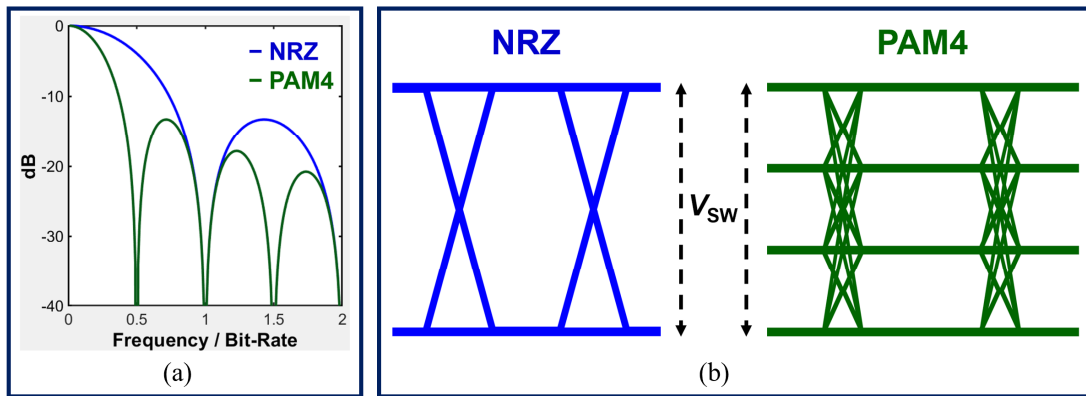


Fig. 1.2. (a) Power spectral density (PSD) plots of NRZ and PAM4. (b) Illustration of the eye diagrams of NRZ and PAM4 for a fixed signal swing  $V_{sw}$ .

words, the slicer design is crucial to the adoption of PAM4 signaling, especially for high data-rate operations where the post-cursor ISI tends to be more salient.

The employment of an optical interconnect adopting PAM4 format simultaneously benefits from the improved spectral efficiency as well as the little modulation-frequency-dependent loss from the optical fibers, thereby considered as a charming candidate to support high-speed and long-distance data communication. However, in optical interconnects, the nonlinearities caused by the generally nonlinear response of optical modulators would detrimentally shrink the eye-openings. Since the peak modulation amplitude is divided into multiple levels in a high-order-modulated system such as PAM4 transceivers, these signal impairments due to the nonlinearities can be very severe, in view of the relatively stringent signal-to-noise ratio (SNR). As depicted in Fig. 1.3, when an optical modulator, micro-ring modulator (MRM) for example, is driven by a linear PAM4 driver, the output levels are expected to be unequally spaced. The presence of unequal eye-openings implies that the overall symbol-error-rate (SER) performance can be severely degraded, attributed to the insufficient SNR of the smallest eye. Accordingly, nonlinearity compensation, or nonlinear equalization, holds the key to accomplishing the higher data rates empowered by the optical interconnects using high-order modulation formats.

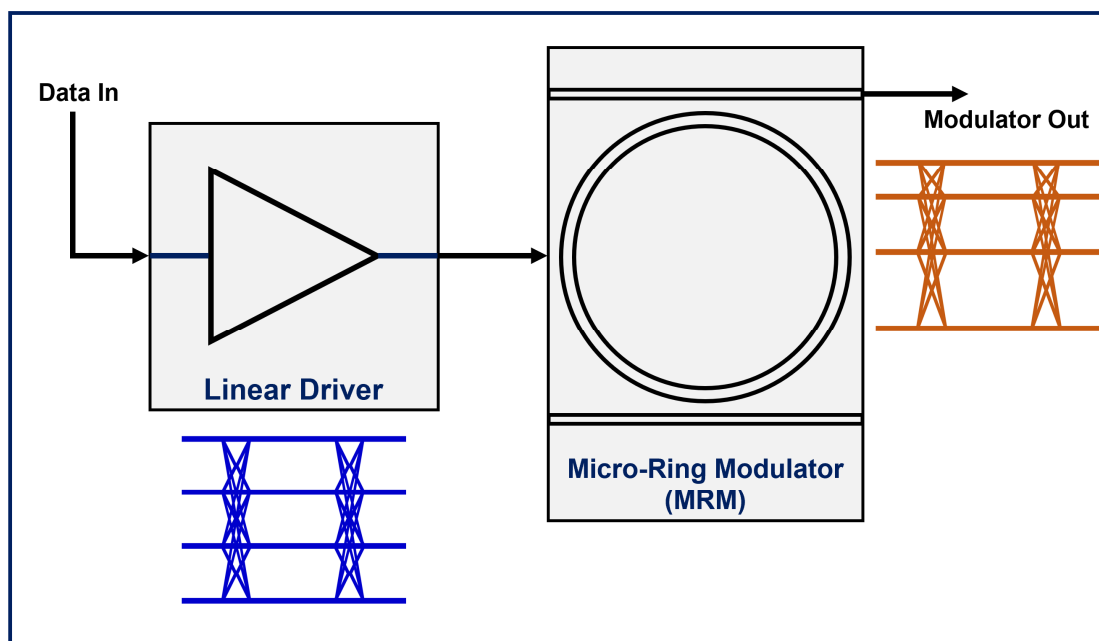


Fig. 1.3. Illustration of the nonlinear response of an optical MRM driven by a linear electrical PAM4 driver, resulting in unequal eye-openings.

### 1.3 Organization

This dissertation presents architectural as well as circuit-level designs and techniques enabling energy-efficient high-speed interconnects. The rest of this dissertation is organized as follows.

In Chapter 2, the basics, features, and implementations of various transmitter-side equalization and receiver-side equalization are summarized. As the receiver-side equalization can significantly improve the receiver sensitivity by effectively mitigating the inter-symbol interference (ISI), the materials presented in this chapter serve as the background and the fundamentals for the subsequent chapters in which different types of equalization techniques are employed to improve the overall performance.

In Chapter 3, a 25-Gb/s avalanche photodetector (APD)-based burst-mode optical receiver is presented. This chapter demonstrates the improvement in receiver sensitivity can be achieved, leveraging a high-gain optical front-end together with the equalization techniques

embedded in the electronic receiver circuits. APD as the very first stage of the optical receiver offers higher gain, compared to a conventional p-i-n diode, and therefore it is possible to improve the SNR at the receiver front-end provided that the thermal noise independent of the APD gain is dominant. The noise analysis of an APD-based optical receiver front-end is available in Chapter 3, where the model of the gain-dependent shot noise of an APD is described. In addition to the employment of an APD, the inclusion of electronic equalizer circuits in the receiver can give rise to further sensitivity improvement, by compensating for the ISI introduced by any slow dynamics in the front-end signal path. In Chapter 3, the proposed equalizer is presented. This equalizer is designed in current-integrating fashion to perform 2-tap FFE and 2-tap DFE, serving as an energy-efficient solution to allowing a relatively low-bandwidth front-end such as the transimpedance amplifier (TIA) to be adopted. With the designed equalizer circuits, it is accomplishable to co-optimize the APD gain along with the gain and bandwidth of the front-end circuits for improving the receiver sensitivity.

Meanwhile, it becomes necessary to make an optical receiver reconfigurable in the scenario where this optical receiver has to respond to multiple transmitters having distinct characteristics. That is, the output data bursts from different transmitters are expected to present distinct dc components, signal swings, and phases to the receiver. Hence, a so-called burst-mode receiver that is reconfigurable to cancel the dc component, control the signal amplitude, and recover the sampling clock phases is required. In Chapter 3, integrating dc comparator and integrating amplitude comparator are proposed to replace the conventional  $RC$  low-pass filter (LPF)-based designs for the purpose of reducing the reconfiguration time. With the help of the proposed integrating dc comparator and the integrating amplitude comparator, the overhead time spent for reconfiguration is significantly reduced, which thus improves the optical link bandwidth as well as latency, especially for a network with frequent switching events.

In Chapter 4, a 60-Gb/s wireline PAM4 receiver with direct 2-tap DFE is presented. Designing a direct DFE to cancel the first few taps of post-cursor ISI at high data rates can

be challenging in view of the tight timing constraints. However, as shown in Chapter 4, for PAM4 signaling, it costs excessive hardware to implement other techniques such as loop-rolling which has been applied to relax the DFE timing constraint. A CMOS track-and-regenerate slicer is proposed as one solution to implementing energy-efficient direct PAM4-DFE at high data rates. In Chapter 4, extensive simulation results of the proposed CMOS track-and-regenerate slicer are presented, in order to study its speed/delay performance along with the DFE timing constraints, and to compare its features with the prevalent StrongArm slicer and current-mode-logic (CML) slicer. The proposed CMOS track-and-regenerate slicer offers three key advantages. First, the clock-to-Q delay is reduced with the employment of the proposed slicer, allowing the DFE timing constraint to be met more easily. Second, the proposed slicer offers rail-to-rail digital-level output swings. Third, the speed performance of the proposed slicer benefits from the ongoing CMOS technology scaling. Other critical circuit blocks in this 60-Gb/s PAM4 receiver are also presented in Chapter 4, including the clock amplifier, continuous time linear equalizer (CTLE), CML summer, duty-cycle correction (DCC) circuits, and common-mode restoration circuits. Despite that the application of the proposed CMOS track-and-regenerate slicer is demonstrated with an electrical wireline PAM4 receiver described in Chapter 4, the advantages of the proposed slicer, which lead to energy-efficient high-speed DFE designs, can be generally leveraged in electrical or optical interconnects to incorporate DFE at high data rates.

In Chapter 5, a series of digital neural-network-enhanced FFEs (NN-FFEs) applicable to PAM4 analog-to-digital converter (ADC)-based optical interconnects is proposed and presented. Optical micro-ring modulators (MRMs) promise improvements in both the link power efficiency as well as the link aggregate bandwidth, by virtue of their relatively compact device sizes, high modulation efficiencies, and potential to support dense wavelength-division multiplexing (WDM) systems. However, when PAM4 signaling is adopted, the generally nonlinear electro-optic modulation of an MRM leads to unequal eye-openings, which necessitates nonlinear equalization for ameliorating the SER performance. In Chapter 5, the nonlinearities of MRMs are first quantitatively characterized in order to investigate the design target of the nonlinear equalization. While the conventional Volterra

equalizers prove to be effective in compensating for the nonlinearities, the required extra multipliers can cost considerable power overhead. Serving as energy-efficient alternatives, the proposed NN-FFE's can achieve similar SER performance with significantly reduced power overheads. The proposed NN-FFE's, which employ custom piecewise linear (PWL) functions, are created learnable with the assistance of open-source machine learning libraries. Details of the noise analysis, power consumptions, and the design framework regarding the nonlinear equalizers are elaborated in Chapter 5. In spite of the fact that the case studies presented in Chapter 5 focus on MRM-based PAM4 optical interconnects, the techniques and methods described in Chapter 5 well extend to the nonlinear equalizer design for different types of optical modulators, or in a broader sense, for other nonlinear channels.

Finally, in Chapter 6, the design considerations and highlights of the receiver circuits presented in this dissertation are summarized, and conclusions are drawn.

## *Chapter 2*

### BACKGROUND

In this chapter, the basics and design considerations of equalization are presented. Transmitter-side linear feedforward equalizer (FFE) is reviewed, and the techniques to further incorporate nonlinear equalization are also elaborated. Afterward, receiver-side linear and nonlinear equalization schemes are described, serving as the foundations for the subsequent chapters.

#### 2.1 Transmitter-Side Feedforward Equalizer (TX-FFE)

Fig. 2.1 shows a general architecture of a linear  $n$ -tap FFE. This linear operation is performed by weighted-summing the data symbols spaced apart in time, which is mathematically described as:

$$Y[k] = \sum_{j=a}^{a+n-1} (W_j X[k+j]) \quad (2.1)$$

where  $j$  is the index,  $n$  is the number of taps,  $a$  and  $k$  are integers,  $X[k]$  and  $Y[k]$  denote the  $k$ -th input data symbol value and the  $k$ -th FFE output, respectively, and  $W_j$  is the tap weight associated with  $X[k+j]$ .

This FFE can be viewed as a pulse-shaping function that allows the signal impairments induced by the channel (e.g., a low-bandwidth channel) to be compensated. More specifically, by making use of the previous as well as the succeeding data symbols, both pre-cursor and post-cursor ISI can be mitigated with the TX-FFE. In the TX-FFE, the delay elements required to implement a multi-tap FFE can be realized with relatively simple digital gates/circuits, whereas the tap weights can be set with digital-to-analog converters (DACs).

While the TX-FFE can be effective in improving the overall eye-openings, two major drawbacks need to be taken into consideration during the design phase. For one thing, as a consequence of allocating a portion of the maximum transmitter swing to the pulse-shaping,



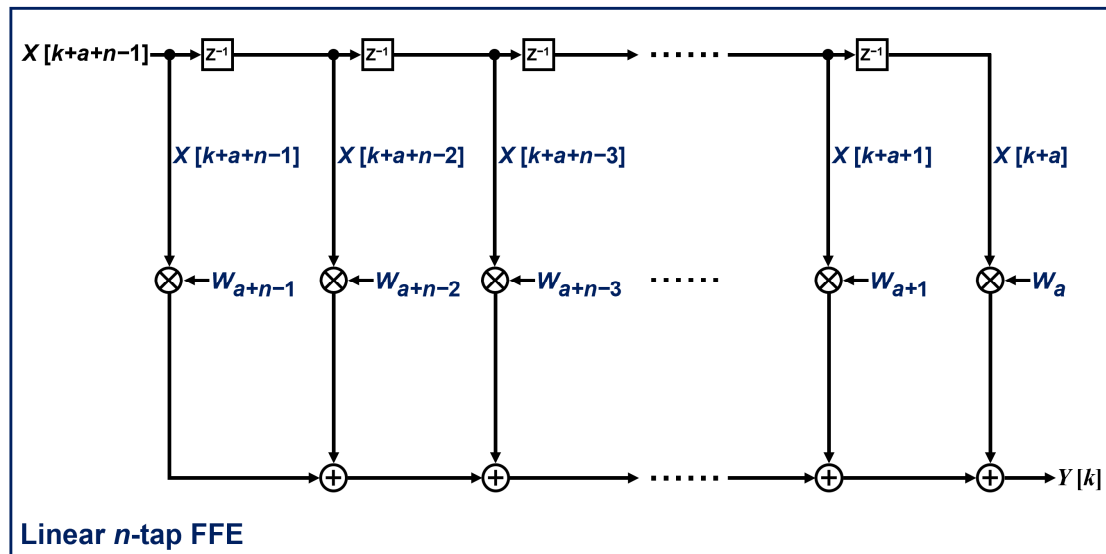


Fig. 2.1. Architecture of a linear  $n$ -tap FFE.

the resultant signal swing is reduced at low frequencies. For the other thing, the overall signal characteristics including the impairments caused by the channel and/or receiver front-end circuits are observable at the receiver side, but not at the transmitter side. Therefore, a back-channel is required in order to adaptively adjust the settings of the TX-FFE.

## 2.2 Transmitter-Side Nonlinear Equalization

The transmitter-side nonlinear equalization aims to enlarge and equalize the eye-openings for achieving better link symbol-error-rate (SER) performance. The pivotal transmitter-side equalization techniques are described in this section with implementation examples. In [5], a vertical-cavity surface-emitting laser (VCSEL)-based NRZ transmitter is modeled, where input data bit 1 and data bit 0 give rise to asymmetrical output pulse responses, attributed to the uneven reactions to the rising and falling edges of modulation. The solution proposed in [5] is shown in Fig. 2.2, which detects the rising and falling edges and correspondingly applies different amounts of pre-emphasis to mitigate the asymmetry. This technique improves the eye-opening and meanwhile ameliorates the transmitter energy efficiency by allowing the VCSEL to be driven at lower bias currents [5]. The foregoing technique suggests that with the inclusion of two auxiliary paths for asymmetrical pre-emphasis, the

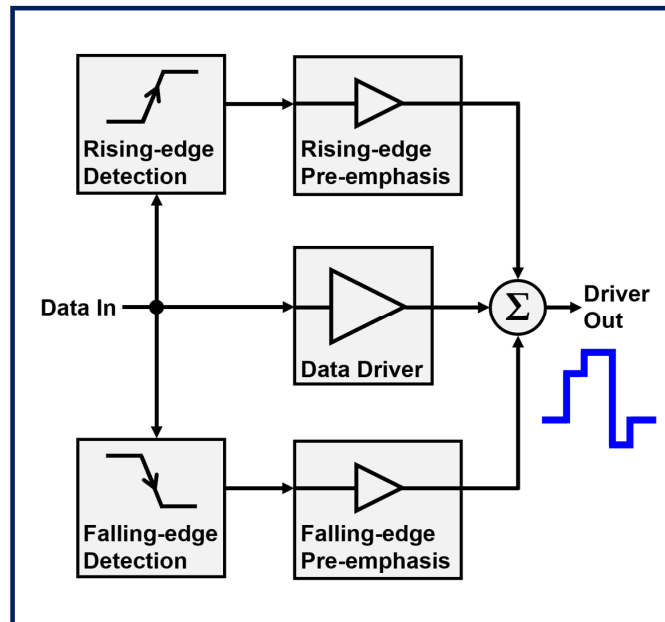


Fig. 2.2. Asymmetric pre-emphasis technique for nonlinear equalization.

nonlinear modulation dynamics in NRZ systems can be addressed. In order to further tackle the unequal eye-openings in high-order-modulated systems, a higher number of signal paths for more accurate compensation is needed. For instance, an optical micro-ring modulator (MRM)-based PAM4 transmitter is reported in [6], where a highly parallelized driver architecture is designed. As depicted in Fig. 2.3, this driver consists of parallel driver slices digitally controlled by a reconfigurable array of lookup tables (LUTs). The LUTs are set to have the analog driver output counteracts the data-dependent MRM nonlinearities. That is, the parallel driver slices act as a segmented electrical digital-to-analog converter (DAC). With a sufficient number of slices/segments, the DAC can incorporate pre-emphasis and pre-distortion for overcoming the nonlinear modulation dynamics and intensity, respectively. The idea of segmentation can be alternatively carried out in the optical domain, thereby simplifying the electrical driver design. Shown in Fig. 2.4 is a two-segment MRM example [7]. By selectively driving the segment(s), the individual contribution to the change of the carrier density within the entire ring and thus the overall modulation of optical intensity at the output can be activated or deactivated. Along with encoders (e.g., LUTs) responsible for digitally setting the selection states, the segmented modulator functions as an optical DAC.

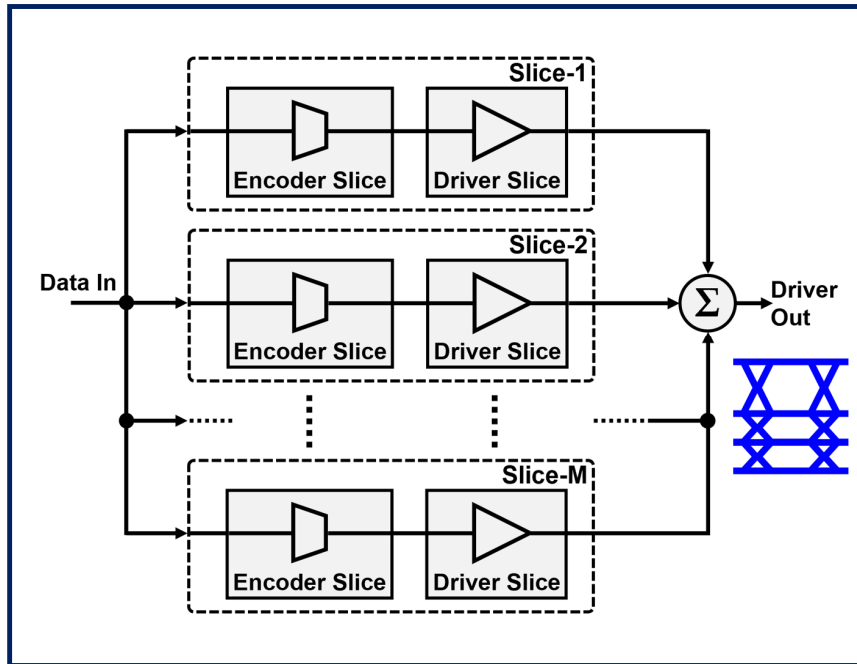


Fig. 2.3. Electrical DAC employing segmented electrical driver slices for pre-distortion/pre-emphasis.

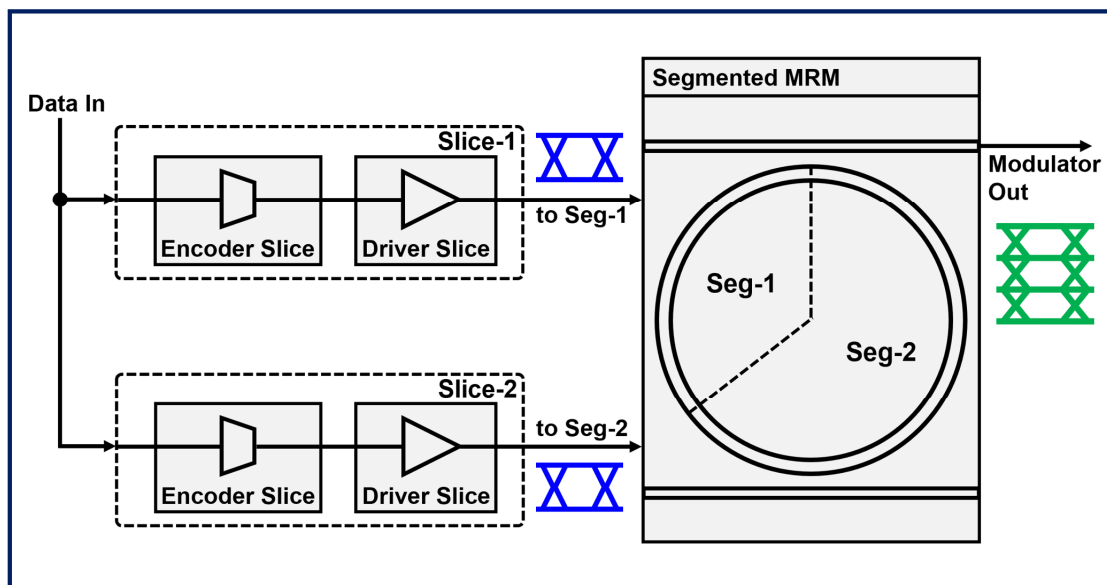


Fig. 2.4. Optical DAC employing a segmented optical modulator along with its driver circuits. A two-segment MRM is shown as an example.

The benefits of segmentation can also be accomplished by using multiple optical modulators; for example, two parallel electro-absorption modulators with uneven lengths are employed in [8]. To improve the DAC resolution for augmented adjustability of compensation, a higher level of segmentation is expected for both electrical DAC and optical DAC. In that case, the former would face challenges in the signal-path bandwidths due to the increased number of electrical connections, whereas the latter would result in overheads in the pin counts and area consumptions [6]. Besides, the optical-domain segmentation mostly fulfills the compensation for the nonlinear modulation intensity, but not for the nonlinear dynamics. At the expense of more driver slices, an electrical DAC can succeed in equalizing both types of nonlinearities by manipulating its driving force. In light of these trade-offs, co-design and co-optimization of electrical drivers along with optical modulators are crucial for effective and energy-efficient transmitter-side equalization.

Similar to the case of transmitter-side linear FFE, the transmitter-side nonlinear equalizer would require a back-channel to capture the accumulated signal impairments including those arising from the channel and/or the receiver front-end circuits. On the contrary, receiver-side equalizers take advantage of the receiving signal paths, allowing the associated adaptations to be fulfilled at the receiver side, while the overall signal characteristics are taken into accounts. The next sections are dedicated to the receiver-side linear and nonlinear equalization schemes along with the design considerations.

### **2.3 Receiver-Side Continuous Time Linear Equalizer (CTLE)**

A continuous time linear equalizer (CTLE) has been widely employed in high-speed serial links. The concept of a CTLE can be understood as a filter that offers gain-boost within a high-frequency band. With this high-frequency boost, or high-frequency peaking in some contexts, the in-band channel loss can be compensated. In consequence, the bandwidth of the overall channel response can be improved. Depending on how the CTLE is implemented, it can fall into the category of either passive or active CTLE. A passive CTLE, as its name suggests, employs passive elements such as resistors, capacitors, and/or inductors in order that the desirable frequency boost is incorporated in the composite frequency response. On

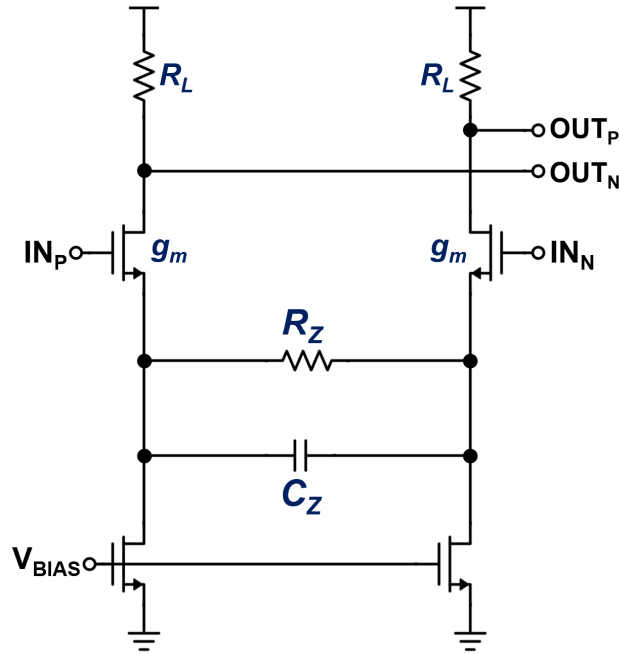


Fig. 2.5. Circuit schematic of a conventional  $RC$  source-degenerated CTLE.

the other hand, active devices (e.g., n-type and/or p-type transistors) are utilized to form an active CTLE which behaves like an amplifier with its gain peaked at the target frequency. A prevalent active CTLE implementation is shown in Fig. 2.5, in which the high-frequency peaking relies on the  $RC$  source-degeneration that introduces a zero in the overall transfer function. The location of the zero in frequency domain, denoted by  $f_z$  is expressed as:

$$f_z = 1 / (2\pi R_Z C_Z) \quad (2.2)$$

where  $R_Z$  and  $C_Z$  are the source-degeneration resistance and capacitance, respectively. As can be seen in (2.2), the location of the zero (i.e.,  $f_z$ ) and thus the peaking frequency is determined by the values of  $R_Z$  and  $C_Z$ . In addition, the low-frequency (small-signal) gain, denoted by  $G_{LF}$ , is derived to be:

$$G_{LF} = (g_m R_L) / (1 + g_m R_Z / 2) \quad (2.3)$$

where  $g_m$  is the transconductance of the transistors, and  $R_L$  is the resistance of the load resistor. Accordingly, the peaking frequency can be adjusted through changing  $R_Z$  and/or  $C_Z$ , while the change in  $R_Z$  also varies the dc/low-frequency gain at the same time.

Two common design considerations are involved with the CTLEs. First, the designed peaking frequency is critical to the CTLE performance, while this peaking frequency can be sensitive to the process, voltage, temperature (PVT) variations. Therefore, it may need a tuning mechanism that correspondingly calibrates the peaking frequency to the optimal value. Second, since the achievable gain at high frequencies is limited by the technology-dependent gain-bandwidth product, designing for a large amount of peaking magnitude (i.e., the ratio of peak gain to the low-frequency gain) often implies that the low-frequency gain is smaller than unity. In other words, instead of amplifying, the aforementioned CTLE attenuates the low-frequency signal components. Since a CTLE effectively expands the overall channel bandwidth with the aim to ameliorate the attenuation of high-frequency signal components, it is capable of mitigating both the pre-cursor ISI and the post-cursor ISI. This attractive feature has motivated the inclusion of CTLE stages as parts of the receiver front-end circuits in prior arts, e.g., [9] and [10].

#### **2.4 Receiver-Side Feedforward Equalizer (RX-FFE)**

The RX-FFE is conceptually identical to the TX-FFE. In both cases, a linear finite impulse response (FIR) filter, as shown in Fig. 2.1, is constructed as the linear combination of multiple inputs spaced apart in time. As described in the previous section (Section 2.1), this FIR filter and thus the RX-FFE are responsible for emphasizing the high-frequency signal components. Equivalently, in the time domain, the RX-FFE shapes the pulse response, aiming for reduced ISI. As a consequence, when the FFE weights are optimized, the RX-FFE is capable of compensating for both pre-cursor and post-cursor ISI. It is also common to have the RX-FFE focus on one type of the ISI. For example, the RX-FFE can be configured to mostly tackle the pre-cursor ISI, while the post-cursor ISI is left for the receiver-side decision feedback equalizer (RX-DFE) to deal with. More details of the RX-DFE will be presented in the next section.

In contrast to the case of TX-FFE, where the data signals are in digital-fashion and hence digital gates/circuits can be employed to implement the delay elements, the received data signals appear in analog-fashion to the RX-FFE. Accordingly, the implementation of a RX-FFE requires the circuits that process multiple analog signals for the weighted sum.

One possible approach is delaying the analog input signal with analog-fashion delay elements in order that multiple analog inputs corresponding to the signals received at different times are presented together to the FFE summer. The analog delay elements can be realized with *LC* delay lines, transmission lines, or active circuit stages [11]. One major concern of the utilization of *LC* delay lines or transmission lines as the delay elements is the relatively high area consumption. Moreover, the losses from cascaded *LC*-line or transmission-line stages, as well as the considerable power consumptions attributed to the low impedance of these lines are of great concern [11], posing challenges to the RX-FFE designs based on analog delay lines. Alternatively, as demonstrated in [11], active transistors can be leveraged to implement the analog delay elements with improved area efficiency. In designing the active delay stages, two pivotal features should be targeted. For one thing, it is important to make the bandwidth of the active delay stages sufficiently large such that the signal impairment is tolerable. For the other thing, it would need calibration or compensation for the PVT variations in order to precisely control the amount of delay.

In contrast to the foregoing approach, where the signals, delay lines, and active delay stages are all in the analog fashion, another way to implement the RX-FFE is carrying out the FFE in the digital domain. In realizing a digital FFE, an ADC, or a time-interleaved ADC bank, is employed in the receiver to first digitize the analog signals into digital samples. In other words, this analog-to-digital conversion allows the input signals to be represented in the digital form of a given resolution, further enabling all the subsequent operations performed in the digital domain. More specifically, the delay elements, summers/adders, multipliers, and thus the FFE can be implemented with digital gates/circuits. The benefits of the digital RX-FFE include the following. First, the digital signal processing shows strong robustness against PVT variations and device mismatches. Second, since the CMOS technology scaling

favors digital circuits, the energy efficiency of the digital FFE improves with the advancement of the technology node. Third, with the assistance of mature computer-aided-design (CAD) tools, the hardware implementations of digital circuits are portable to different technologies through the automated synthesis. Furthermore, with the ease of delaying digital signals with CMOS gates, it is feasible to implement a long-tap digital FFE that offers superior equalization capability targeting high-loss channels. For example, in [9], a 31-tap digital FFE is incorporated in an ADC-based receiver, empowering 112-Gb/s data transmission over a channel with 37.5-dB loss at Nyquist frequency. For the purposes of meeting the timing constraints and/or optimizing the power consumption of a digital FFE, design techniques including pipelining, parallelization, and power supply reduction can be further applied.

Despite that a digital RX-FFE can lead to favorably strong and robust equalization, the demand for a high-speed ADC in the receiver data path may result in considerable power overhead. Consequently, while the superiority of a digital RX-FFE in realizing a long-tap FFE is still appreciated, this digital-fashion FFE may not be the optimal option if only a short-tap FFE is needed. In that case, the third method for the FFE implementation, which utilizes multi-phase sampling, serves as a promising candidate. The principle of the FFE based on multi-phase sampling is described as what follows. The analog input signals appearing at different times are sampled and then respectively held for the subsequent weighted summing performed at the same time. To implement this function, sample-and-hold circuits are employed and commonly clocked with multi-phase sub-rate clocks. The distinct phases of the clocks correspond to the different sampling instants. In Section 3.4, an implementation example, known as the double sampling technique, will be presented. The double sampling takes two samples spaced with one UI and sums them with desirable weights, essentially functioning as a discrete-time 2-tap FFE. This discrete-time implementation of FFE, enabled by the multi-phase sampling, can achieve improved power efficiency over its analog FFE counterpart by excluding the analog delay elements. Additionally, compared to the decision feedback equalizer with infinite impulse response (DFE-IIR), which will be described in next subsection (Section 2.5.4), the FFE based on



multi-phase sampling can also be more energy-efficient since the output multiplexers required in the DFE-IIR can be eliminated. Notwithstanding the aforementioned attractive power efficiency, it would not be straightforward to apply the multi-phase sampling technique to long-tap FFE implementations, since preserving the sampled analog value with high accuracy for a relatively long time (e.g., many UIs) is challenging.

To sum up, configured to have high-pass characteristics, RX-FFEs emphasize the high-frequency signal components and hence ameliorate the signal impairments due to ISI. However, since the noise and crosstalk are high-pass-filtered by the RX-FFEs as well, the high-frequency noise/crosstalk is amplified in the meantime, which is accordingly referred to as noise/crosstalk enhancement in the literatures. With the presence of the noise/crosstalk enhancement, RX-FFEs can still improve the overall eye-openings and bit-error-rate performance by removing the ISI or making the residual ISI insignificant.

## **2.5 Receiver-Side Decision Feedback Equalizer (RX-DFE)**

### **2.5.1 Direct DFE-FIR**

The concept of a decision feedback equalizer (DFE) is depicted in Fig. 2.6, in which the post-cursor ISI appearing in the uncompensated pulse response can be mitigated by the feedback signal. This compensation is made possible by first making correct decisions on the previously received signals and then correspondingly adjusting the polarity as well as the magnitude of the feedback signal in order to counteract the post-cursor ISI. More specifically, the architecture shown in Fig. 2.6 is known as DFE-FIR in that the feedback path consists of an FIR filter. With an  $n$ -tap FIR filter employed in the feedback, an  $n$ -tap DFE can be constructed, enabling the compensation for  $n$ -tap post-cursor ISI.

Unlike an FFE capable of addressing both pre-cursor and post-cursor ISI, a DFE can only tackle the post-cursor ISI, as a consequence of the need of decoding the previous symbols. Nonetheless, DFEs offer several appealing benefits. For one thing, by mitigating the post-cursor ISI, DFEs effectively emphasize the high-frequency signal components, whereas the noise/crosstalk enhancement for FFEs would not be a concern for DFEs, attributed to the

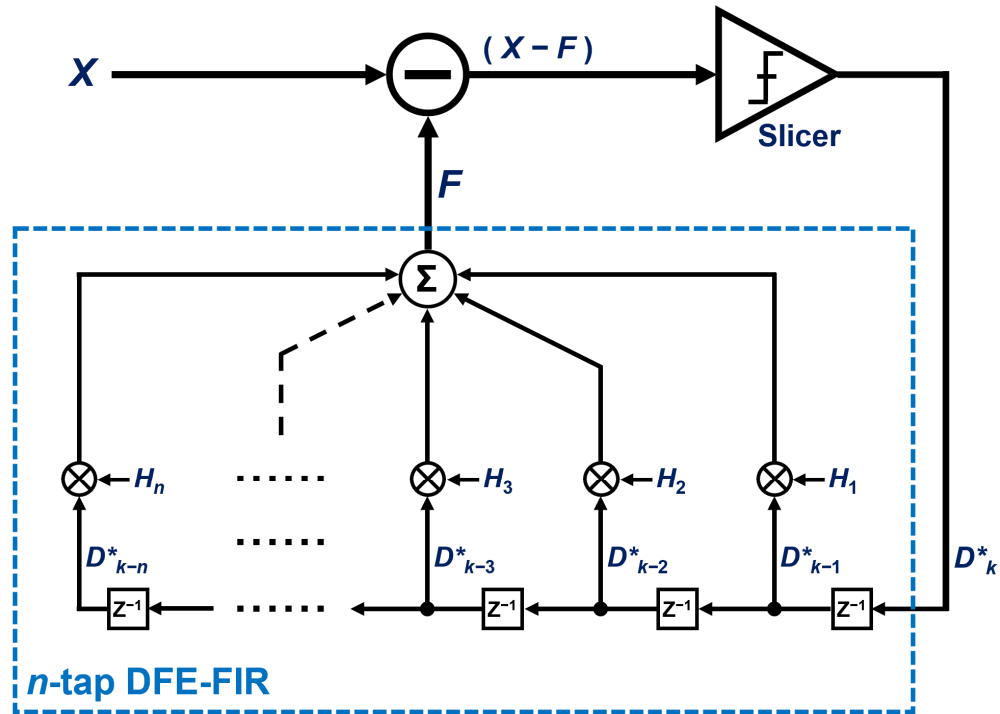


Fig. 2.6. Architecture of a direct DFE, with an  $n$ -tap FIR filter in the feedback path, referred to as an  $n$ -tap DFE-FIR.

digital-level output signals offered by well-designed decision circuits. For another thing, DFEs can succeed in compensating for the post-cursor ISI stemming from the reflections, when there are impedance discontinuities existing in the signal path. Especially for the cases where the reflections cause spectral notches, the efficacy of a DFE can be superior to that of an FFE [12].

In light of these advantages, it has been a favorable option to include a DFE in the receiver. Nevertheless, DFE implementations for high-speed operations demand efforts dedicated to meeting the stringent timing constraint. Referring to Fig. 2.6, this DFE architecture falls into the category of direct DFE, where the resolved data signals (i.e., the decisions) are scaled and directly fed back to the summer. The timing delays within this feedback loop lead to the timing constraints for successful post-cursor ISI compensation. For the  $N$ -th post-cursor ISI compensation, the timing constraint in a direct DFE design can be expressed as:

$$T_{\text{CKQ}} + T_{\text{dh}N} + T_{\text{settle}} + T_{\text{setup}} < N \times 1\text{UI} \quad (2.4)$$

where  $T_{\text{CKQ}}$  is the clock-to-Q delay of the slicer,  $T_{\text{dh}N}$  is the propagation delay of the  $N$ -th DFE tap,  $T_{\text{settle}}$  is the settling time of the summer, and  $T_{\text{setup}}$  is the setup time of the slicer. With the details of these timing delays presented in Chapter 4, it can be seen from (2.4) that the first-tap DFE poses the most stringent timing constraint, which can be the bottleneck in implementing first-tap DFE at high data rates. Therefore, as will be presented in Chapter 4, the key to realizing a direct first-tap DFE is reducing the timing delay term(s) appearing in (2.4) such that all the operations are finished within 1 UI.

### 2.5.2 Loop-Unrolling DFE

An alternative DFE architecture, commonly known as the loop-unrolling DFE, is shown in Fig. 2.7, in which only the first tap is unrolled as a simple illustration. The main idea of a loop-unrolling DFE relies on that all the possible equalized results are pre-computed, among which only one of them will be selected by a multiplexer depending on the previous decisions. The major benefit enabled by the loop-unrolling architecture is the relaxed timing constraint. For the first-tap loop-unrolling DFE, its timing constraint becomes [13]:

$$T_{\text{CKQ}} + T_{\text{setup}} + T_{\text{mux}} < 1\text{UI} \quad (2.5)$$

where  $T_{\text{CKQ}}$  is the clock-to-Q delay,  $T_{\text{setup}}$  is the setup time, and  $T_{\text{mux}}$  is the propagation delay of the multiplexer (MUX). By comparing (2.5) and (2.4) when  $N = 1$ , it suggests that the loop-unrolling technique eases the DFE implementations at high data rates, since  $T_{\text{mux}}$  is smaller than  $(T_{\text{dh}1} + T_{\text{settle}})$  in most cases.

However, as displayed in Fig. 2.7, it costs extra hardware (e.g., slicers) and potentially higher power consumption to realize the loop-unrolling DFEs. More specifically, the number of required slicers increases exponentially with the number of taps unrolled. That is, if  $N$ -tap DFE is designed with the loop-unrolling fashion in NRZ systems, then the demand for  $2^N$  slicers is expected. This hardware requirement is even more demanding for a high-order

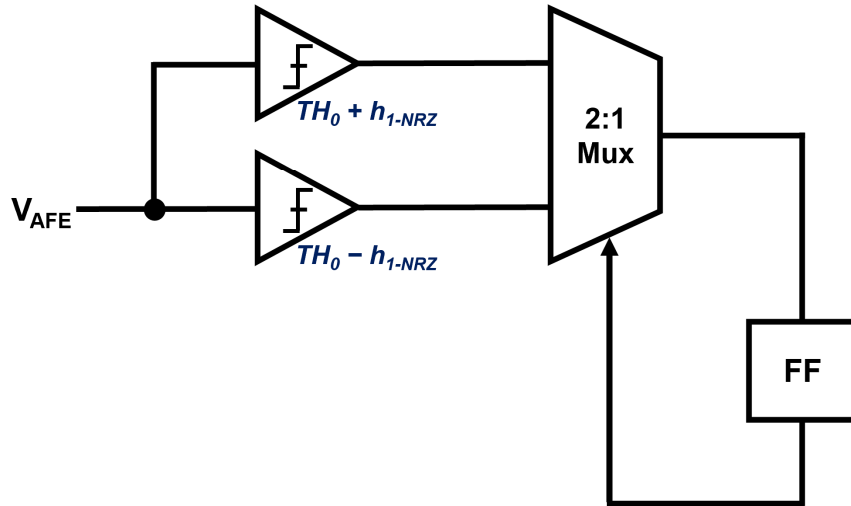


Fig. 2.7. Loop-unrolling DFE with the first tap unrolled for NRZ systems.

modulation format. For instance, implementing  $N$ -tap loop-unrolling DFE in PAM4 systems results in the significant increase in the required number of slicers, proportional to  $4^N$ . Corresponding to the three distinct voltage thresholds in the PAM4 systems, it would need 12 slicers, 3 multiplexers, and one thermometer-to-binary decoder in each deserialized data path, even if only one tap of the DFE is unrolled, as shown in Fig. 2.8.

In view of that unrolling a large number of taps in NRZ systems or unrolling the first few taps within a high-order-modulated system can both give rise to prohibitively expensive hardware and power consumptions, it is more common to apply the loop-unrolling technique for the first few DFE tap(s) in NRZ systems.

### 2.5.3 Look-Ahead Multiplexing DFE

To further relax the timing constraint of a feedback loop involving multiplexers, the look-ahead multiplexing technique has been developed. Proposed in [14], the principle of this look-ahead multiplexing technique can be explained with the following equations [14]:

$$D_n = H_n D_{n-1} + L_n \overline{D_{n-1}} \quad (2.6)$$

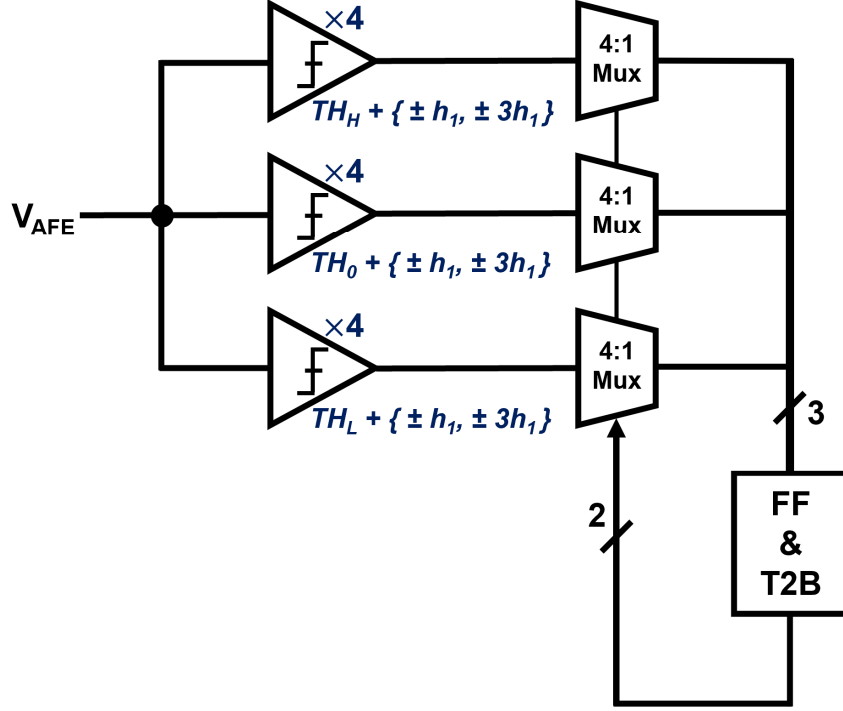


Fig. 2.8. Loop-unrolling DFE with the first tap unrolled for PAM4 systems.

where the subscripts  $n$  and  $n - 1$  denote the timing order, at time  $n$  and  $n - 1$ , respectively;  $D$  is the multiplexer output;  $H$  and  $L$  are the two inputs of the multiplexer. As expressed in (2.6), the value of  $D_n$  depends on the multiplexer output at the previous time along with its complement, i.e.,  $D_{n-1}$  and  $\overline{D_{n-1}}$ . Similarly,

$$D_{n-1} = H_{n-1} D_{n-2} + L_{n-1} \overline{D_{n-2}} \quad (2.7)$$

where the dependence on the previous output at the time  $n - 2$  is also observed in deciding the value of  $D_{n-1}$ . By substituting  $D_{n-1}$  in (2.6) with (2.7), the following expression for  $D_n$  can be derived [14]:

$$D_n = (H_n H_{n-1} + L_n \overline{H_{n-1}}) D_{n-2} + (H_n L_{n-1} + L_n \overline{L_{n-1}}) \overline{D_{n-2}} \quad (2.8)$$

where the value of  $D_n$  is now dependent on  $D_{n-2}$  and its complement  $\overline{D_{n-2}}$ , along with the computed terms in the parentheses. By comparing (2.6) with (2.8), it shows the look-ahead

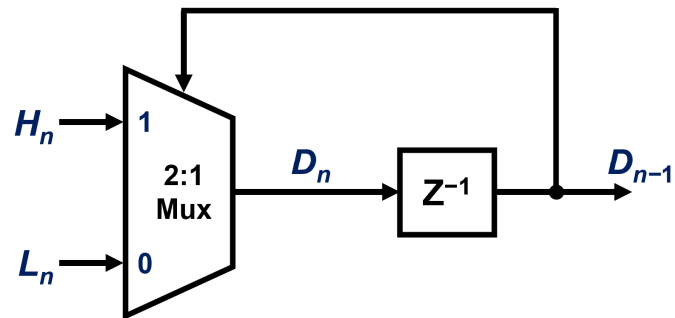


Fig. 2.9. Implementation of a 2-to-1 multiplexer loop.

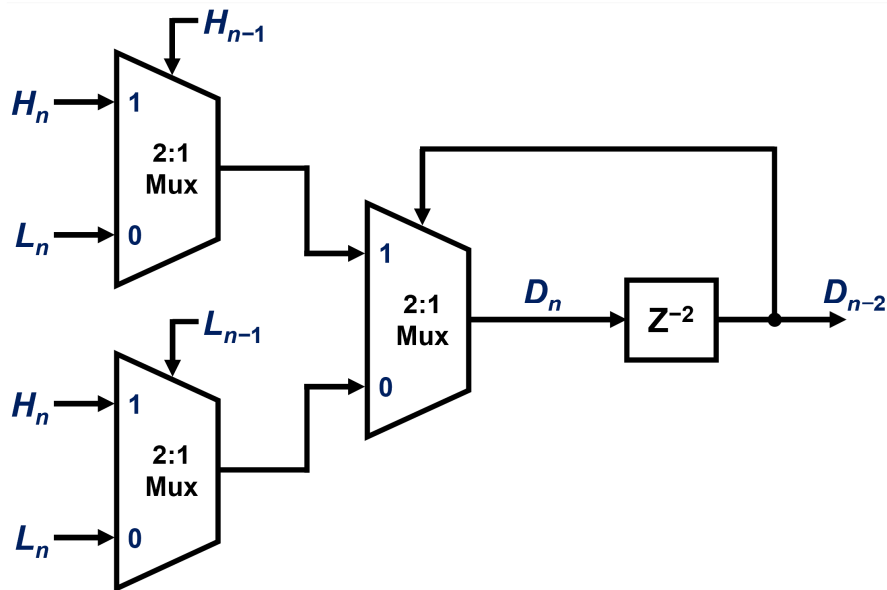


Fig. 2.10. Implementation of a 2-to-1 multiplexer loop, with look-ahead factor of 1.

multiplexing technique brings the key benefit that the timing constraint can be significantly relaxed, as the iteration bound is doubled at the expense of extra hardware. These hardware implementations are illustrated in Fig. 2.9 and Fig. 2.10.

As pointed out in [12], in order to implement the look-ahead multiplexing technique, a larger number of multiplexers needs to be invested. The number of required multiplexers increases

with the look-ahead factor, and the increase becomes more drastic when a high-order modulation format is adopted [12].

#### **2.5.4 Decision Feedback Equalizer with Infinite Impulse Response (DFE-IIR)**

The previously presented DFE architectures can be considered as the variants of the DFE-FIR shown in Fig. 2.6, where each DFE tap is dedicated to mitigating the ISI of one specific post cursor. In consequence, when it comes to a pulse response with long-tail post-cursor ISI (i.e., with considerable post-cursor ISI over a relatively large time-span), equalizing the long-tail post-cursor ISI with the DFE-FIR architecture would require a large number of taps and thus cost considerable hardware and power overheads. For the purpose of avoiding a large number of DFE taps, a decision feedback equalizer with infinite impulse response (DFE-IIR) can serve as a promising alternative in certain scenarios.

The architecture of a DFE-IIR is shown in Fig. 2.11. Considering the case where the long-tail post-cursor ISI profile can be approximated with the response of an IIR filter, it is feasible to include the IIR filter in the feedback path in order that the resulting feedback signal resembles the targeted long-tail post-cursor ISI. By subtracting the feedback signal with summers, most or parts of the long-tail post-cursor ISI can be compensated. The most common implementation examples are those where the channels behave like first-order low-pass filters, and therefore simple *RC* low-pass filters can be employed as the IIR feedback filters. The employment of a DFE-IIR does not conflict with the inclusion of a DFE-FIR. In effect, the simultaneous use of a DFE-IIR along with a DFE-FIR can potentially lead to improved equalization effectiveness.

The benefits of utilizing a DFE-IIR together with a DFE-FIR have been demonstrated in [15], in which the first post-cursor ISI is compensated by a 1-tap DFE-FIR whereas the long-tail ISI is compensated by a 2-tap DFE-IIR. Compared to a design only with 2-tap DFE-IIR, the addition of the 1-tap DFE-FIR makes the equalization performance less sensitive to the loop delay as well as the coefficient variations [15]. Meanwhile, the 2-tap DFE-IIR consisting of two IIR filters with different time constants (i.e., different bandwidths) offers superior performance, when compared to a 1-tap DFE-IIR with only one filter contributing to the

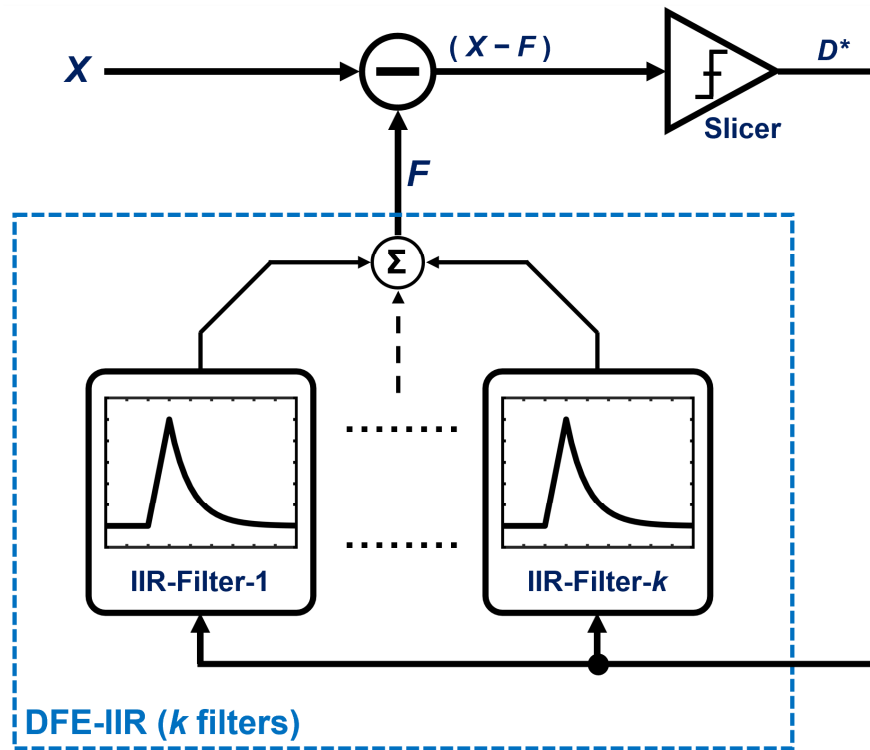


Fig. 2.11. Architecture of a DFE-IIR equalizer, with  $k$  filters included in the feedback path. A full-rate implementation is shown.

approximation for long-tail post-cursor ISI [15]. From the implementation perspective, it is noteworthy that a multiplexer needs to be included as a part of the feedback data path in a sub-rate DFE-IIR design. For instance, the DFE-IIR in [15] is implemented in the half-rate fashion; consequently, a 2:1 differential multiplexer responsible for multiplexing the data from two half-rate paths is necessary, as shown in Fig. 2.12.

## 2.6 Receiver-Side Nonlinear Equalization

In contrast to the transmitter-side equalization which requires a back-channel for its adaptation, receiver-side equalization techniques, including those targeting nonlinearity compensation, are capable of tackling the accumulated signal impairments without a back-channel. In other words, the linearity and/or bandwidth degradation resulting from the channel and the receiver front-end circuits can be captured at the receiver side and further ameliorated by the receiver-side equalizers. Moreover, when an ADC is included in the



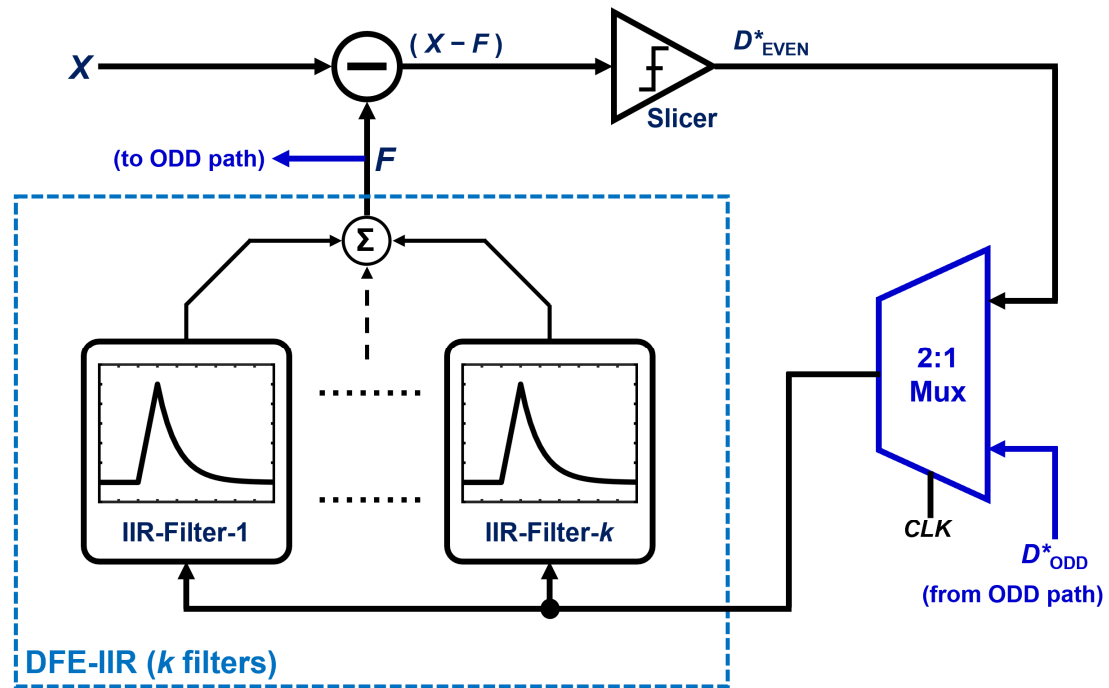


Fig. 2.12. Architecture of a DFE-IIR equalizer, with  $k$  filters included in the feedback path. A half-rate implementation is shown.

receiver, receiver-side linear and nonlinear equalizers can be implemented in the digital domain. As with the case of digital RX-FFE described previously, a digital nonlinear equalizer holds strong immunity to the PVT variations and at the same time benefits from the CMOS technology scaling.

One conventional nonlinear equalizer is famous as the Volterra equalizer. As its name suggests, the Volterra equalizer is based on the Volterra series, and the equalizer coefficients, or kernels in some contexts, are optimized so as to counteract the existing nonlinearities. A third-order Volterra equalizer can be mathematically expressed as:

$$\begin{aligned}
 y(n) = & \sum_{k_1=0}^{m_1-1} h_1(k_1) x(n - k_1) + \sum_{k_1=0}^{m_2-1} \sum_{k_2=0}^{k_1} h_2(k_1, k_2) \prod_{j=1}^2 x(n - k_j) \\
 & + \sum_{k_1=0}^{m_3-1} \sum_{k_2=0}^{k_1} \sum_{k_3=0}^{k_2} h_3(k_1, k_2, k_3) \prod_{j=1}^3 x(n - k_j)
 \end{aligned} \tag{2.9}$$

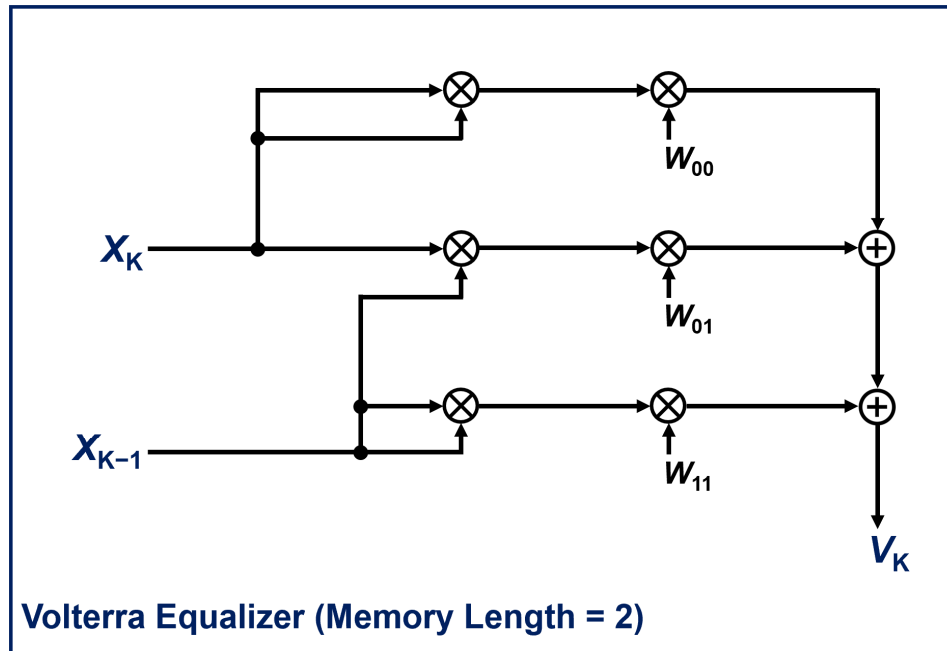


Fig. 2.13. Implementation of a second-order Volterra equalizer with memory length 2.

where  $x(n)$  is the  $n$ -th input sample;  $y(n)$  is the  $n$ -th sample of the Volterra equalizer output;  $m_1, m_2$ , and  $m_3$  are the memory lengths of the first, second, and third order terms, respectively;  $h_1, h_2$ , and  $h_3$  respectively denote the first, second, and third order Volterra kernels.

As can be observed from (2.9), the operation of a conventional Volterra equalizer relies on generating multiplicative high-order terms. Hence, a digital multiplier is identified as a critical building block in implementing a conventional Volterra equalizer. Furthermore, it can also be inferred from (2.9) that the number of required multipliers increases dramatically with the memory length as well as the order of the Volterra equalizer. For example, with  $m_1 = 2$  in (2.9), two multipliers are required to generate the first-order terms, while with  $m_2 = 2$  in (2.9), as shown in Fig. 2.13, six multipliers are required to generate the second-order terms. When  $m_1$  and  $m_2$  in (2.9) are both increased to 3, generating the first-order terms needs three multipliers, whereas generating the second-order terms needs twelve multipliers. In view of that the multiplicative computations for the high-order terms can cost considerable power

overhead, Chapter 5 elaborates design techniques for nonlinear equalization, which aim at a reduced number of employed multipliers and thus improved energy efficiency.

## AVALANCHE PHOTODETECTOR (APD)-BASED BURST-MODE OPTICAL RECEIVER

### 3.1 Overview

Optical interconnects have wide applications in modern data communication and computing systems, including data center networks. The roadmaps for optical interconnects in data centers [16] require significant improvements in various metrics. Within the span of a decade, it is proposed that the speed of optical links in the data centers increases by a factor of 25, the energy efficiency is improved by a factor of 5, and the optical switching speed reduces from 10 ns to 100 ps [16]. In order to realize the envisioned specifications, efforts have been incited to not only advance the high-speed optical devices such as modulators and photodetectors but also innovate the electronic circuit design for offering a superior interface and better energy efficiency, e.g., [17]–[21]. In this paper, an optical receiver, which leverages the advancement of avalanche photodetector (APD) and new electronic circuit topologies for high sensitivity and fast reconfigurations, is presented.

Despite the small modulation-frequency-dependent loss introduced by the optical fibers, modulation-frequency-independent signal attenuation and proportional losses (for multimode fiber, the loss is about 1.5 dB/km for 1300-nm signals; for single mode fiber, the loss is about 0.5 dB/km for 1310-nm signals) can be considerable in an optical network where long fibers and a large number of connectors, couplers, or splitters are involved. To overcome the attenuation and losses, the laser power needs to be augmented. With a given level of attenuation along the signal path, improvement in energy efficiency of optical links can be achieved with the availability of high-sensitivity receivers. Designing a high-sensitivity optical receiver using an APD along with the energy-efficient equalization techniques implemented in modern CMOS technology is one of the main goals of this paper.

In a rapidly reconfigurable optical network, different data bursts originating from different transmitters can present distinct dc, amplitude, and phase characteristics, as illustrated in Fig.

3.1(a). A burst-mode receiver (BMRX) capable of performing reconfigurations to adapt itself to the variability, prior to the real data transmission, is essential. Fig. 3.1(b) shows a simplified timing diagram of the burst-mode reconfiguration scheme; the receiver needs to cancel the dc offset, control the signal amplitude for linear operations, and also recover the sampling clocks, before the transmission of the data payload. The aforementioned reconfigurations lead to an overhead time whenever a different data burst arrives, and consequently, the link latency and bandwidth can be improved by reducing the overhead, i.e., the overall reconfiguration time, especially for a network where switching events occur frequently. *RC* low-pass filter (LPF)-based designs are conventionally applied to extract the dc and amplitude information [19], [22], whereas the inevitable tradeoff between the tracking time and the settling behavior of *RC* LPF forms a bottleneck in reducing the reconfiguration time. Prior arts have employed various design techniques to improve the settling time. For instance, the work in [39] uses a feedback-type automatic offset compensation (AOC) loop with switchable bandwidth to remove the input dc offset in less than 75 ns for 10-Gb/s operations. A feed-forward type AOC is applied in [40] achieving 25.6-ns response time for 10-Gb/s operations with tradeoffs in accuracy and power consumption, as indicated in [39]. A calibration state machine is designed along with *RC* LPF in [19], which completes the search for the settings associated with dc component cancellation in 12.5 ns at 25-Gb/s operations. We propose an integrating dc comparator and an integrating amplitude comparator in this paper to enable fast cancellation of the dc offset, and rapid signal amplitude control, respectively. The proposed integrating dc and amplitude comparators eliminate the *RC* settling time constraints, and as will be shown in Sections 3.5.2 and 3.5.3, the minimum comparison time is reduced to two unit intervals (UIs), empowering significant acceleration of the burst-mode reconfiguration and scaling with the data rate. Furthermore, due to the nature of performing integration, the proposed integrating dc and amplitude comparators do not require the clock and data recovery circuits (CDR) to be locked in advance.

This paper is organized as follows. Section 3.2 reviews the basics of APD, its advantages, and challenges. Section 3.3 presents the overall APD-based receiver architecture. Section 3.4

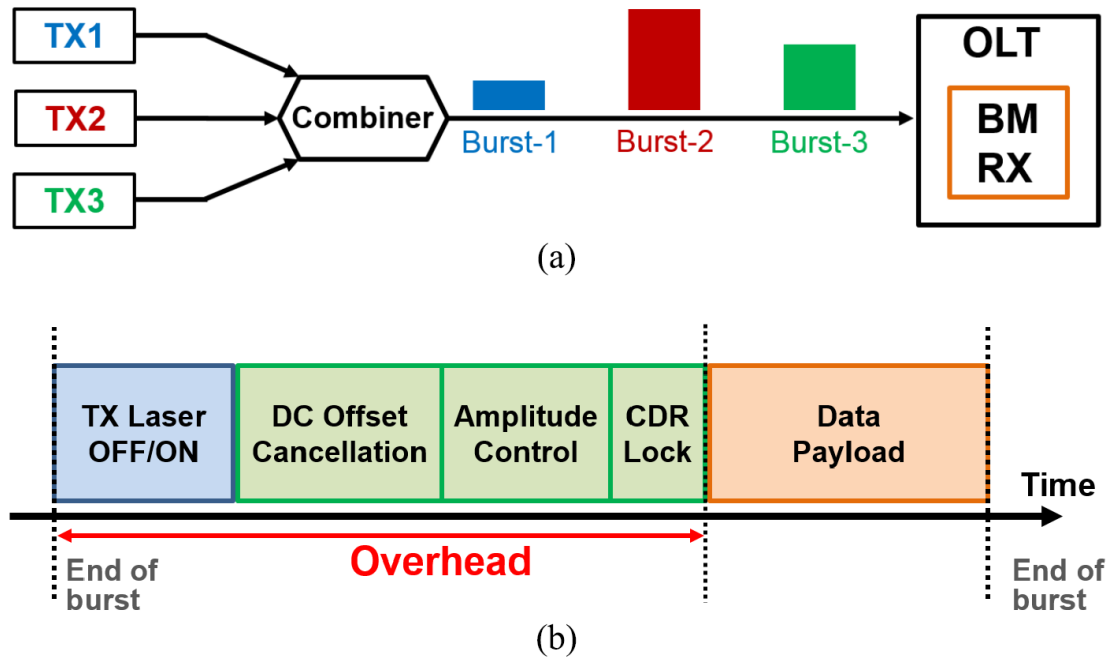


Fig. 3.1. (a) Transmission of distinct data bursts originated from different transmitters to a single optical line terminal (OLT), including a burst-mode optical receiver (BMRX). (b) Simplified timing diagram of burst-mode reconfiguration scheme, in which the dc offset cancellation and amplitude control are the focus of this paper.

describes the equalization circuits designed in current-integrating fashion. Section 3.5 explains the operation of the burst-mode reconfiguration loops and elaborates the principles and implementations of the proposed integrating dc and amplitude comparators. The experimental results of this burst-mode optical receiver are shown in Section 3.6, and finally, Section 3.7 summarizes this paper with performance comparisons and conclusions.

### 3.2 Avalanche Photodetector (APD)

Friis' formula for noise figure [23] suggests that a high-gain stage at the front end is favorable in suppressing the noise contribution from succeeding stages to the overall signal-to-noise ratio (SNR). This motivates the use of APD since APD offers gain that increases the photocurrent by a multiplication factor of  $M$  as the very first stage of the receiver, and the ongoing advancements in the gain-bandwidth product of APD [24]–[26] have made APD more and more suitable for high-speed data communication. Nevertheless, since the APD gain arises from the generation of secondary electron-hole pairs through the impact

ionization process, and these pairs are generated at random times [27], the shot noise of APD is enhanced by the excess noise factor,  $F$ , given by

$$F = kM + (1 - k) (2 - 1/M) \quad (3.1)$$

where  $k = \alpha_e/\alpha_h$  if  $\alpha_h > \alpha_e$ , or  $k = \alpha_h/\alpha_e$  if  $\alpha_e > \alpha_h$  by definition, while  $\alpha_e$  and  $\alpha_h$  denote the impact ionization coefficients for electrons and holes, respectively [27]. With the incident optical power represented by  $P$ , the dark current represented by  $I_d$ , the magnitude of electron charge represented by  $q$ , the responsivity of the photodetector represented by  $R$ , the effective noise bandwidth of the receiver represented by  $\Delta f$ , the thermal noise power represented by  $N_T$ , the shot noise power, denoted by  $N_S$ , and the SNR of an APD-based front end can be, respectively, written as: [27], [28]

$$N_S = 2qM^2F (I_d + RP) \Delta f \quad (3.2)$$

$$\text{SNR} = (MRP)^2 / (N_S + N_T) \quad (3.3)$$

A few observations can be made from (3.1) to (3.3). First, when  $M$  is set to 1,  $F$  equals 1 in (3.1), implying the absence of excess shot noise, and the resulting expressions for (3.2) and (3.3) correspond to the case of using a p-i-n photodetector. Second, provided that the thermal noise is dominant over the shot noise, i.e.,  $N_T \gg N_S$ , the signal power increases with the gain ( $M$ ) quadratically, and hence, the improvement in SNR by a factor of approximately  $M^2$  can be achieved as long as the gain-independent thermal noise keeps dominating the noise contribution. Therefore, compared to a p-i-n photodetector with similar bandwidth, APD considerably benefits the receiver sensitivity in the thermal-noise-limited regime. On the contrary, in the case of being shot-noise limited, i.e.,  $N_S \gg N_T$ , it can be inferred from (3.2) and (3.3) that the SNR can no longer be improved by increasing  $M$ , and as a matter of fact, the SNR is degraded by the excess noise factor  $F$ , in comparison with a p-i-n photodetector having similar bandwidth. The foregoing suggests that there exists an optimum value of gain  $M$ , which gives rise to the maximum SNR; the optimum value of  $M$  can be found by solving (3.2) and (3.3). Fig. 3.2 shows the SNR improvements versus  $M$  with a given level of optical

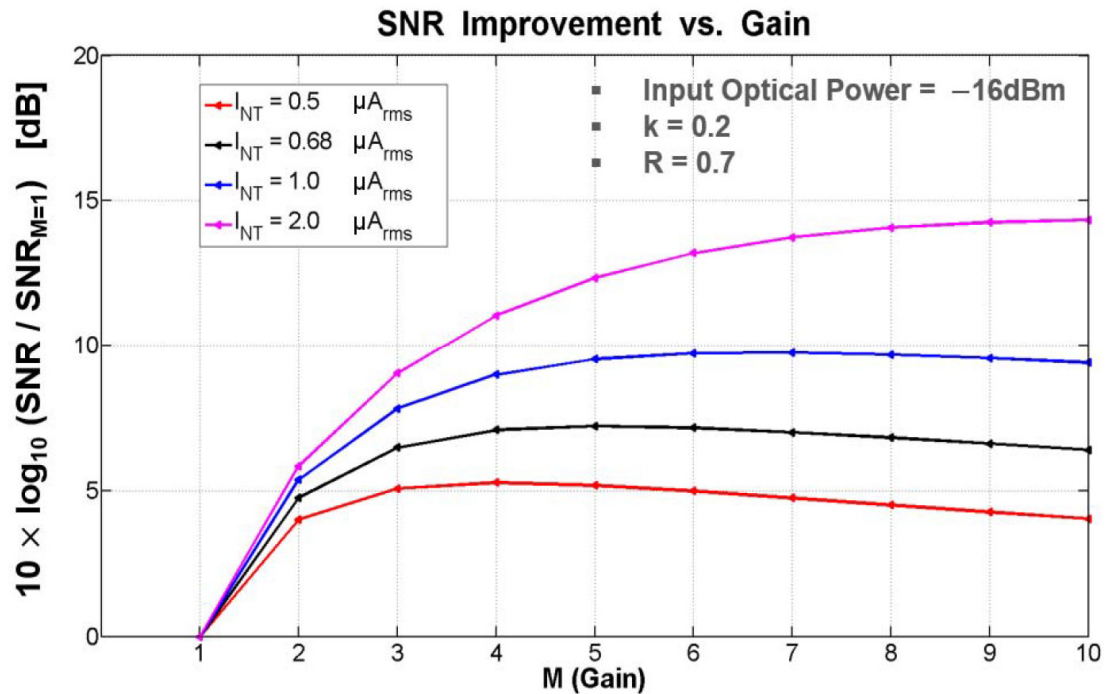


Fig. 3.2. SNR improvements (in decibels) versus  $M$  with a given level of optical input power ( $-16$  dBm).  $k = 0.2$ ,  $R = 0.7$ , and different amounts of input-referred thermal noise are used in the computations.

input power while different amounts of input-referred thermal noise ( $I_{NT}$ ) are present. In this design, the input-referred noise current of the receiver from the simulation is  $0.68 \mu A_{rms}$ , and the overall responsivity of APD is set to be  $4$  A/W, corresponding to a multiplication factor or gain of  $5.7$  approximately.

In addition to the enhanced shot noise, the bandwidth of APD generally decreases with the gain because of the longer avalanche build-up time [26]. As the effective signal power can be compromised by the excess inter-symbol interference (ISI) due to the lower bandwidth, equalizer (EQ) circuits are included in this APD-based optical receiver for the purpose of ameliorating speed limitations formed by the APD gain-bandwidth tradeoff and the  $R_{in}C_{in}$  time constants as well, where  $R_{in}$  denotes the input resistance of the receiver, and  $C_{in}$  denotes the total capacitance at the receiver input.



### 3.3 APD-Based Optical Receiver Architecture

The architecture of the burst-mode optical receiver is shown in Fig. 3.3(a). The single-ended photocurrent is converted into differential voltage outputs by the analog front end (AFE), consisting of a variable current source (VCS) to subtract the dc component of the photocurrent, a three-stage inverter-based transimpedance amplifier (TIA), a differential pair-based single-ended-to-differential amplifier (S2D), two-stage current-steering variable gain amplifier (VGA), and a transconductance-C LPF (gm-C LPF) with 100-kHz bandwidth in a negative feedback loop for residual offset cancellation and combating low-frequency drifts. The circuit schematic of the VCS is shown in Fig. 3.3(b), where the value of  $V_{BIAS}$  and the ON/OFF states of the switches are determined by 8-bit digital setting (b0:b7). The 8-bit control of VCS is implemented in a binary-weighted fashion, and its tuning range can be adjusted by varying the tail current source of the V2I shown in Fig. 3.12(b). The idea of keeping the resolution (LSB) at 2% – 4% of the peak-to-peak ac current amplitude, proposed in [19], is adopted in this VCS design. The circuit schematic of the three-stage inverter-based TIA is shown in Fig. 3.3(c), and the feedback resistors,  $R_{F1}$  and  $R_{F2}$ , are designed to be 1.2 k and 275  $\Omega$ , respectively. In view of that the value of  $R_{F1}$  impacts on the SNR performance and the EQ specifications, the design considerations of  $R_{F1}$  are described together with the EQ in Section 3.4, while the value of  $R_{F2}$  is chosen such that the third-inverter stage with feedback resistor acts as an amplifier, and that  $R_{F2}$  does not considerably affect the overall AFE bandwidth. In addition, to better interface with the current-mode logic (CML) used in succeeding stages, the second-inverter stage in the TIA is sized so as to have the common-mode output voltage of TIA is  $\sim 635$  mV under 1-V supply. A conventional differential amplifier is used to implement the S2D, as shown in Fig. 3.3(d). The S2D is designed to have voltage gain 1.5 V/V, output common-mode voltage  $\sim 730$  mV, and  $-3$ -dB bandwidth 28 GHz when loaded with the VGA in this paper. The circuit schematic of the VGA is shown in Fig. 3.3(e), in which  $V_{B0}$  is a fixed bias voltage, while  $V_{B1}$  and  $V_{B2}$  are determined by 5-bit digital setting (b8:b12) such that a fixed amount of current  $I_{CM} = I_G + I_R$ , is steered between the branches with and without gain. The purpose of having a fixed value of  $I_{CM}$  is to keep the common-mode output voltages the same, independent of the gain setting. With the current-steering tuning mechanism and without adjusting the values of the load resistors, the

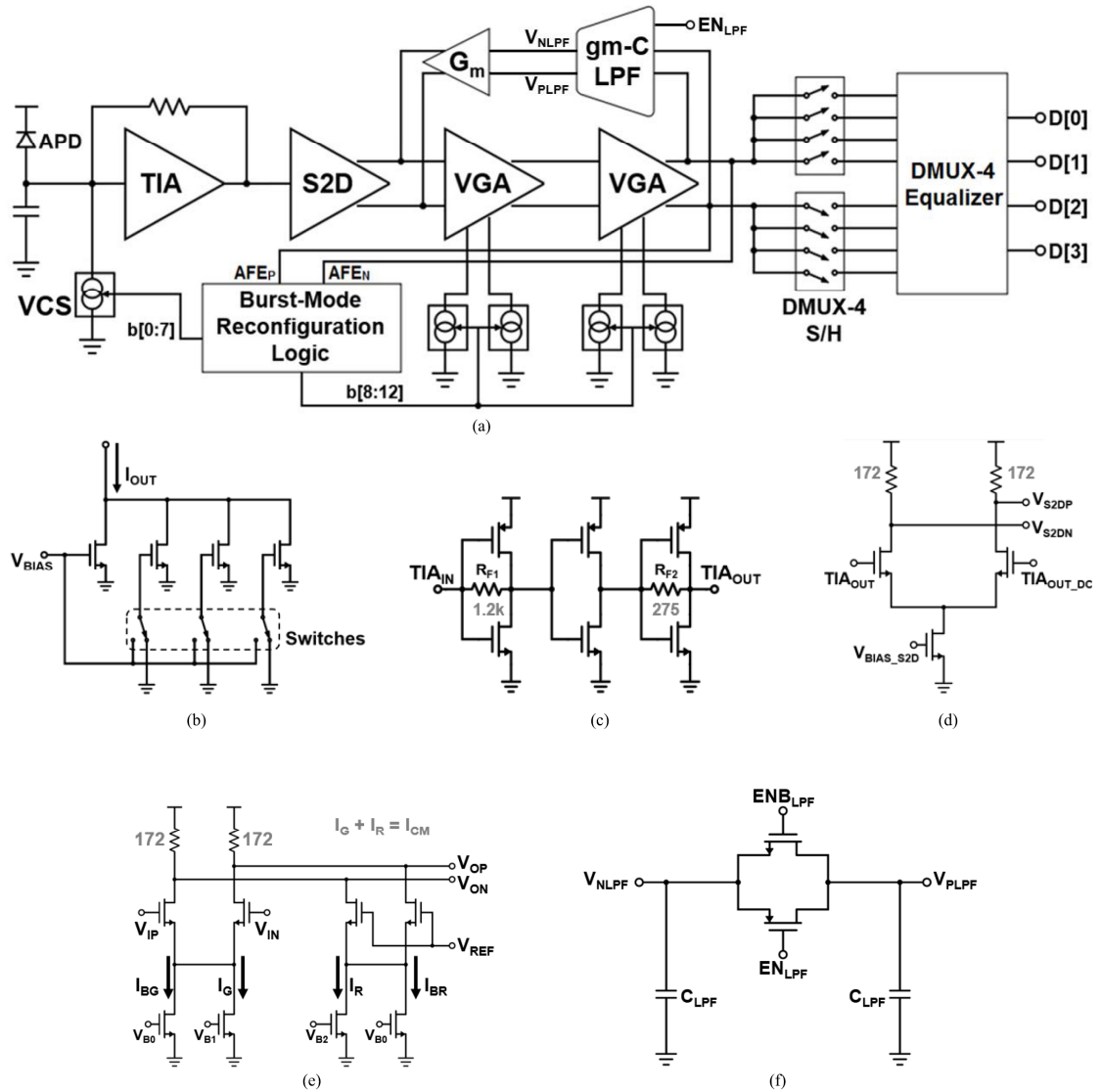


Fig. 3.3. (a) Architecture of the BMRX. (b) Circuit schematic of the VCS. (c) Circuit schematic of the three-stage inverter-based TIA.  $R_{F1} = 1.2 \text{ k}\Omega$  and  $R_{F2} = 275 \text{ }\Omega$  nominally in this design. (d) Circuit schematic of the single-ended-to-differential amplifier (S2D), with load resistors set to  $172 \text{ }\Omega$  in this design. (e) Circuit schematic of the current-steering VGA, with load resistors set to  $172 \text{ }\Omega$  in this design. (f) Circuit schematic of the enable/disable control scheme for the LPF loop.

bandwidth can be kept sufficiently constant among all gain settings for the VGA. The tuning range of the VGA gain per stage is from 0.95 to 1.67 V/V, and the 5-bit control is implemented in thermometer code fashion. Specifically, for the two-stage VGA in this design, when the 5-bit digital setting steps from (0, 0, 0, 0, 0), (0, 0, 0, 0, 1), (0, 0, 0, 1, 1), . . . ,

(1, 1, 1, 1, 1), the gain of the two-stage VGA is increased by a factor of 1.25 per step with the  $-3$ -dB bandwidth of the two-stage VGA kept at  $\sim 20$  GHz. From simulations, the 1-, 2-, and 3-dB compression points in the gain of each VGA stage are 227, 310, and 368 mV, respectively. The enable/disable control scheme for the LPF loop is shown in Fig. 3.3(f). When  $EN_{LPF}$  is set to logical low and  $ENB_{LPF}$  is set to logical high, the LPF loop is disabled by having  $V_{NLPF} \approx V_{PLPF}$ , introducing approximately zero offset to the AFE. The output of AFE is deserialized (1-to-4) by a bank of four sample-and-hold (S/H) switches, clocked by four quarter-rate clock phases. The S/H switch is implemented with a single transistor (pMOS) with a dummy transistor in series to mitigate the effects of charge injection as in [30]. Followed by a dedicated set of EQ and slicer, also clocked by the quarter-rate clock phases, each deserialized voltage sample is recovered to digital logic level. When a new data burst arrives with a “1010. . .” preamble pattern, the on-chip searching logic is designed to sequentially determine the optimum digital setting of (b0:b12) with respect to two goals. One is to cancel the dc offset and, in the meantime, retain the dc bias point by matching the dc component of the photocurrent with the current from VCS. The other is to control the signal amplitude by adjusting the gain of VGA, in order that linear operation is maintained, and the setting of the EQ circuits does not need to be updated with different data bursts possessing distinct power levels.

### 3.4 Equalizer Design

Increasing the value of the shunt-feedback resistor used in the TIA benefits in higher gain and lower noise at the receiver front end at the expense of eventually pushing the dominant pole toward low frequency, particularly with the presence of the capacitance from APD and wire-bond pad. When the frequency of the dominant pole is significantly smaller than the data rate, the long-tail post-cursor ISI is induced in the pulse response. The signal and noise analysis of a TIA front end employing an inverter with a shunt-feedback resistor, and the effects of varying the shunt-feedback resistor value on the TIA bandwidth have been studied in [35]. With the aim of optimizing the receiver sensitivity, in this paper, the shunt-feedback resistor [ $R_{F1}$  in Fig. 3.3(c)] is increased to the extent that the ISI can be effectively cancelled or mitigated by the succeeding EQ. With  $R_{F1}$  designed to be 1.2 k $\Omega$ , the three-stage TIA

achieves 67.16-dB $\Omega$  dc gain, 7.4 GHz  $-3$ -dB bandwidth, and the resultant  $-3$ -dB bandwidth of the AFE is 6 GHz from the simulation. The pulse response at the AFE outputs with  $-16$ -dBm optical modulation amplitude (OMA) input is simulated to determine the equalization scheme and the EQ coefficients, as shown in Fig. 3.4(a), where the peak value is  $\sim 253$  mV. In this paper, an EQ performing two-tap (including the main cursor) feed-forward equalization (FFE) and two-tap decision feedback equalization (DFE) in current-integrating fashion is designed such that the long-tail ISI can be mostly removed by the two-tap FFE, while the residual first and second post-cursor ISI are cancelled by the two-tap DFE, as illustrated in Fig. 3.4(b). Although FFE amplifies high-frequency noise, the sensitivity can be improved when the benefit arising from reducing the ISI by FFE surpasses the penalty of the enhanced noise. The pulse responses at the AFE outputs are also simulated with different input power levels in the range from  $-16$ - to  $-11$ -dBm (OMA), along with their corresponding gain settings of VGA to verify the following inequality is satisfied:

$$V_{\text{Main}} - \sum_k |ISI_k| > 7 \times (V_{\text{Noise}}) + 30 \text{ mV} \quad (3.4)$$

where  $V_{\text{Main}}$  denotes the main cursor magnitude;  $ISI_k$  denotes the residual ISI that is  $k$  UIs apart from the main cursor; the factor, 7, refers to the target bit-error-rate (BER)  $< 10^{-12}$ , and 30 mV is left as the decision margin for the data slicers.

The double-sampling technique, reported and analyzed in [29] and [30], serves as one form of implementing two-tap FFE in the discrete-time domain. It takes two signal samples spaced with one UI and sums up the two samples with appropriate weights. As described in [30], the double-sampling technique is effective in equalizing a channel that well resembles a first-order  $RC$  low-pass system since the long-tail ISI can be cancelled by having the following satisfied:

$$\beta_{\text{DS}} = 1 - \exp(-T_b / T_{\text{RC}}) \quad (3.5)$$

in which  $T_b$  is the bit interval,  $T_{\text{RC}}$  is the  $RC$  time constant, and  $(\beta_{\text{DS}} - 1)$  is the ratio of the summing coefficient of the previous sample to that of the current sample. In addition, the

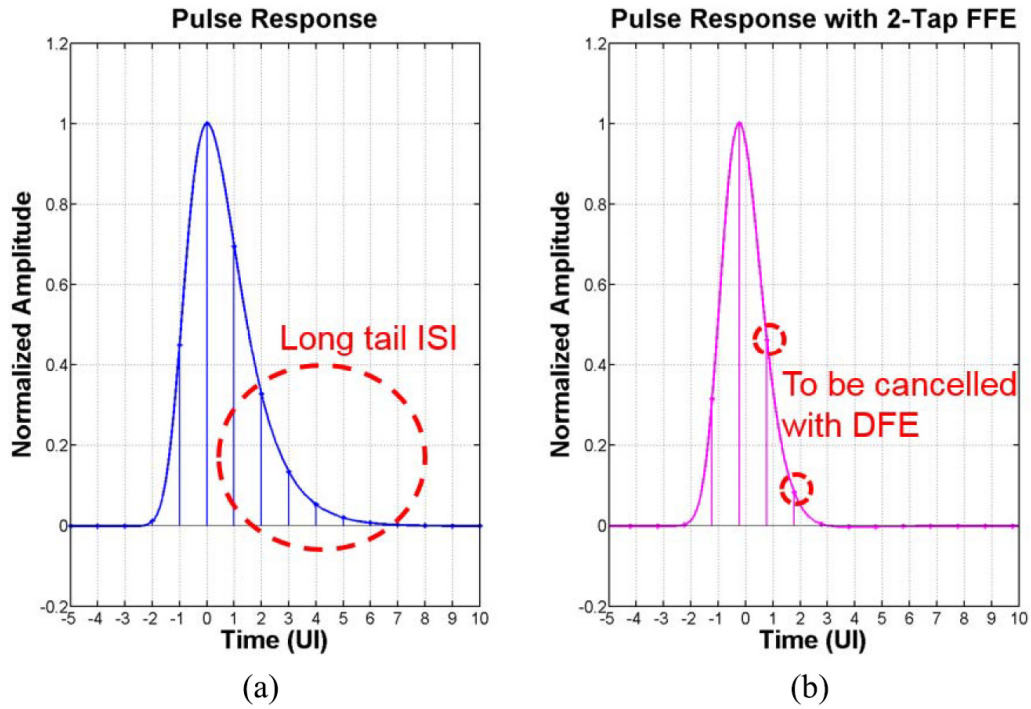


Fig. 3.4. (a) Pulse responses at the AFE outputs before applying equalization. (b) Pulse responses at the AFE outputs after applying ideal two-tap FFE.

double-sampling technique is energy efficient in comparison to both an infinite impulse response DFE (DFE-IIR) and an analog FFE by virtue of the dispensability of the output multiplexer after deserialization as well as the implementation of analog delay elements. In this design, the resistively loaded summer in [30] is replaced with a current-integrating summer to improve the settling time, and another DFE tap (second-tap DFE) is included. The schematic of the EQ is shown in Fig. 3.5, consisting of a current-integrating summer connected to the two-stage regenerative slicer embedding the first-tap DFE. The clock phases are designed for quarter rate operations, similar to [37], and such that  $SUM_P[n]$  and  $SUM_N[n]$  nodes shown in Fig. 3.5 are pre-charged to the supply voltage prior to the current integration over a single UI. At the end of the integration phase, the differential output voltage ( $SUM_P[n] - SUM_N[n]$ ) is the weighted sum or the equalized value, as the result of performing two-tap FFE together with the second-tap DFE. Specifically,

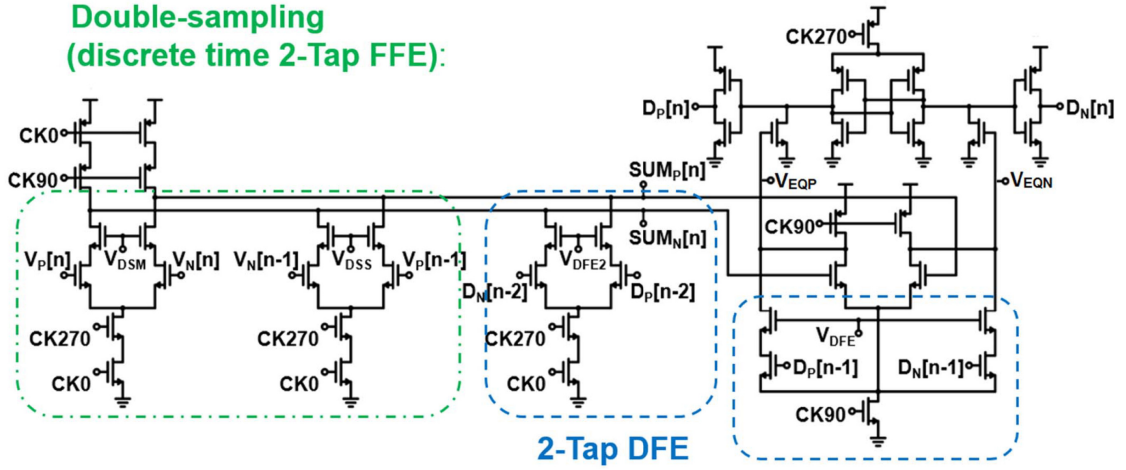


Fig. 3.5. Schematic of the EQ performing double-sampling and two-tap DFE.

$$\begin{aligned}
 \text{SUM}_P[n] - \text{SUM}_N[n] &= \alpha \times (V_P[n] - V_N[n]) \\
 &\quad + \beta \times (V_P[n-1] - V_N[n-1]) \\
 &\quad + \gamma \times (D_P[n-2] - D_N[n-2])
 \end{aligned} \tag{3.6}$$

where  $V_P[n]$  and  $V_N[n]$  are the differential S/H outputs of the current sample;  $V_P[n-1]$  and  $V_N[n-1]$  are the differential S/H outputs of the previous sample spaced with one UI ahead;  $D_P[n-2]$  and  $D_N[n-2]$  are the recovered complementary digital data bits two UIs ahead;  $\alpha$  and  $\beta$  are the FFE coefficients, and  $\gamma$  is the coefficient for the second-tap DFE. The FFE and second-tap DFE coefficients,  $\alpha$ ,  $\beta$ , and  $\gamma$ , are adjusted by varying the gate voltages of the cascoding transistors  $V_{DSM}$ ,  $V_{DSS}$ , and  $V_{DFE2}$ , respectively, in Fig. 3.5, as in [38]. Similarly,  $D_P[n-1]$  and  $D_N[n-1]$  are the recovered complementary digital data bits one UI ahead, and the first-tap DFE coefficient,  $\delta$ , is adjustable by varying  $V_{DFE}$ . The gate voltages are set by voltage digital-to-analog converters (VDACs), and the resultant tap weight ranges of the FFE and DFE (i.e.,  $\beta/\alpha$ ,  $\gamma/\alpha$ , and  $\delta/\alpha$ ) can be set from 0 to 0.8, with 0.025 resolution. The nonlinearity of the integrating summer increases with the differential input signal level. From simulations, the error is increased to  $\sim 10\%$  of the ideal sum, when the differential input levels (i.e.,  $V_P[n] - V_N[n]$  and  $V_P[n-1] - V_N[n-1]$ ) are increased to 330 mV. When the input levels are further increased to 400, 450, and 500 mV, the error is increased to 15.5%, 19.2%, and 23.4%, respectively. The limited accuracy of the integrating summer does have the negative effects on implementing precise equalization;

however, the employed EQ design allows the SNR target shown as (3.4) to be fulfilled within the target dynamic range. As the first-tap DFE is embedded in the two-stage regenerative slicer, the cancellation of the first post-cursor ISI is carried out at the internal nodes of the slicer,  $V_{EQP}$  and  $V_{EQN}$ , labeled in Fig. 3.5. In this design, the direct feedbacks used in [31] are employed. The settled outputs of one regenerative latch are directly fed as inputs to two other EQs for two-tap DFE operation, and the loop-unrolling DFEs are not required by exploiting the overlaps of the evaluation phases of the two adjacent slicers.

### 3.5 Burst-Mode Reconfiguration Loops

The block diagram of the burst-mode reconfiguration loops is shown in Fig. 3.6. During the preamble phase, the reconfiguration is started with an external pulse signal (PUL\_IN) and is finished in 14 reconfiguration clock (RCK) cycles. The on-chip search algorithm applies successive approximation register (SAR) logic, with each clock cycle dedicated to the sequential decision of 1 bit of digital setting, and one additional cycle inserted between those devoted to b7 and b8. The inserted cycle allows reliable dc offset cancellation before the search for the gain setting since the gm-C LPF is enabled to cancel the residual offset at the completion of setting b7. With the enable/disable control scheme shown in Fig. 3.3(f), the capacitors in effect memorize nothing related to the results of the VCS loop as  $V_{NLPF} \approx V_{PLPF}$  throughout the time, when the LPF is disabled. Accordingly, as soon as the LPF is enabled, it starts to help with cancelling the residual offset. Similar to other applications of SAR algorithm, e.g., SAR analog-to-digital converter (ADC), the SAR algorithm applied in this paper relies on comparators to resolve 1 bit of digital setting, and the maximum speed at which the SAR algorithm can run depends on the delay within the loop. Therefore, integrating dc comparator and integrating amplitude comparator are proposed to reduce the minimum comparison time to two UIs, such that the loop delay is no longer limited by the  $RC$  settling time of conventional  $RC$  LPF-based designs. When the preamble data stream is present, the integrating dc comparator compares the dc levels of the AFE outputs, whereas the integrating comparator compares the signal amplitude with reference amplitude. The results are amplified to a digital logic level by the slicers following the comparators, and the VCS or VGA is accordingly adjusted, depending on

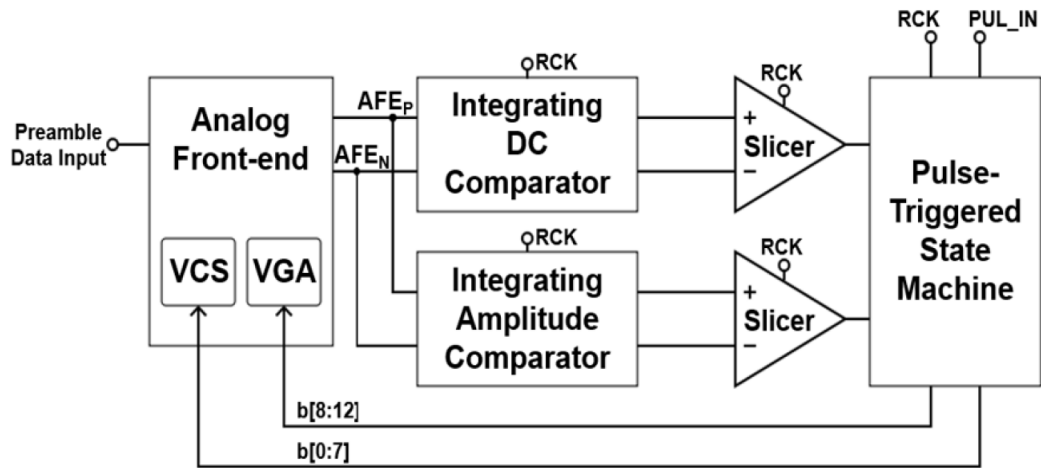


Fig. 3.6. Block diagram of the burst-mode reconfiguration loops.

which reconfiguration loop is on duty. The slicer follows the topology of the double-tail latch-type voltage sense amplifier, proposed in [36], and the slicers in the reconfiguration loops are designed with the specifications as follows. The input-referred noise is  $0.33 \text{ mV}_{\text{rms}}$ ; the sensitivity at 6.25-GHz operation is better than  $100 \mu\text{V}$  for input common-mode voltages varying from 0.4 to 0.7 V; and the offset is 5 mV from Monte Carlo simulations and can be effectively calibrated by introducing the offset into the preceding integrating dc or amplitude comparator. Sections 3.5.1–3.5.3 first describe a customized state machine as part of the SAR search algorithm and elaborate the functions and advantages of the proposed integrating dc and amplitude comparators which have critical contributions to improve the reconfiguration loop delays and hence the link bandwidth as well as latency in burst-mode operations.

### 3.5.1 Pulse-Triggered State Machine

The pulse-triggered state machine is designed for high-speed operation with the goal that each bit of the digital setting ( $b_0:b_{12}$ ) does not react to the slicers in the reconfiguration loops until the corresponding pulse arrives. Additional function with enable/disable logic is implemented, offering options to use either the predefined setting set by an external field-programmable gate array or the setting determined by the reconfiguration loops. Fig. 3.7 shows the block diagram of the pulsed-triggered state machine. Setting the enable signal



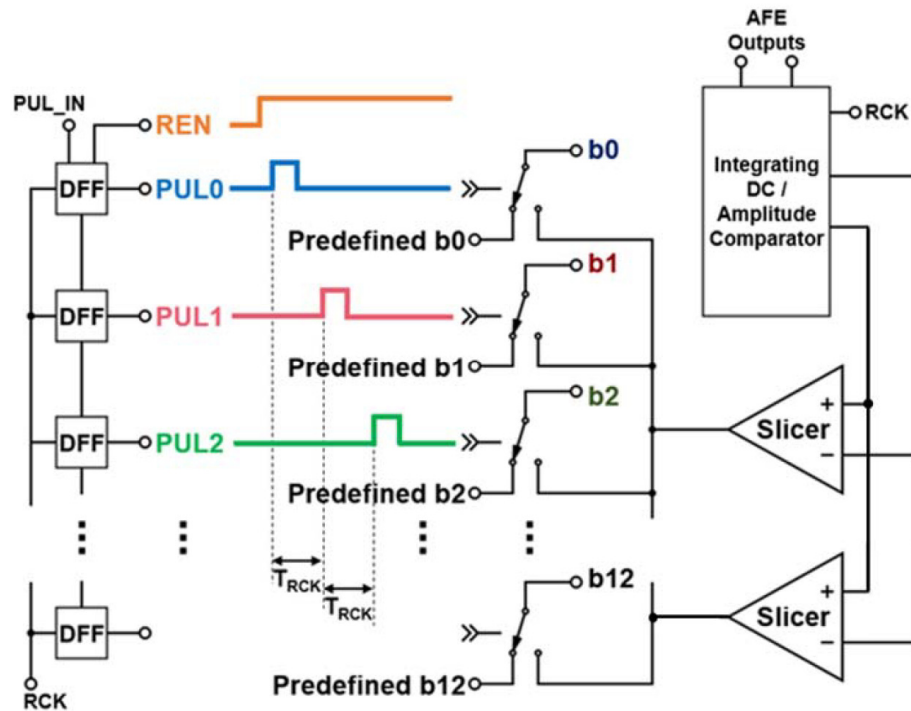


Fig. 3.7. Block diagram of the pulse-triggered state machine.

(REN) to logical low disables the reconfiguration loops, and the predefined digital setting will be used throughout. Setting REN to logical high enables the reconfiguration loops, and a chain of nonoverlapping pulses spaced with one RCK cycle ( $T_{RCK}$ ) is generated, selecting the bit to be overwritten by the slicer, one after another. In other words, as REN is set to be logical high, the predefined values of the digital setting are to be sequentially overwritten. For instance, with REN set to high, b0 keeps its predefined value when its corresponding digital control signal, PUL0, is initially low. When PUL0 rises to high, the register of b0 starts to take in the slicer output. Before PUL0 goes back to low, the regenerative slicer settles and overwrites the original predefined value of b0. This value written by the slicer is held afterward, unless the predefined value is reloaded by setting REN to low. By design, the rising edges of the pulses are aligned with those of the RCK, and the misalignment induced by process variations can be compensated with an on-chip digitally controlled delay line.

### 3.5.2 Integrating DC Comparator

Conventional first-order  $RC$  LPFs are commonly applied to extract dc information. As shown in Fig. 3.8(a), the slicer directly compares the LPF voltage levels and amplifies the difference to digital logic level. The result is then taken as 1 bit of the digital setting for VCS during the reconfiguration process. Nonetheless, as illustrated in Fig. 3.8(b), there is an inevitable tradeoff between the tracking time and settling behavior. With the  $RC$  time constant set to be 0.1 ns, as shown in blue, it can be observed that considerable ripples, which make the comparison result less reliable, are introduced. In contrast, with the  $RC$  time constant set to be 1 ns, as shown in red, it fails to track the dc component in 1.5 ns. This  $RC$  settling time constraint presents a bottleneck in speeding up the SAR logic, and thus the burst-mode reconfiguration since the unsettled voltage levels do not accurately reflect the effect of the last adjustment of VCS. As a consequence, comparing the unsettled voltage levels can lead to the nonoptimal setting of the VCS at the end of the reconfiguration process. The integrating dc comparator, as shown in Fig. 3.9(a), is proposed to replace the  $RC$  LPF. The pMOS pair charges the outputs to the supply voltage when the RCK is low, resetting the differential output voltage approximately zero. When RCK becomes high, the integration of the respective input voltage is effectively performed as the summation of the discharging current on the load capacitance ( $C_{LOAD}$ ), i.e., the voltage drop at the output. Since the input waveform is programmed to have “1010. . .” preamble pattern, the voltage drop at the output contains the information of the input dc level with the integration period set to even numbers of UIs. The simulation result, as shown in Fig. 3.9(b), illustrates the principle of operation. With the integration period (half of the RCK period in this design) set to two UIs and proper common-mode design, the polarity of the differential output voltage indicates which input has higher dc level at the end of the integration period. In addition, it is insensitive to the timing alignment between the RCK and the preamble data stream due to the nature of performing integration, and therefore, the locking of CDR in advance is unnecessary. The slicer following the integrating dc comparator further amplifies the differential output voltage to digital levels, overwriting 1 bit of digital setting to adjust the current of VCS. The proposed integrating

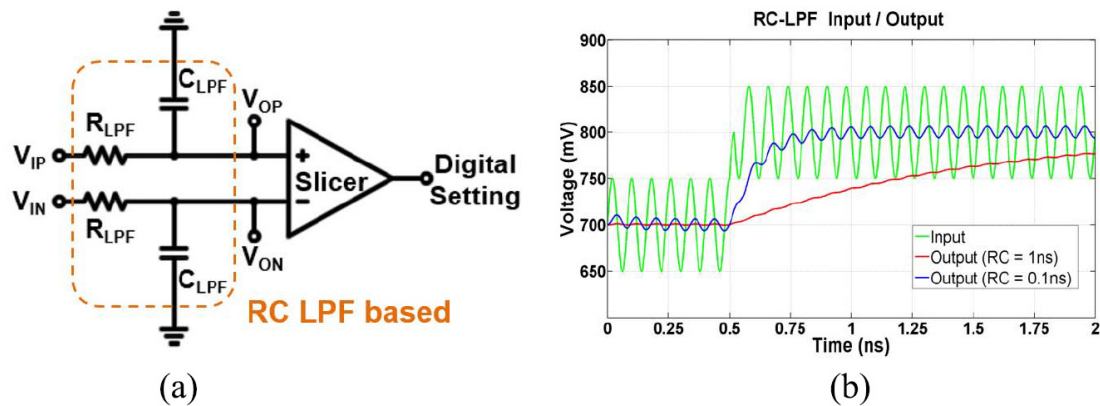
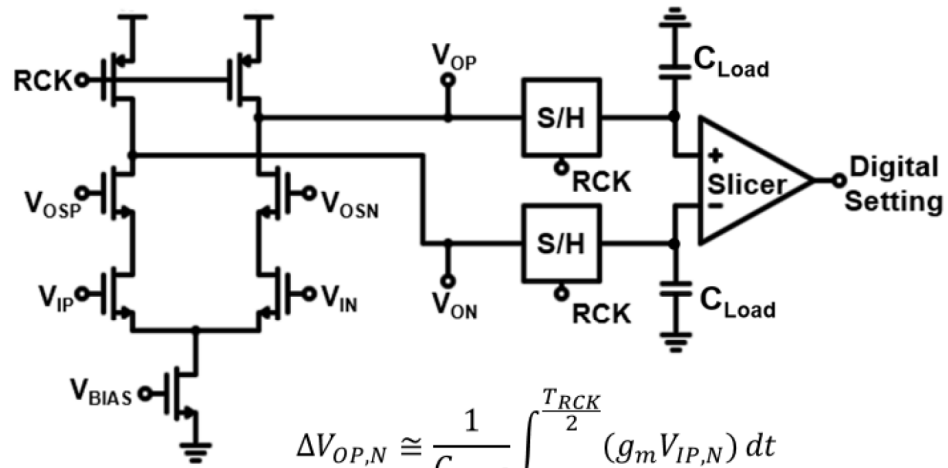


Fig. 3.8. (a) Conventional  $RC$  LPF-based dc comparator. (b) Simulation results showing the tradeoff between tracking time and settling behavior.

dc comparator eliminates the  $RC$  settling time constraint and the minimum integration time; namely, the minimum comparison time can be set to be two UIs by integrating only one pair of 1 and 0. To make the fast dc offset cancellation loop more precise, the offset from the integrating dc comparator itself can be calibrated by adjusting the gate voltages of the cascoding transistors ( $V_{OSP}$  and  $V_{OSN}$ ) with VDACS. As other current-integrating designs, the common-mode integration could cause problems, if the common-mode voltage drops at the outputs are undesirably large such that the transconductance ( $g_m$ ) of the input pairs becomes significantly smaller as the integration carries out. To avoid the aforementioned issue, the common-mode output voltages are designed in order that 150 mV is left as the margin for the input pairs from being out of the saturation region. In addition, the tail bias current can be varied by adjusting its gate voltage  $V_{BIAS}$ . Finally, the effects of non-50% duty cycle clocks on the integration results are simulated, as shown in Fig. 3.9(c), suggesting that  $\pm 10\%$  of duty cycle distortion does not have a significant impact on the calibration accuracy on account of the invariant polarity or sign of the integration results.

### 3.5.3 Integrating Amplitude Comparator

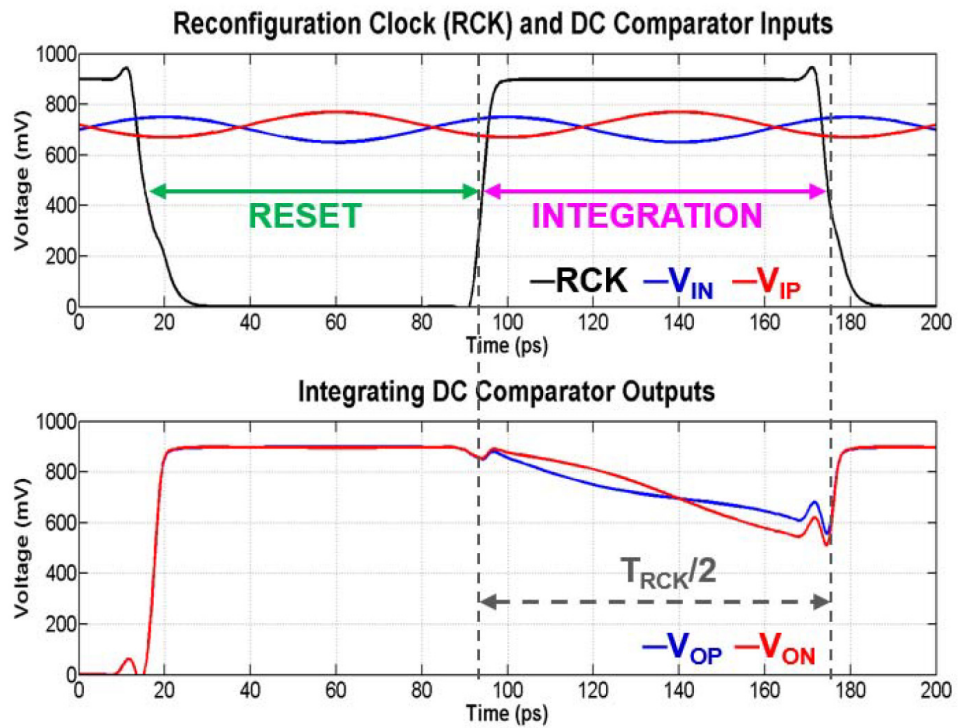
An automatic gain control (AGC) loop needs the information of signal amplitude in order to adjust the gain along the signal path. This purpose is conventionally implemented by using  $RC$  LPF-based peak detectors, e.g., [22], to measure the value or the level-shifted



$$\Delta V_{OP,N} \cong \frac{1}{C_{Load}} \int_0^{\frac{T_{RCK}}{2}} (g_m V_{IP,N}) dt$$

$\propto V_{IP,N(DC)}$ , if integrating over even number pairs of 1 and 0

(a)



(b)

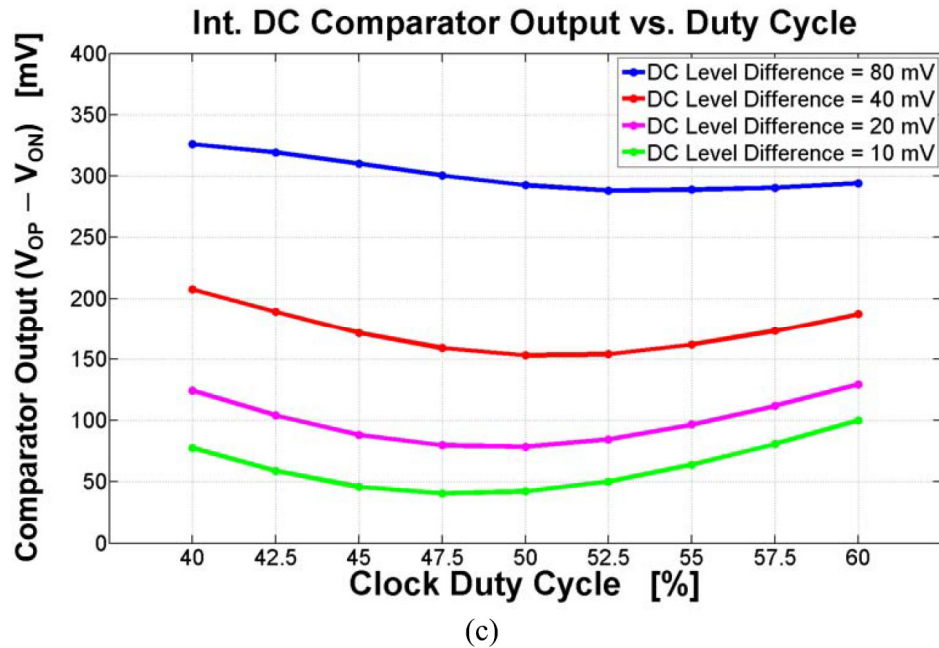
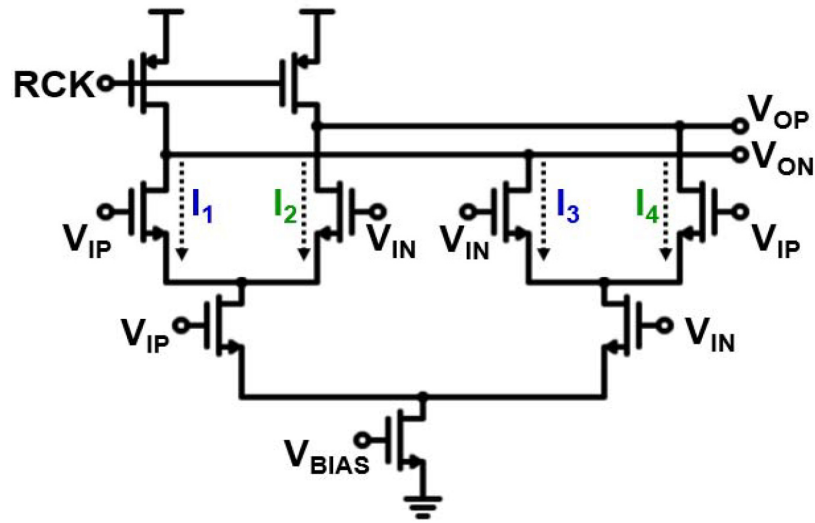
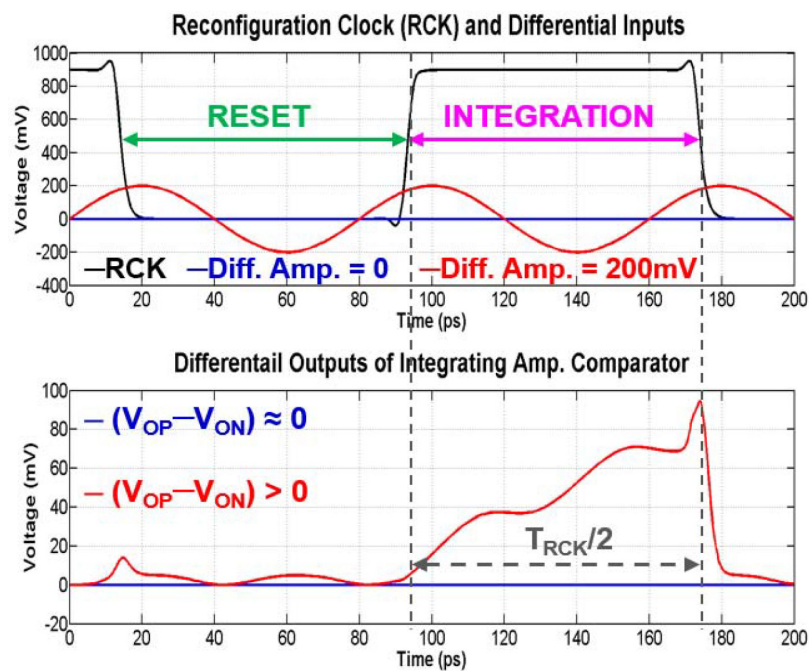


Fig. 3.9. (a) Circuit schematic of the proposed integrating dc comparator. (b) Simulation results showing the operation of the proposed integrating dc comparator, where the dc level of  $V_{IN}$  is lower than that of  $V_{IP}$  by 20 mV. (c) Integrating dc comparator differential output voltage versus different clock duty cycles with four distinct dc-level differences.

value of the peak amplitude. Similar to the first-order  $RC$  LPF, as described previously, the inevitable tradeoff between tracking time and settling behavior limits the reconfiguration speed, as the next adjustment of the gain setting may not be correctly resolved if the peak detectors are not settled. In this paper, the integrating amplitude comparator is proposed to replace the conventional peak detectors in the AGC loop and to enable rapid signal amplitude control along with the SAR search algorithm. The circuit schematic of the building block in the proposed integrating comparator is shown in Fig. 3.10(a), while its principle of operation is illustrated in Fig. 3.10(b). With the same RCK used in the integrating dc comparator, the outputs are pre-charged to the supply voltage when RCK is low, and the differential output voltage is thus reset to approximately zero prior to the rise of RCK. During the integration phase, i.e., when RCK is high,  $V_{OP}$  and  $V_{ON}$  are both being discharged, with a potentially equal or very different amount, depending on the differential input amplitude. In Fig. 3.10(b), provided that the mismatches introduced by the process



(a)



(b)

Fig. 3.10. (a) Building block of the proposed integrating amplitude comparator. (b) Simulation results showing the operation of the building block in the proposed integrating amplitude comparator.

variations are negligible or calibrated,  $I_1 = I_2 = I_3 = I_4$  when  $V_{IP} = V_{IN}$  by symmetry, and consequently, as shown in blue, the zero differential input amplitude leads to zero

differential output voltage ( $V_{OP} - V_{ON} \approx 0$ ), at the end of integration. By contrast, in the case that the differential input amplitude is large, as shown in red in Fig. 3.10(b),  $I_1$  conducts most of the tail bias current during the half preamble period (one UI) when  $V_{IP} > V_{IN}$ , while  $I_3$  conducts most of the tail bias current during the other half preamble period when  $V_{IP} < V_{IN}$ . Since both  $I_1$  and  $I_3$  discharge the same node  $V_{ON}$ , a relatively large differential output voltage ( $V_{OP} - V_{ON}$ ), after the integration over one full preamble period (two UIs), is expected, due to the significantly more voltage drop at  $V_{ON}$ . The biasing and the sizes of the differential pairs are further optimized in order that the value of ( $V_{OP} - V_{ON}$ ) at the end of the integration phase increases with the differential amplitude of the input, regardless of the timing alignment between the input preamble waveform and the RCK. As shown in Fig. 3.11(a), a replica stage is connected to the outputs with opposite polarity, converting the differential amplitude of its input into the value of ( $V_{ON} - V_{OP}$ ) instead, at the end of the integration phase. Therefore, one stage will compete with the other during the integration phase in deciding the sign of ( $V_{OP} - V_{ON}$ ), and the result directly indicates which stage sees the input signal with larger differential amplitude, given that the common-mode voltages of the inputs are identical, and the offsets are negligible. A reference preamble waveform possessing “1010. . .” pattern is derived from the rail-to-rail clock signals and its amplitude is programmable but fixed during the reconfiguration process. By comparing the preamble waveform from the AFE outputs with the reference preamble waveform, the proposed integrating amplitude comparator removes the need for peak detectors, offering much faster updates to the VGA gain setting. The rapid signal amplitude control is achieved by incorporating the proposed integrating amplitude comparator with the SAR search algorithm such that the amplitude of the AFE outputs converges toward the reference amplitude in a designed number of RCK cycles. In this paper, differential pair-based buffers are included at the inputs of the proposed integrating amplitude comparator to implement common-mode rejection, with the main benefit that the proposed gain reconfiguration loop is insensitive to the residual dc offset. Finally, similar to the case of integrating dc comparator, the effects of non-50% duty cycle clocks on the integration results are also simulated, as shown in Fig. 3.11(b), again suggesting that  $\pm 10\%$  of duty cycle distortion does not have a significant impact on the calibration accuracy on account



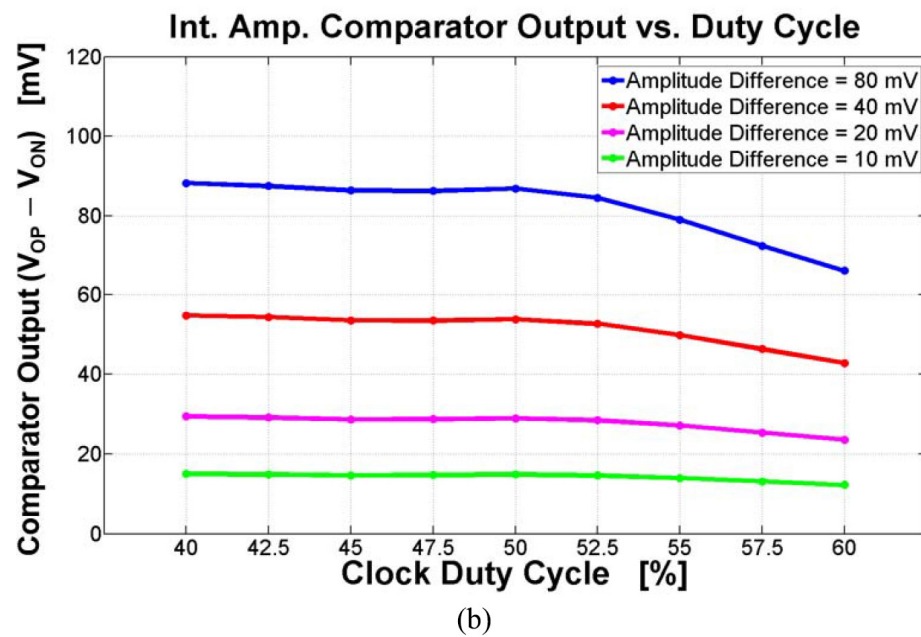
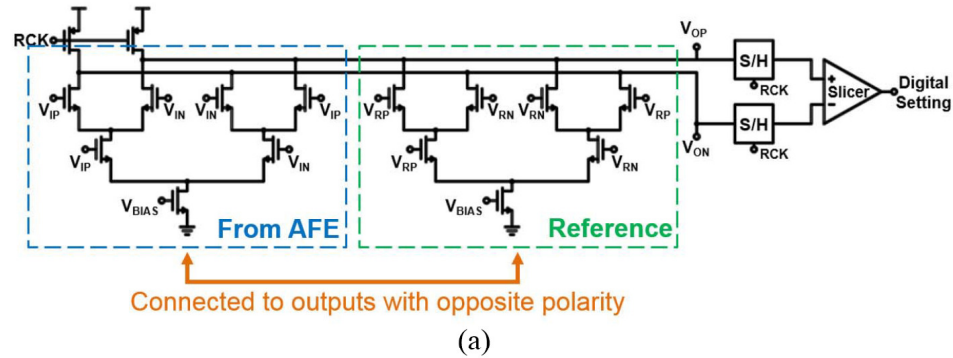


Fig. 3.11. (a) Circuit schematics of the proposed integrating amplitude comparator. (b) Integrating amplitude comparator differential output voltage versus different clock duty cycles with four distinct amplitude differences.

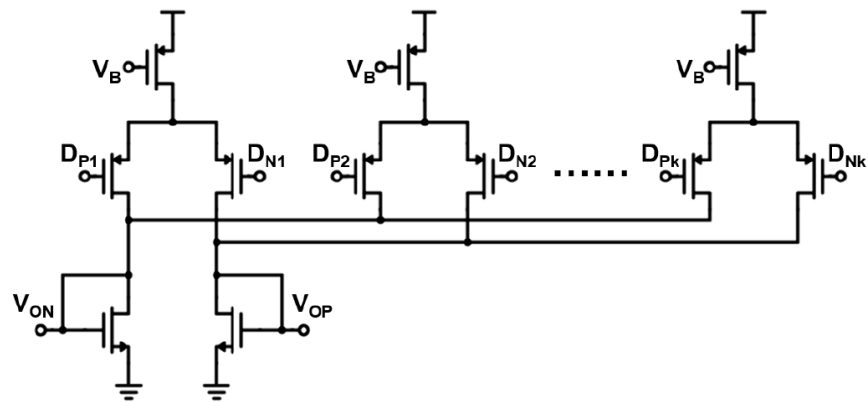
of the invariant polarity or sign of the integration results.

### 3.5.4 Analog Settling Time Reduction

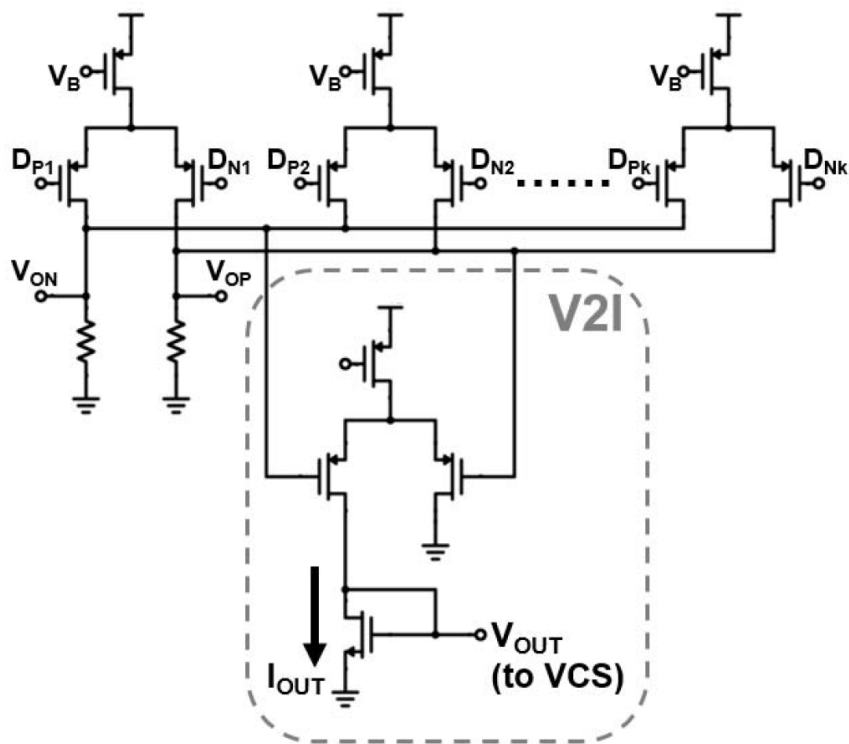
Despite the bottleneck formed by the  $RC$  settling time constraint in speeding up the reconfiguration loop is eliminated by the proposed integrating dc and amplitude comparators, the analog settling time still takes part in determining the maximum speed at which the SAR logic can operate. The analog settling time is destined, as the effects of



updating the digital setting of VCS or VGA cannot be immediately settled and ready for the next point in the SAR search process. Even though the analog settling occurs concurrently, it is informative to identify the analog settling time as two parts. The first one resides in the AFE and strongly depends on the bandwidth of the AFE. One possible way to reduce the analog settling time of the AFE, which is not implemented in this paper, is adding switches to decrease the load resistance at each or selected stages during the reconfiguration process. For instance, the resistance of the shunt-feedback resistor in TIA or the load resistors in VGA can be effectively reduced by turning on the switches in parallel, when the reconfiguration is in progress. The drawback of the aforementioned method is that the increase in bandwidth by decreasing the load resistance generally implies the reduction in gain, and hence, the dc offset and signal amplitude are both expected to be smaller, compared with those in the case, when the parallel switches are absent. The other part of analog settling time is associated with the settling of the bias currents in VCS and in the current-steering VGA. Fig. 3.12(a) shows the schematic of a conventional current-mirror-based current digital-to-analog converter (DAC), where the digital inputs steer the currents into or out from the current mirror at the output. This topology using a current mirror is suitable for high-speed operation, i.e., with short settling time, only if the current mirror conducts a relatively high current such that the diode-connected transistor acts as a resistor with relatively low resistance. Accordingly, the DAC with a current mirror can be used in reconfiguring the current-steering VGA in that the currents flowing through the current mirrors are expected to be within 1–3 mA. By contrast, the DAC using a current mirror should not be directly used in reconfiguring the VCS, in view of the fact that the target dc component of the photocurrent, which is to be subtracted with the current flowing through the VCS, is on the order of 100  $\mu\text{A}$ . The schematic of the proposed solution is shown in Fig. 3.12(b), where the DAC is loaded with resistors with low resistance, and the differential output voltage is then taken as the differential input of a voltage-to-current (V2I) converter. The resistively loaded DAC has a lower and invariable  $RC$  time constant, in contrast to the DAC loaded with current mirrors, benefiting the settling time whenever a new digital setting is applied. A mirroring ratio of 7:2 is used at the output current mirror of the V2I to further avoid relatively small bias current flowing into the node  $V_{\text{OUT}}$  labeled



(a)



(b)

Fig. 3.12. (a) Circuit schematic of the conventional DAC loaded with diode-connected transistors for current mirroring. (b) Circuit schematic of the DAC loaded with low-resistance resistors. The differential output voltage is taken as the differential input of a voltage-to-current (V2I) converter.

in Fig. 3.12(b). The 95% settling time at  $V_{OUT}$  is measured to be 36.64 ps from the simulation. The V2I converter not only provides isolation of the output node from the bank

of switches used to steer the currents but also allows the VCS to operate in different dynamic ranges by simply varying its tail bias current. Although nonlinearity can be introduced by the V2I converter, it is not an issue in this design where the resolution for the target dynamic range is sufficient and the convergence to the level closest to the ideal one is accomplished by the feedback loop of SAR search algorithm.

### 3.5.5 Simulation Results

With the foregoing designs and optimizations, the RCK period for 25-Gb/s operation can be set to be four UIs, in which two UIs are dedicated to the integration phase, i.e., comparison time, while the other two UIs are devoted to resetting the integrating comparators and the settling time after an update to the digital setting is applied. A typical simulation result is shown in Fig. 3.13 for illustration. The outputs of the AFE are initially far away from each other because of the large dc offset. As the dc offset gets cancelled, they become closer to each other. Afterward, the amplitude starts to grow and remain at a desirable level. The whole burst-mode reconfiguration process takes a fixed number of RCK cycles, 14 cycles in this design, and thus finishes in  $14 \times 4$  UIs. Specifically, with the pulse-triggered state machine described in Section 3.5.1, the digital settings can only be sequentially overwritten or reconfigured within 14 clock cycles. After the 14 clock cycles, all settings cannot be further changed since their digital values are held and stored by latches until the next reconfiguration process. When a quarter-rate (6.25 GHz) clock is used for 25-Gb/s operations, the reconfiguration takes place within the time span of  $14 \times 160$  ps = 2.24 ns. Similarly, if a 3.125-GHz (1/8 of the data rate) clock is used for 25-Gb/s operations, the reconfiguration is completed in  $14 \times 320$  ps = 4.48 ns.

## 3.6 Experimental Results

The chip is fabricated in 28-nm CMOS technology. Fig. 3.18(a) shows the die micrograph of the core circuitry. The experiment setup is shown in Fig. 3.14. The receiver chip is wire-bonded to an APD die whose gain ( $M$ ) at 1310 nm is adjustable via the reverse-bias voltage. A continuous-wave laser is modulated by a high-speed Mach–Zehnder modulator with PRBS-7 data pattern and coupled to the APD through a single-mode fiber. An oscilloscope

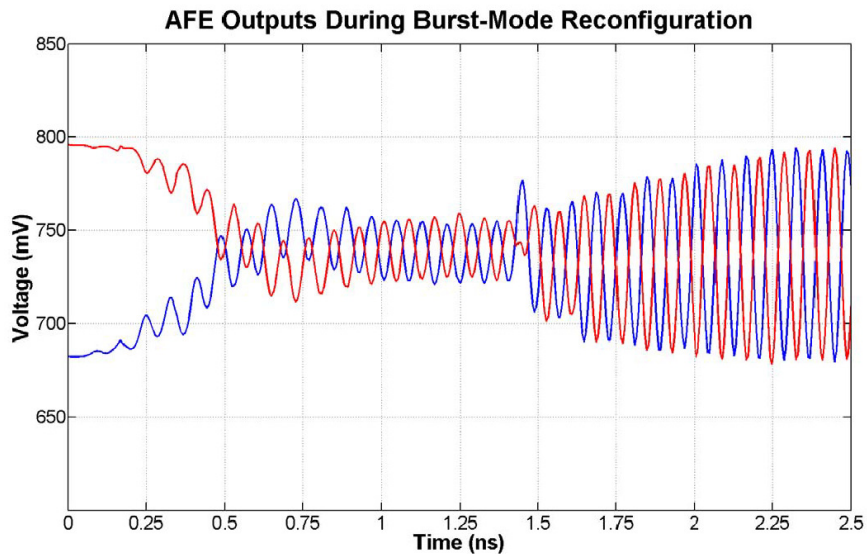


Fig. 3.13. Simulated AFE outputs in burst-mode reconfiguration.

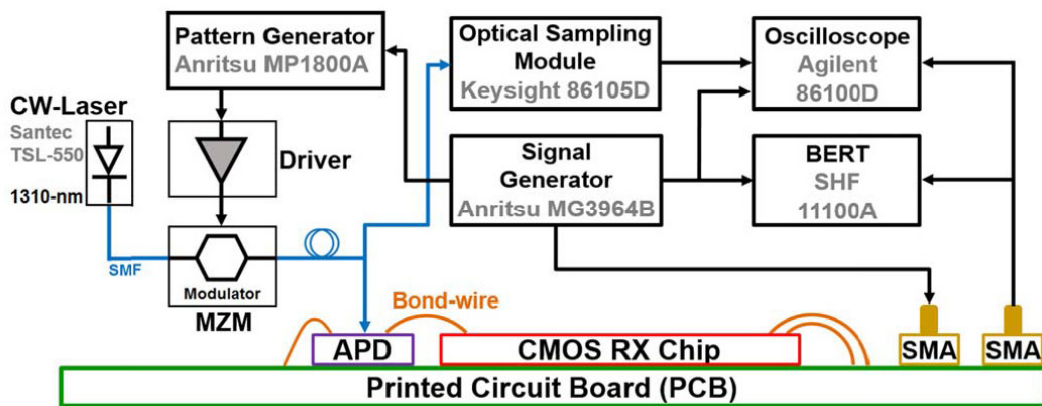


Fig. 3.14. Block diagram of the experiment setup.

for monitoring the input optical data signal and the output electrical data signal is set up, and an external BER tester is used to measure the BER. The best sensitivity,  $-16$ -dBm (OMA), is achieved at 25 Gb/s with PRBS-7 input pattern when the reverse-bias voltage of APD is set to be 16 V, under which the overall responsivity, including the multiplication factor, of the APD is 4 A/W, while the  $-3$ -dB bandwidth of the APD optical response, excluding the input resistance and capacitance of the electronic chip, is  $\sim 20$  GHz. The off-chip decoupling capacitors are included on the printed circuit board (PCB) to minimize the

variation of the APD bias voltage. The measured bathtub curve with  $-16\text{-dBm}$  (OMA) at  $25\text{ Gb/s}$  with PRBS-7 input pattern is shown in Fig. 3.15, showing  $0.2\text{ UI}$  horizontal opening for BER less than  $10^{-12}$ . In order to verify the function of the proposed integrating dc and amplitude comparators together with the reconfiguration loops, the waterfall plot for PRBS-7 input with fixed EQ setting found with  $-16\text{-dBm}$  input is shown in Fig. 3.16, and a dynamic range of  $5\text{ dB}$  is achieved. Outside the dynamic range, the BER is improved when reducing the RCK frequency from quarter rate to one-eighth of the data rate. Similar to the SAR ADC designs, a single decision error during the SAR search process can lead to deviation from the optimum convergence point. The extra time granted by reducing the RCK frequency primarily helped the AFE to settle more completely, and by which the chance of having a decision error on account of the unsettled inputs is reduced. The limiting factor of the dynamic range in this paper lies in the current-steering VGA since each stage of the VGA is designed to have only  $2.45\text{-dB}$  dynamic range of gain with its  $-3\text{-dB}$  bandwidth kept approximately constant. When tested with PRBS-31 input pattern, the best sensitivity measured at  $25\text{ Gb/s}$  is degraded to  $-15.3\text{-dBm}$  (OMA), and the waterfall plot for PRBS-31 with fixed EQ setting found with  $-15.3\text{-dBm}$  input is shown in Fig. 3.17. Finally, the power consumption and the breakdown at  $25\text{ Gb/s}$  are shown in Fig. 3.18(b). The AFE consumes  $12.2\text{ mW}$ , including  $1.2\text{ mW}$  by APD; the EQ consumes  $4.3\text{ mW}$ , and

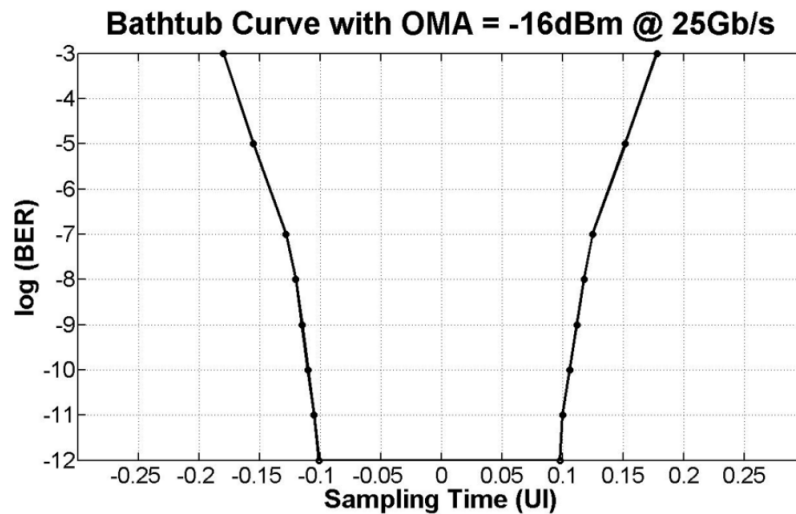


Fig. 3.15. Bathtub curve measured with  $-16\text{-dBm}$  OMA at  $25\text{ Gb/s}$ .

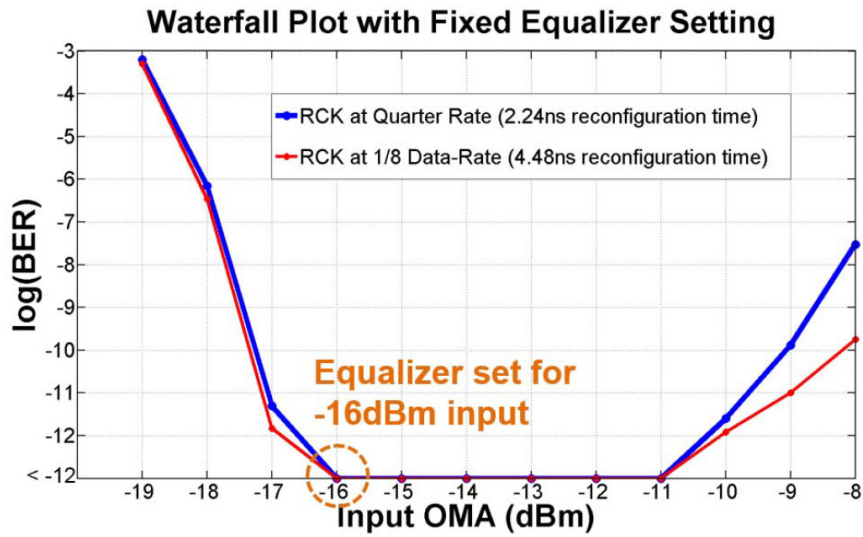


Fig. 3.16. Waterfall plot with fixed EQ setting at 25 Gb/s, with PRBS-7.

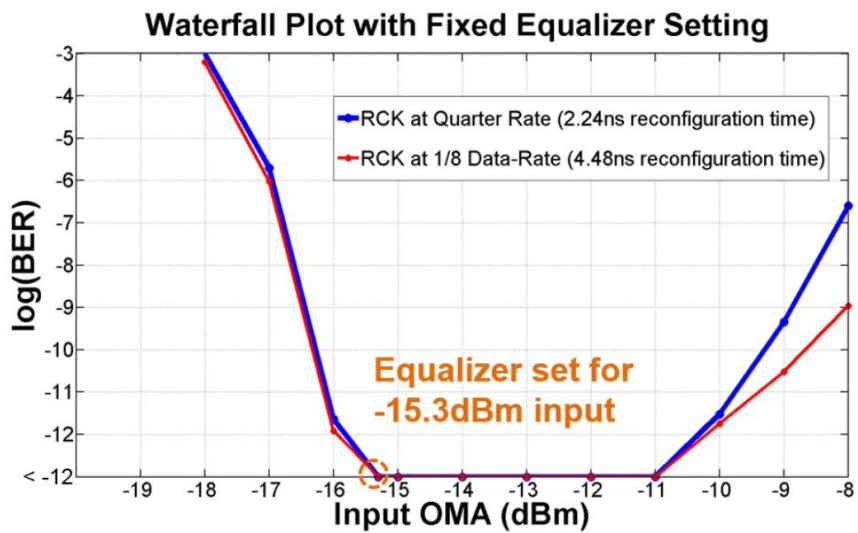


Fig. 3.17. Waterfall plot with fixed EQ setting at 25 Gb/s, with PRBS-31.

the clock and data buffer consume 17.7 mW. In total, 34.2 mW is consumed by the receiver data-path, and 1.37-pJ/b energy efficiency is achieved.

### 3.7 Summary

The APD-based burst-mode optical receiver applies current-integrating equalization and

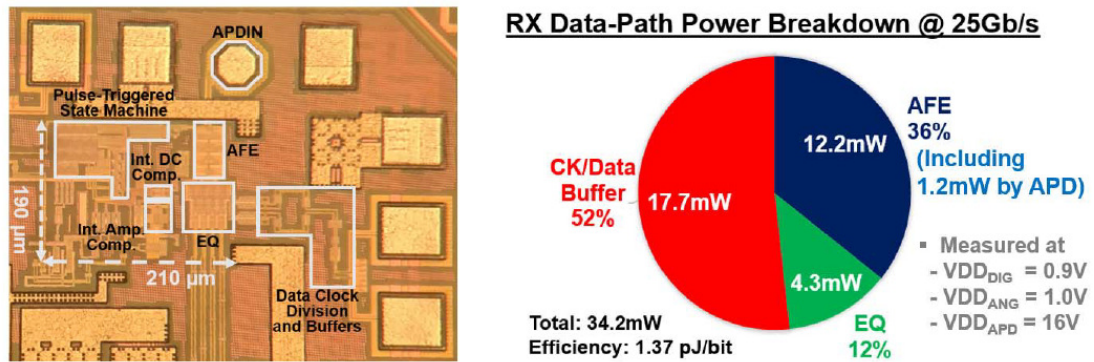


Fig. 3.18. (a) Micrograph of the core circuitry, including the pad wire-bonded to the APD (APDIN), AFE, quarter-rate EQ (EQ), integrating dc comparator (Int. dc Comp.), integrating amplitude comparator (Int. Amp. Comp.). (b) Power consumption breakdown of the receiver data path at 25 Gb/s.

	This work	JSSC' 2015 [19]	RFIC' 2014 [32]	ISSCC' 2017 [33]	VLSI' 2017 [34]	
Technology	28nm	32nm SOI	28nm	14nm FinFET	14nm FinFET	
Data Rate (Gb/sec)	25	25	25	32-64	25	32
Efficiency (pJ/bit)	1.37	4*	0.17	1.4 @64G	1.59	1.41
PD Capacitance (fF)	55	100	8	69	69	
PD Responsivity (A/W)	4	0.5	0.8	0.52	0.52	
Reconfiguration Time (ns)	2.24	12.5**	N/A	N/A	N/A	
Sensitivity (dBm)	-16	-10.9	-12.8***	-13 @32G	-13.8	-12.4

\*Including clock and data recovery circuitry. \*\*The 12.5 ns is fully dedicated to cancelling the DC component. \*\*\*Calculated with 6-dB optical coupling loss.

Table 3.1. Performance summary and comparisons of optical receivers.

achieves  $-16$ -dBm (OMA) sensitivity at 25 Gb/s with 1.37-pJ/b energy efficiency. The proposed integrating dc comparator and integrating amplitude comparator significantly relax the settling time constraints, enabling 2.24-ns reconfiguration time at 25 Gb/s. The performance and comparisons with the state-of-the-art are summarized in Table 3.1.

## PAM4 WIRELINE RECEIVER WITH 2-TAP DIRECT DECISION FEEDBACK EQUALIZATION (DFE)

### 4.1 Overview

Four-level pulse amplitude modulation (PAM4) signaling has become an attractive option for high-speed data communication links where the channels suffer from severe bandwidth limitation, by virtue of its halved Nyquist frequency in comparison with that of non-return-to-zero (NRZ) modulation. In other words, the PAM4 signaling improves the spectral efficiency over that of NRZ, by encoding two bits of information, often referred to as the most significant bit (MSB) and the least significant bit (LSB), into one symbol. The consequent advantages of using PAM4 as a substitution for NRZ include the following: the bandwidth requirements for the channel and the front-end circuits are both reduced, and the circuits for clock generation and distribution can operate at the halved frequency. These advantages can potentially lead to higher data rates and/or lower power consumptions. However, there are new challenges stemming from the nature of multilevel signaling when designing PAM4 transceivers. Specifically, with a fixed transmitter swing that is divided into the multiple levels, the receiver needs to resolve the transmitted bits from signals that have lower strength. The foregoing infers to two important design challenges, which this work focuses to address. One is the more demanding sensitivity of the decision circuitry, as will be elaborated in later paragraphs. The other is the necessity of canceling the inter-symbol interference (ISI), since the ISI resulting from strong symbols, for example, (MSB, LSB) = (+1, +1), can intrude detrimental interference on the nearby weak symbols, for example, (MSB, LSB) = (-1, +1), and cause undesirable data eye closure as a result. The same proportion of ISI level can be, on the contrary, tolerable in the cases of NRZ modulation in that the bipolar symbols, or bits, have nominally identical magnitude of signal swings.

Depending on the architecture of the receiver, analog-based equalization and/or analog-to-



digital converter (ADC)-based equalization can be employed. In both scenarios, the incorporation of a decision feedback equalizer (DFE) is often an appealing option, as a DFE can succeed in compensating post-cursor ISI without amplifying crosstalk and noise. Recent examples include an ADC-based PAM4 receiver [41], designed in 16-nm FinFET CMOS, utilizing analog CTLE together with 24-tap feedforward equalizer (FFE) and 1-tap DFE implemented in digital domain, and the transceiver achieves bit-error-rate (BER) less than  $1\text{E}-8$  at 56 Gb/s over a channel with 31-dB loss at 14 GHz (Nyquist frequency). With the feasibility of integrating hybrid analog and digital equalization including long-tap FFE, ADC-based receiver architectures have been designed for longer reach or channels with loss greater than 30 dB at Nyquist [41]–[44]. On the other hand, an analog-based 40–56-Gb/s PAM4 receiver in 16-nm FinFET CMOS [45], targeting chip-to-module and board-to-board cable interconnects, mitigates the channel loss of 10 dB at 14 GHz and reflections, by incorporating CTLE and direct 10-tap DFE in analog domain. Compared to [41] that equalizes  $> 30$ -dB loss at Nyquist with ADC-based architecture, this analog-based receiver [45] designed for 10-dB loss at Nyquist achieves BER of less than  $1\text{E}-12$  at 56 Gb/s but consumes  $\sim 40\%$  less power [45]. These previous designs suggest that for short reach applications where channel losses can be less than 10 dB at Nyquist, an ADC-based receiver may not be the optimal solution in consideration of both the hardware and power that need to be invested.

Despite the usefulness of including a DFE as part of a PAM4 receiver in the analog fashion, as demonstrated in [45]–[47], improving the energy efficiency of an analog-based PAM4-DFE at high data rates remains challenging. First, compared with NRZ receivers, the reduced eye-height in PAM4 receivers (by a factor of  $\sim 3$  in the absence of nonlinearity and with fixed transmitter swing) sets a more stringent limit for the sensitivity of the slicer used for resolving the symbols and making decisions. Furthermore, the sensitivity requirement generally becomes more difficult to meet, given tighter timing constraints, such as at higher data rates and/or with lower decision latency requirements in feedback loops. Second, at least three slicers are required with respect to the three distinct thresholds, and therefore, the power consumed by the slicers and the loading presented by the slicers are of much

greater concern in designing PAM4 receivers. Moreover, as can be seen in Fig. 4.1, which compares the implementation of direct 1-tap PAM4-DFE with that of 1-tap loop-unrolling PAM4-DFE, the loop-unrolling technique demands significantly more hardware. Even if only one tap is unrolled, it needs 12 slicers, three multiplexers, and one thermometer-to-binary (T2B) decoder for each deserialized branch (e.g., 24 slicers, six multiplexers, and two T2B decoders in total for a half-rate design). Since the number of slicers increases exponentially with the number of taps unrolled, the loop-unrolling technique is prohibitively costly in hardware and power consumption for a high data rate PAM4 receiver, suggesting that the speed or the delay performance of the slicers is critical. As illustrated in Fig. 4.2, a stringent timing constraint that requires all the operations to be finished within 1 UI is set, when attempting to directly close the decision feedback loop for the first tap. Although the signal propagation and settling happen concurrently in reality, it is informative and useful to conceptually distinguish them into the setup time of the slicer, clock-to-Q delay of the slicer, the propagation delay of the DFE tap, and the settling time of the summer. Details and interpretations of these timing constraints are presented in Section 4.4. In particular, since the clock-to-Q delay of the slicer takes up a considerable portion in the 1-UI constraint, as will be shown in Section 4.3, the improvement in slicer delay helps to close the loop at higher data rates or to relax the summer design such that no excess power-bandwidth tradeoffs or area-consuming inductors are required for reducing the summer settling time. Therefore, this work aims to demonstrate the idea of implementing an energy-efficient PAM4 receiver with direct DFE loops by improving the slicer performance.

This article, expound upon [60], is organized as follows. Section 4.2 presents the overall PAM4 receiver architecture, where each subsection describes the circuits that serve as key building blocks in the analog front-end (AFE) and in the clock path. Section 4.3 reviews the operations and features of prevalent slicer topologies and describes the proposed slicer in detail. Section 4.4 elaborates the timing constraint for completing the DFE loops. Experimental results of this PAM4 receiver are shown in Section 4.5, and finally, Section 4.6 summarizes this work with performance comparisons and conclusions.

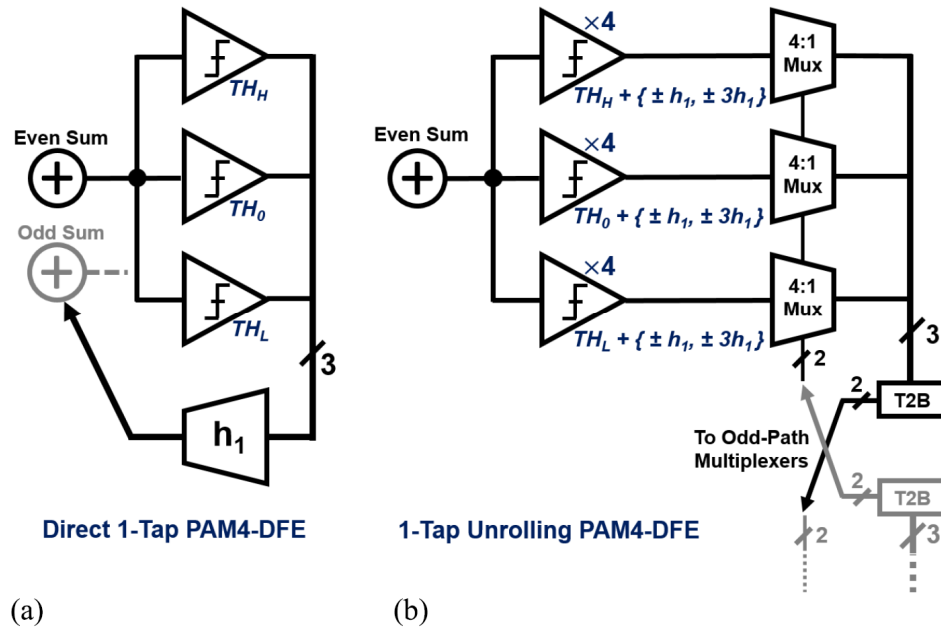


Fig. 4.1. Hardware implementation of PAM4-DFE in half-rate designs. Only the even data-path is shown for clarity, where  $TH_H$ ,  $TH_0$ , and  $TH_L$  are the three distinct threshold levels, and  $h_1$  corresponds to the first post-cursor ISI. (a) Direct 1-tap PAM4-DFE. (b) Loop-unrolling 1-tap PAM4-DFE.

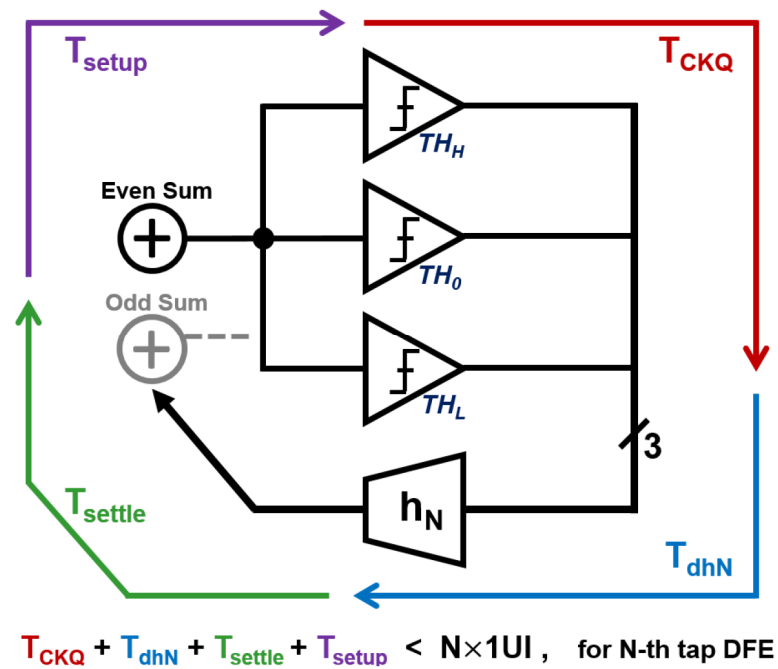


Fig. 4.2. Timing constraint for a direct DFE design for  $N$ -th post-cursor ISI compensation.

## 4.2 Receiver Architecture

### 4.2.1 Overall Architecture

The overall architecture of the PAM4 receiver is shown in Fig. 4.3. The AFE is composed of two stages of continuous time linear equalizer (CTLE) and two half-rate summers. The outputs of each summer are connected to four proposed CMOS track-and-regenerate slicers, among which one is responsible for the eye monitor (EM), and the other three slicers are dedicated to recovering the analog summer outputs to the corresponding 3-bit thermometer-coded digital levels. With the proposed slicers, direct 2-tap DFE is implemented. The 3-bit thermometer-coded outputs are first directly fed back to the summer in the other data path for the first tap of DFE, and then with 1-UI delay, fed back to the summer in the same data path for the second tap of DFE. The digital-level slicer outputs are further demultiplexed (1-to-32) for external and on-chip EM and BER counters (BERCs) to evaluate the eye-opening and BER performance, respectively. The clock path takes in an external pair of half-rate differential clock signals and amplifies them to rail-to-rail levels with on-chip duty cycle correction (DCC). Clock buffers (CKBUFs) and a digitally adjustable delay line (DL) are included on the chip, serving as the interfaces with the clocked slicers to provide rail-to-rail clock signals for data recovery as well as the required clock phases for eye monitoring.

The following sections describe the details associated with the design of CTLE in Section 4.2.2, half rate summers in Section 4.2.3, linearity characterizations in Section 4.2.4, current mode logic (CML)-to-CMOS clock converter in Section 4.2.5, and DCC circuits in Section 4.2.6. The details of the proposed slicer are presented in Section 4.3.

### 4.2.2 CTLE

CTLE is included in the receiver to mitigate both pre-cursor ISI and post-cursor ISI, as the coverage of the direct 2-tap DFE design is limited to the first and second post-cursor ISI. Fig. 4.4(a) shows the schematic of the CTLE, which adopts the conventional topology of *RC* source-degenerated differential amplifier with digital controllability. The high-

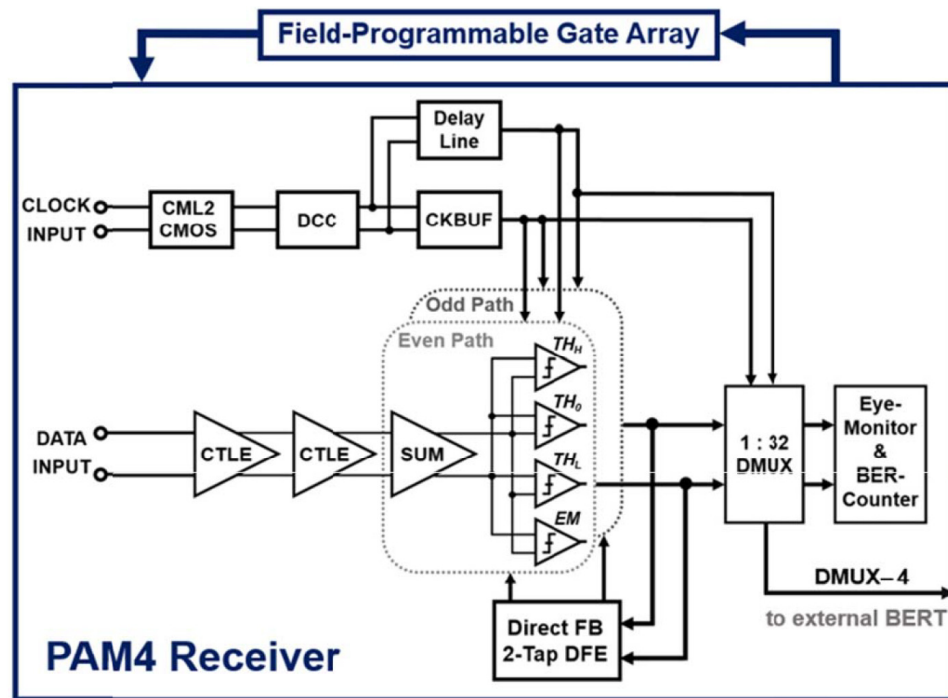


Fig. 4.3. Overall architecture of the PAM4 receiver.

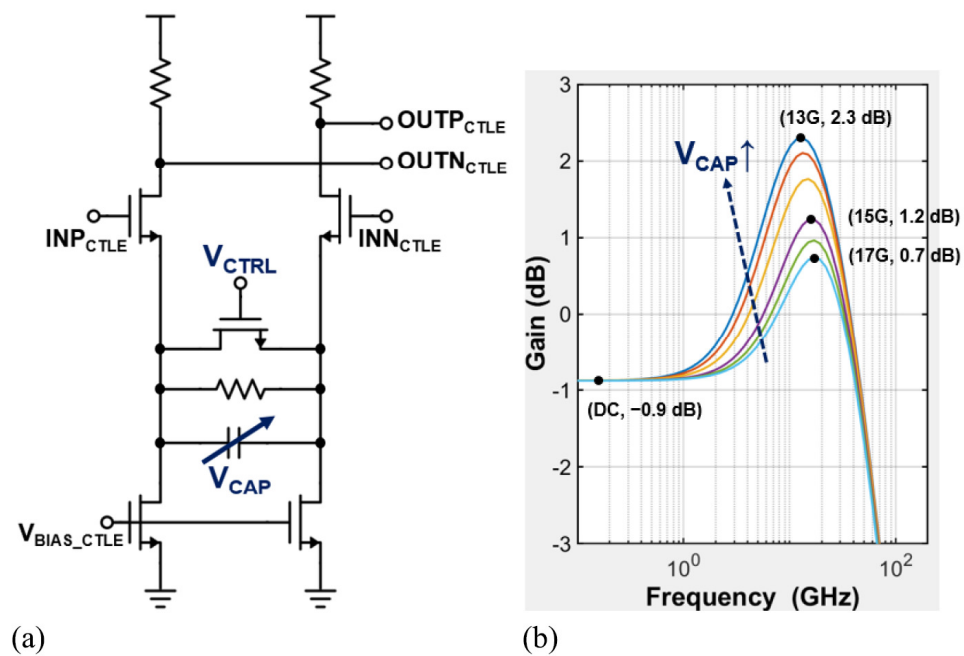


Fig. 4.4. (a) Schematic of the source-degenerated CTLE. (b) Simulated frequency response of the CTLE (single stage) with different settings of  $V_{CAP}$ .

frequency peaking can be enabled or disabled by setting  $V_{CTRL}$  to be logic low or logic high, respectively. As shown in Fig. 4.4(b), the peaking frequency is digitally adjustable by varying the voltage level of  $V_{CAP}$ . Since the source-degenerated resistance remains unchanged, the dc gain of the CTLE is approximately 0.9 (V/V), independent of the setting of  $V_{CAP}$ . Without the inclusion of inductors, the frequency boost at 15 GHz is simulated to be 2.1 dB for a single CTLE stage. The voltage level of  $V_{CAP}$  is set by an 8-bit on-chip voltage digital-to-analog converter (DAC), and the implementation of which follows the conventional resistor ladder R-2R architecture as presented in [41]. The voltage DAC therefore provides a dc voltage with 8-bit resolution between the ground (0 V) and a reference voltage,  $V_{HIGH}$ , where the value of  $V_{HIGH}$  can be changed via a pad connected to an external voltage source. In this prototype, an on-chip voltage DAC bank consisting of duplications of the aforementioned 8-bit voltage DAC is responsible for generating the digitally adjustable voltage levels. For further reduction in the area overhead, the resolution of each voltage DAC can be individually optimized with respect to the associated circuit blocks.

### 4.2.3 Summer

The summers used in the PAM4 receiver fall in the category of resistively loaded CML summer, and the architecture incorporating 2-tap DFE summation is shown in Fig. 4.5. Resistive source-degeneration is employed for linearity improvement. Depending on the previous two symbols resolved by the three data slicers; that is, the corresponding six thermometer-coded digital signals in differential fashion, the six tail currents are respectively steered to one of the two load resistors to perform DFE summation. To maintain the common-mode voltage level at the summer outputs irrespective of the DFE setting, all these tail currents are summed and mirrored to a common-mode restoration block which injects the currents evenly from the supply into the summing nodes ( $OUTP_{SUM}$  and  $OUTN_{SUM}$ ). The common-mode restoration allows the threshold setting and delay performance of the slicers to be independent of the DFE setting.

The schematic of the common-mode restoration circuits is shown in Fig. 4.6(a). It is similar

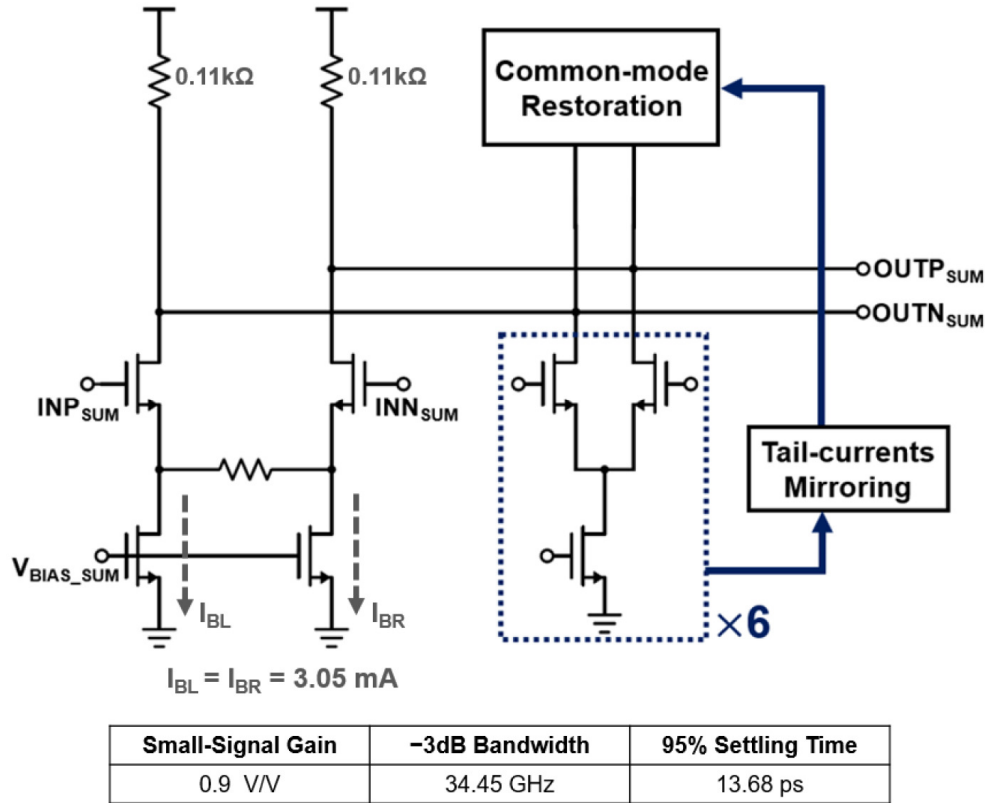


Fig. 4.5. Architecture and performance of the summer for 2-tap DFE.

to that in a prior art [45], while an additional function is included in this work for offset compensation. The common-mode restoration currents,  $I_{CMP}$  and  $I_{CMN}$ , are nominally half of the sum of all DFE currents that is,  $(3I_{DFE1} + 3I_{DFE2})/2$ . The offset cancellation currents,  $I_{OSP}$  and  $I_{OSN}$ , are individually adjustable to compensate the accumulated dc offset of the CTLE stages and the summer. In this prototype, a closed offset-cancellation loop is not implemented, and the values of  $I_{OSP}$  and  $I_{OSN}$  are adjusted with on-chip voltage DACs. Due to the finite output resistance of the current sources and current mirrors, larger errors can be introduced when the currents to be copied become larger. As shown in Fig. 4.6(b), simulations have been carried out to study the deviations from the target common-mode voltage level, with distinct settings of DFE currents. It can be seen that without the common-mode restoration circuits, the output common-mode level of the summer drops roughly linearly with the increase of DFE currents; the voltage drop of output common-

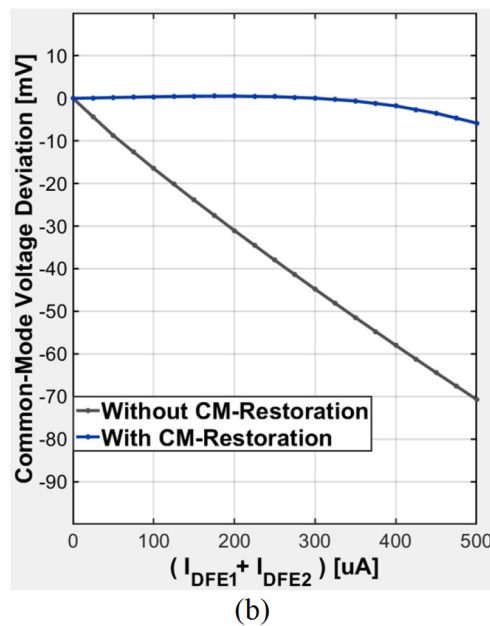
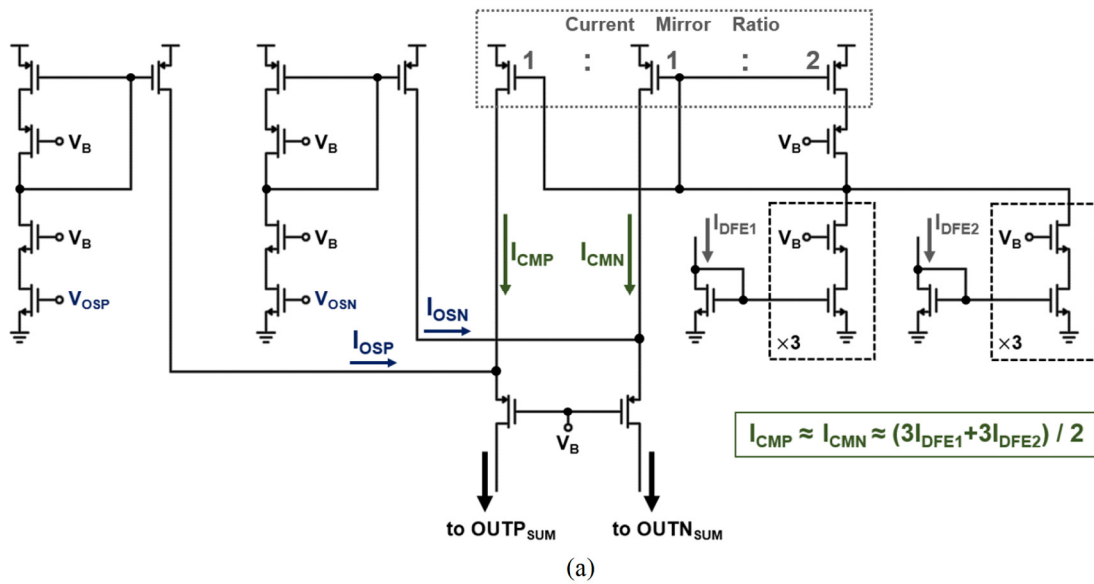


Fig. 4.6. (a) Schematic of the common-mode restoration circuits. (b) Simulated performance of the common-mode restoration circuits, showing the deviation from the target common mode with and without the common-mode restoration circuits.

mode level is approximately 70 mV when  $(I_{DFE1} + I_{DFE2}) = 500 \mu\text{A}$ . By contrast, when the common-mode restoration circuits are connected, the voltage drop of output common-mode level is less than 6 mV when  $(I_{DFE1} + I_{DFE2}) = 500 \mu\text{A}$ , and a relatively constant output



common-mode level is sustained across the range shown in Fig. 4.6(b).

#### 4.2.4 Linearity Characterizations

Since the receiver front-end linearity performance is crucial for multilevel signaling, the linearity of the summer and that of the CTLE are respectively examined via the evaluations on the output eye linearity (EL) versus the input amplitude. Fig. 4.7(a) shows the nomenclature for PAM4 eye diagrams along with the definition used for PAM4 EL, where  $V_{\text{Amp}}$  denotes the peak-to-peak input amplitude;  $E_{\text{H}}$ ,  $E_{\text{M}}$ , and  $E_{\text{L}}$  measure the eye heights of the upper eye, the middle eye, and the lower eye, respectively. Clean PAM4 signals without level mismatch (i.e.,  $EL = 1$ ) are applied to the input of the summer, and the output EL of the summer is recorded for each given input amplitude. Similarly, to test the CTLE linearity, PAM4 signals of different amplitudes with  $EL = 1$  are generated, whereas these signals go through a channel with 4-dB loss at 15 GHz before being applied to the input of the CTLE. The two-stage CTLE is correspondingly configured to provide ~4-dB boost at 15 GHz, which is the target amount of peaking, as described in Section 4.2.2. The simulated results at different process corners are shown in Fig. 4.7(b) for the summer, and Fig. 4.7(c) for the CTLE. As variable gain amplifiers (VGAs) are not included in this prototype, the EL remains above 90%, only when  $V_{\text{Amp}}$  is not greater than ~450 mV.

#### 4.2.5 CML-to-CMOS Clock Converter

Fig. 4.8(a) shows the schematic of the CML-to-CMOS clock converter. It consists of a differential amplifier and two stages of ac-coupled inverter-based clock amplifier. The use of ac coupling capacitor and inverter with the input node connected to the output node via a resistor ensures that the dc level of the clock signals is biased to around half of the supply voltage. The CML-to-CMOS clock converter is able to amplify incoming sinusoidal clock signals to rail-to-rail (i.e., CMOS levels) at various clock frequencies, provided that the amplitude of the input sinusoidal signals is sufficiently large. Fig. 4.8(b) summarizes the minimum required peak-to-peak amplitudes at different frequencies such that the swings of the clock signals at the converter output are larger than 50–850 mV. In particular, for

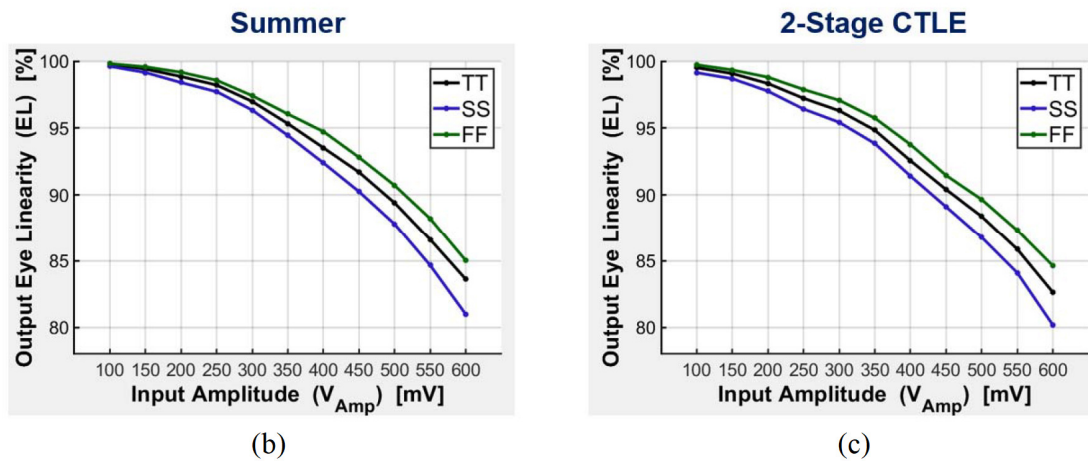
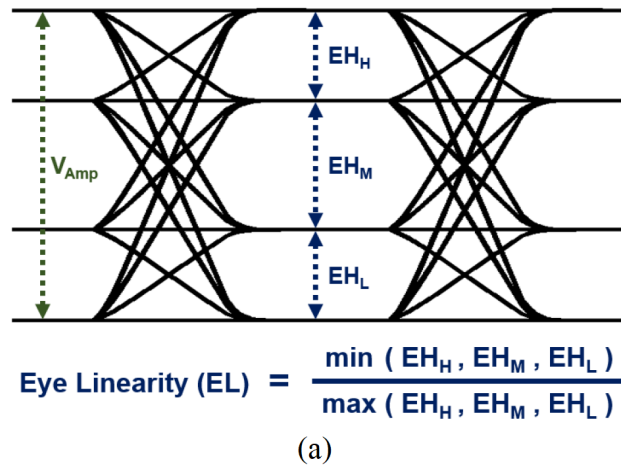


Fig. 4.7. (a) Nomenclature for PAM4 eye diagrams and the definition for PAM4 EL. (b) Simulated linearity performance of the summer. (c) Simulated linearity performance of the CTLE.

15-GHz clock signals, 24-mV<sub>pp</sub> input amplitude is needed for output swing larger than 50–850 mV, and 40-mV<sub>pp</sub> input amplitude further increases the output swing from approximately ground (0 V) to the supply voltage (900 mV). By providing larger input amplitudes, this CML-to-CMOS clock converter can work at higher frequencies.

#### 4.2.6 DCC Circuits

Duty-cycle distortion effectively induces unequal time frames for the operations (e.g., sampling, or data recovery) in different data paths, and therefore duty-cycle distortion can

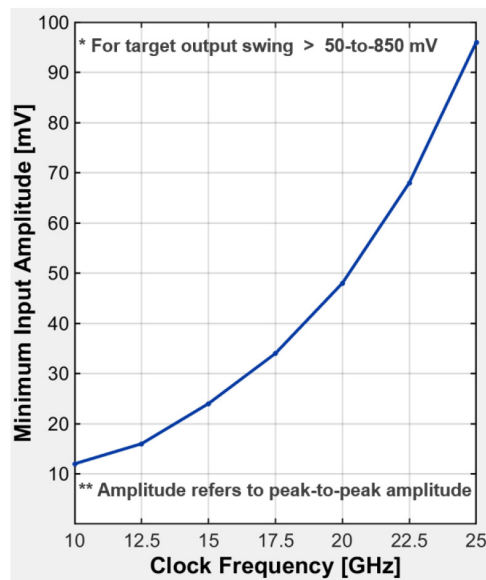
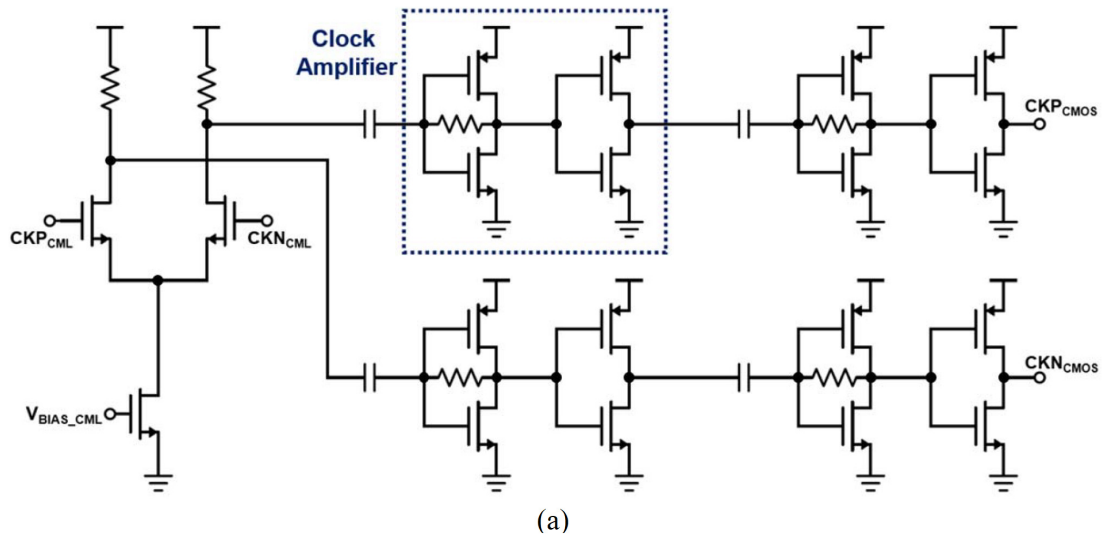


Fig. 4.8. (a) Schematic of the CML-to-CMOS clock converter. (b) Simulated minimum required input peak-to-peak amplitude with different input clock frequencies for the CML-to-CMOS clock converter.

be highly undesirable for high data-rate designs where the performance such as BER is sensitive to the unwanted reduction or imbalance of the timing allocation. In light of the negative effects of the duty-cycle distortion, DCC circuits are designed and implemented on the chip. Fig. 4.9(a) presents the schematic of the DCC circuits. The duty-cycle is

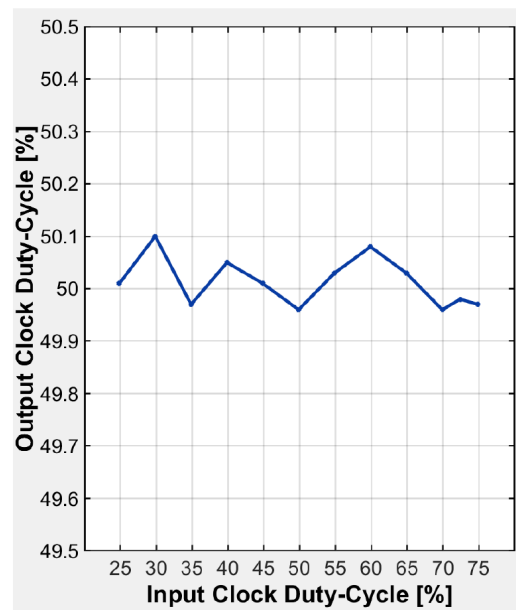
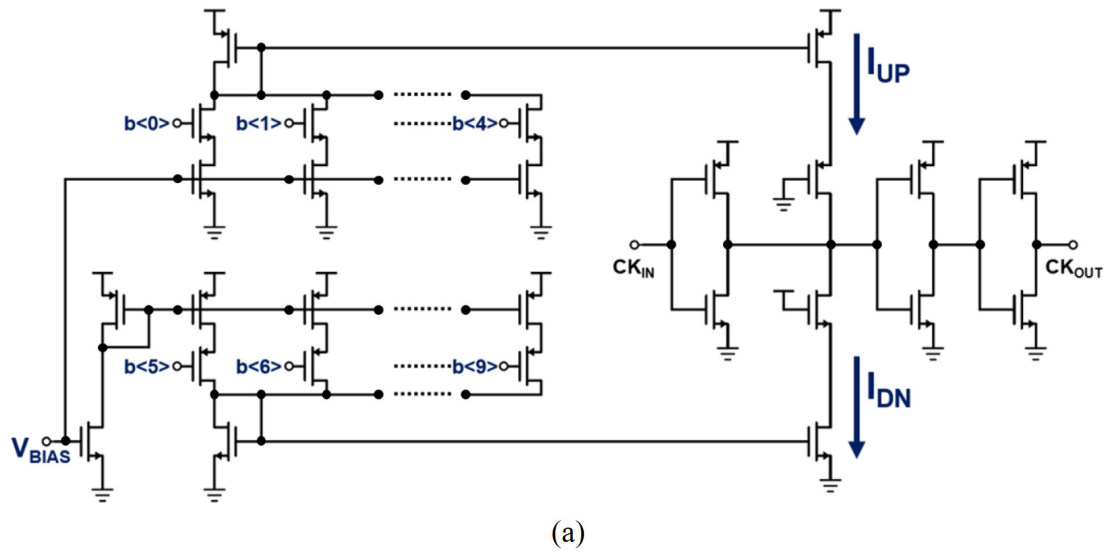


Fig. 4.9. (a) Schematic of the DCC circuits. (b) Simulated performance of DCC with 15-GHz clock signals.

adjusted by varying the amounts of the currents,  $I_{UP}$  and  $I_{DN}$ , which are digitally programmable by 10 bits,  $b\langle 9:0 \rangle$ . In addition, to be capable of accommodating both large duty-cycle distortion and fine-tuning, the value of  $V_{BIAS}$ , which sets the current level of the

current sources, is designed to be also digitally adjustable with an on-chip DAC. Simulation results of the DCC at 15 GHz are shown in Fig. 4.9(b). With the simultaneous programmability of the values of  $I_{UP}$ ,  $I_{DN}$ , and  $V_{BIAS}$ , the DCC is able to correct the input clock signal with duty-cycle of 25%–75% such that the duty-cycle of the output clock signal is very close to 50% with errors not greater than 0.1%. Provided that the duty-cycle of the input clock source to the receiver chip is 50%, Monte Carlo simulations show that the resultant duty-cycle at the outputs of the on-chip clock path varies from 48.46% to 52.48%. Accordingly, the presented DCC well covers the range due to process variations, and also has the competence to accommodate an input clock source whose duty-cycle deviates from 50%. In this work, an on-chip adaptive closed loop for setting the DCC is not implemented, but the setting is swept with the aim of optimizing the measured BER instead.

### 4.3 Slicer Design

#### 4.3.1 Slicer Overview

Voltage comparators, also known as slicers, or sense amplifiers in some contexts, have served widely in mixed-signal circuits and systems, including ADCs, adaptive configuration loops, memory access circuitry, and data receivers. A variety of slicer topologies with their practical utility have been demonstrated. In [48], CML slicers appeared in the implementation of a 6-bit ADC, and later the CML slicer topology has been frequently employed in data receivers [46], [47], [49], [50]. A CMOS latch-type comparator, known as the StrongArm and originally studied in memory circuits [51], became popular due to its often negligible static power consumption and the competence to generate rail-to-rail output swings. The StrongArm has found broad application in both low-power architectures and high-speed receivers, and its mechanism appears to incite inventions or variants of CMOS latch-type slicers. A 2-stage topology called double-tail latch-type voltage sense amplifier is presented in [36], which enhances the capability of operating at lower supply voltages and input common-mode voltage levels, by having fewer stacking transistors and separate tail currents for the input stage and the latch stage.

The slicers used in [52] and [53] are both essentially variants of the double-tail latch-type slicer [36], with an augmented function to incorporate 1-tap DFE summation. Another 2-stage slicer is reported in [35], where it is mentioned that increasing the common mode for the same clock-to-Q delay is enabled. Compared with the StrongArm, the aforementioned latch-type slicers ([35], [36], [52], [53]) attempt to conform the delay performance among an extended range of supply voltage or input common-mode levels, without much emphasis on considerably reducing the achievable delay. As illustrated in Fig. 4.2, the clock-to-Q delay performance of the slicers plays a critical role in closing the DFE loops. The next subsection describes the features of the particularly prevalent two slicer circuits, that is, the StrongArm and the CML slicer, and discusses the potential improvements.

### 4.3.2 Prevalent Slicer Topologies

The schematic of the StrongArm is shown in Fig. 4.10(a). It is designed in a dynamic CMOS latch fashion and its typical operation is illustrated in Fig. 4.11. When the clock (CK) is logic low, the outputs are both being charged to the supply value such that the differential output is reset to approximately zero. When the clock becomes logic high, the StrongArm samples the differential input and then the differential output is regenerated toward rail-to-rail with the help of the positive feedback offered by the cross-coupled pairs. A few observations can be made after closely examining the simulated waveforms. First, attributing to the reset mechanism, there is always certain time that needs to be spent for the differential output signal to grow from approximately zero. Second, since the time allocated for regeneration is limited by the data rate, the regeneration started with a higher level is very beneficial in that at the end of the regeneration phase, the differential output swing can be considerably larger and the delay to achieve digital level is also significantly less. In other words, for high data-rate operations, the time required for the output signals of the StrongArm to grow from approximately zero to the level that can be identified as digital outputs may not be sufficient. The above observations motivate the idea to design a slicer which instead of resetting tracks the polarity of the differential input signal such that the regeneration can proceed with a higher signal level. Another prevalent slicer topology

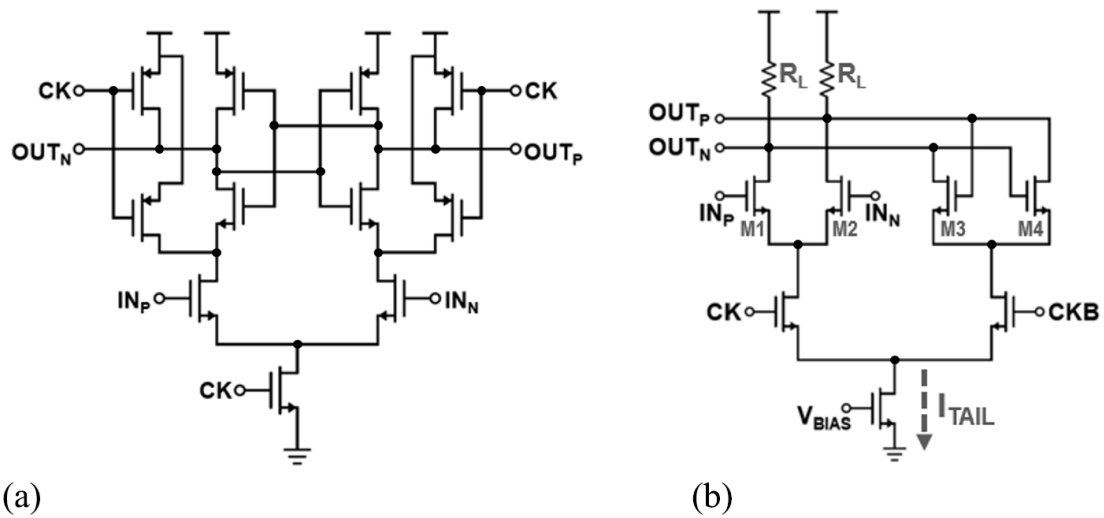


Fig. 4.10. Prevalent slicer topologies. (a) StrongArm slicer. (b) CML slicer.

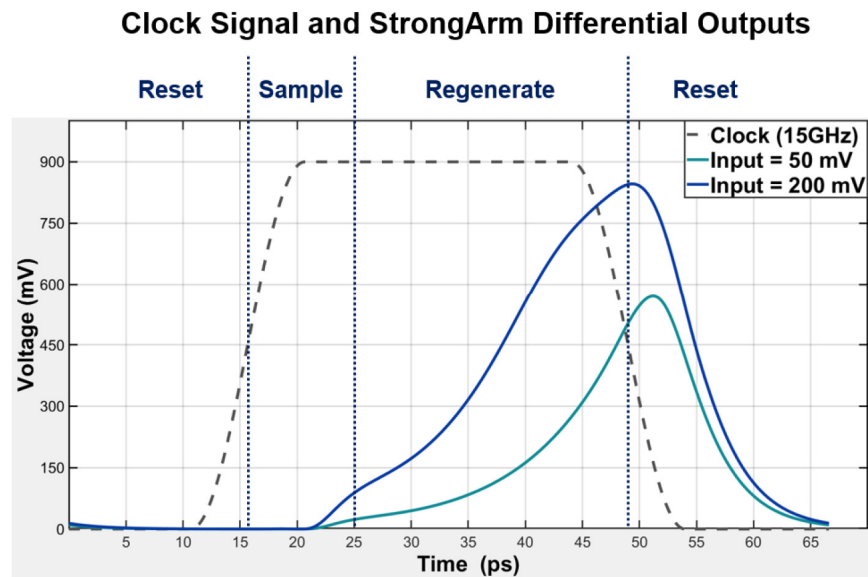


Fig. 4.11. Simulated waveforms showing the typical operations including the reset, sample, and regenerate phases of the StrongArm slicer.

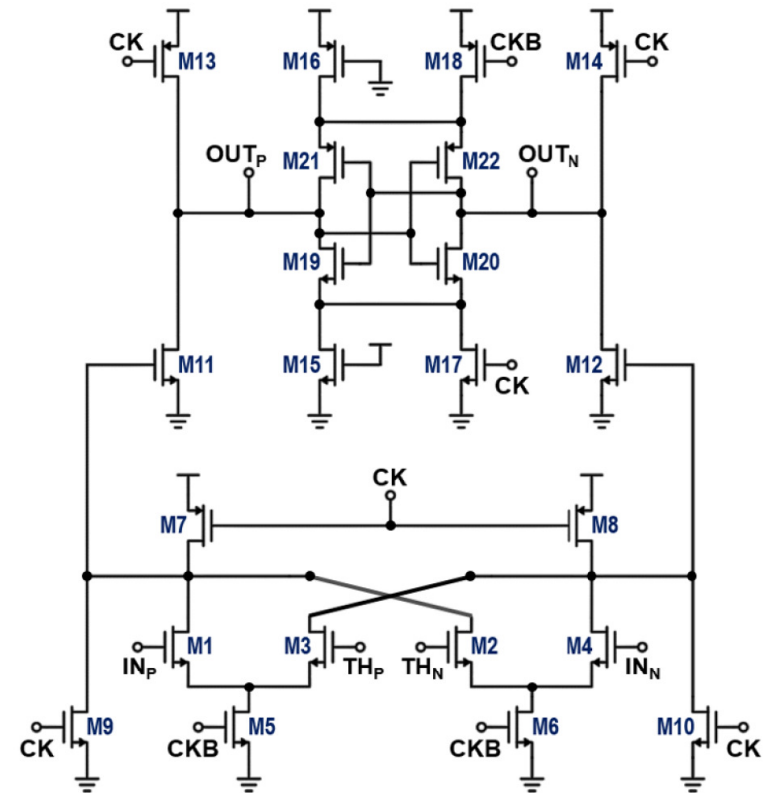
shown in Fig. 4.10(b), known as the CML slicer, leverages the idea of tracking the inputs. However, as the output swing magnitude of the CML slicer cannot exceed the product of the tail current and the load resistance, ( $I_{TAIL} \times R_L$ ), a number of drawbacks are associated with the CML slicer when implementing a PAM4-DFE with it. For one thing, since the

output swing is not rail-to-rail, the CML slicer may not be directly compatible to the relatively energy-efficient CMOS gates for delaying or buffering the resolved data, and a potential solution by inserting CML-to-CMOS amplifiers would increase the total delay. For another thing, the smaller output swing offers less strength to steer the DFE currents, and therefore, the sizes of the differential pairs of the DFE current branches cannot be minimized, which equivalently adds restrictions in minimizing the load capacitance at the summer outputs. Furthermore, referring to Fig. 4.10(b), because M1 and M2 are directly connected to M3 and M4, when the CML slicer is designed for larger output swing with large tail current, it leads to relatively high power consumption, and at the same time presents large input and output capacitances.

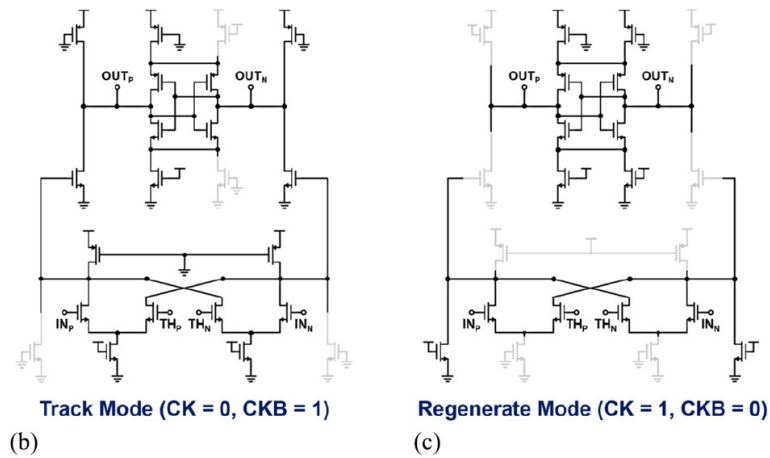
### 4.3.3 CMOS Track-and-Regenerate Slicer

In view of the above, a CMOS track-and-regenerate slicer is proposed and designed, aiming to improve the clock-to-Q delay as well as the output swing. When the DFE is implemented with the proposed slicer, digital-level outputs are directly available and the settling time specification of the summer is relaxed in consequence of the reduced slicer delay, enabling an energy-efficient DFE design that operates at high data rates. The overall circuit schematic of the proposed CMOS track-and-regenerate slicer is shown in Fig. 4.12(a). The proposed slicer tracks the differential input instead of being reset, and it regenerates the differential output to rail-to-rail levels. Designed in CMOS dynamic latch fashion, the proposed slicer is suitable for technology scaling and can be viewed as having three-stage configuration. The first stage, consisting of M1–M10, works as a dynamic differential amplifier. M11–M14 form the second stage, which serves as a buffer to provide some isolation between the first and the third stage. The third stage, M15–M22, is essentially dynamically controlled cross-coupled pairs that are responsible for regenerating the signal with positive feedback. Fig. 4.12(b) and (c), respectively, illustrates the operation of the proposed slicer during the two complementary clock phases. When CK is logic low and CKB is logic high, as in Fig. 4.12(b), M1–M8 and M11–M14 perform the tracking function with M9, M10, M17, and M18 turned off, and they overwrite the latch outputs ( $OUT_P$  and  $OUT_N$ ). M15 and M16 are kept always on and conduct relatively weak currents to avoid





(a)



(b)

(c)

Fig. 4.12. Proposed CMOS track-and-regenerate slicer. (a) Overall circuit schematic. (b) Proposed slicer in track mode. (c) Proposed slicer in regenerate mode.

the cross-coupled pairs (M19–M22) recovering from being completely off while allowing the outputs to be easily overwritten. In the other half of clock cycle, that is, when CK is

logic high and CKB is logic low, as in Fig. 4.12(c), the tracking function is stopped with the outputs of the first stage being cleared and the second stage disabled. The outputs of the first stage are cleared by M9 and M10 which discharge the output node voltages toward zero, and hence eventually turn off M11 and M12. With M11 and M12 turned off by M9 and M10, respectively, M13 and M14 turned off with the rise of CK, the second stage is quickly disabled, isolating the continuously changing inputs from the latch outputs which shall be regenerated toward rail-to-rail levels with respect to the polarity that has been tracked. At the same time, the cross-coupled pairs conduct significantly more currents by turning on M17 and M18, empowering strong positive feedback for the regeneration. It is noteworthy that clearing the outputs of the first stage with M9 and M10 during the regenerate-mode also helps with tracking the inputs for the next tracking phase, thanks to the fact that the first stage itself does not memorize the results from the previous tracking phase. In this work, the threshold level of the slicer is determined by the gate voltages  $TH_P$  and  $TH_N$ , and each slicer has its own individually adjustable threshold generator. The threshold levels are programmable from an external field-programmable gate array (FPGA) and set by on-chip voltage DACs that are described in Section 4.2.2. The slicer offset can be compensated by setting  $TH_P$  and  $TH_N$  correspondingly for a given threshold.

#### 4.3.4 Simulation Results

For the purpose of demonstrating the features of the proposed CMOS track-and-regenerate slicer and comparing the proposed slicer with the StrongArm which is also compact in size and suitable for technology scaling, extensive simulations have been carried out and the results are presented as follows. Fig. 4.13 illustrates the large-signal behavior and clock-to-Q delay performance of the proposed CMOS “track-and-regenerate” slicer along with the “reset-and-regenerate” StrongArm slicer. The input signals to the slicers are shown in Fig. 4.13(a), representing a worst-case pattern when a weak negative symbol, that is, (MSB, LSB) = (-1, +1), is between a long sequence of strong positive symbols, i.e., (MSB, LSB) = (+1, +1). Using this input pattern, the worst-case delay performance for PAM4 signaling can be evaluated and the memory effect in the proposed slicer with non-resetting mechanism is examined. As shown in Fig. 4.13(b), in contrast to the conventional CML

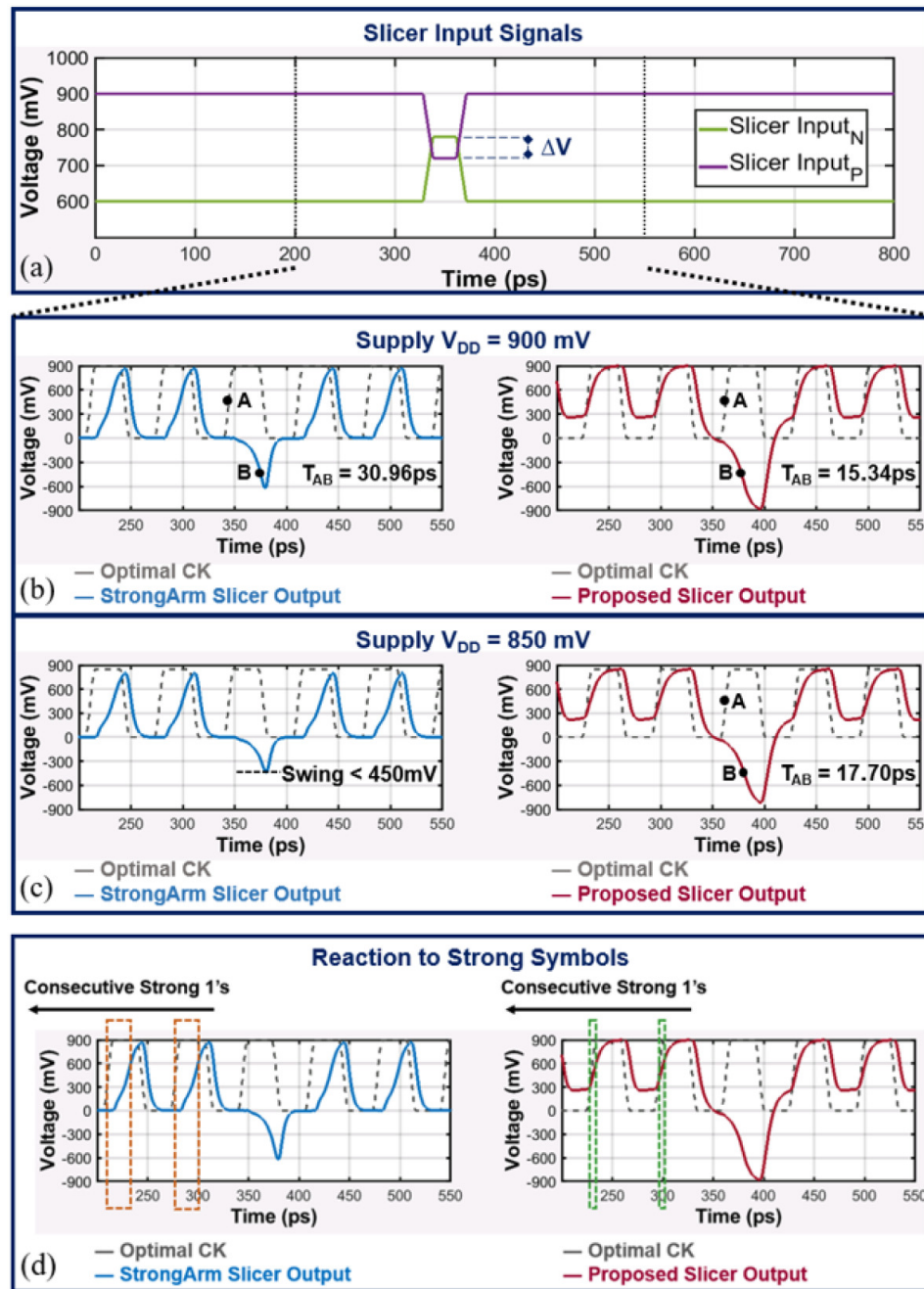
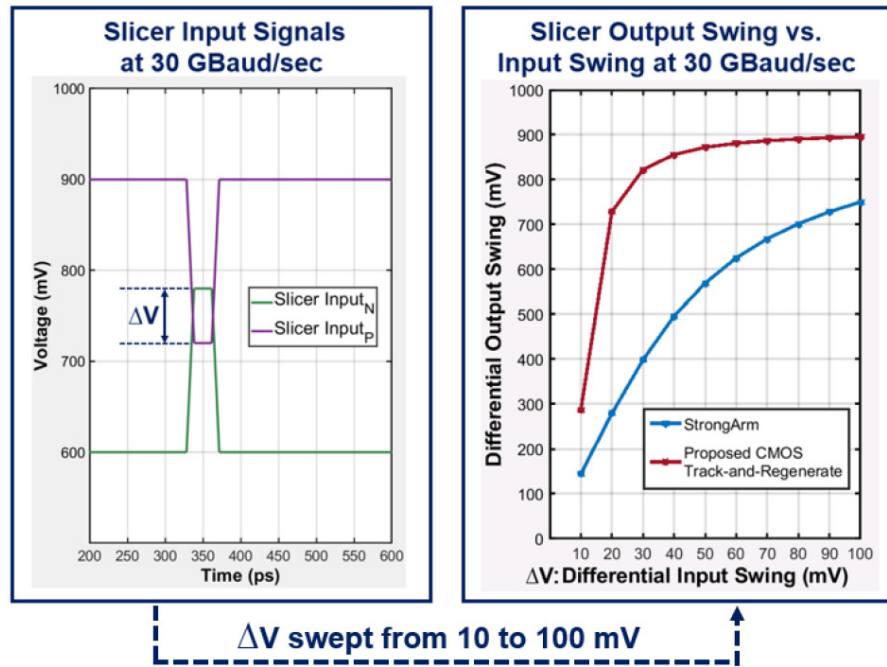


Fig. 4.13. Simulations and comparisons of the large-signal performance between the reset-and-regenerate StrongArm and the proposed CMOS track-and-regenerate slicer. (a) Input signals to the slicers. (b) Optimal clock signals and the resulting output waveforms of the slicers with 900-mV supply. (c) Optimal clock signals and the resulting output waveforms of the slicers with 850-mV supply. (d) Faster reaction to strong symbols with the proposed slicer.

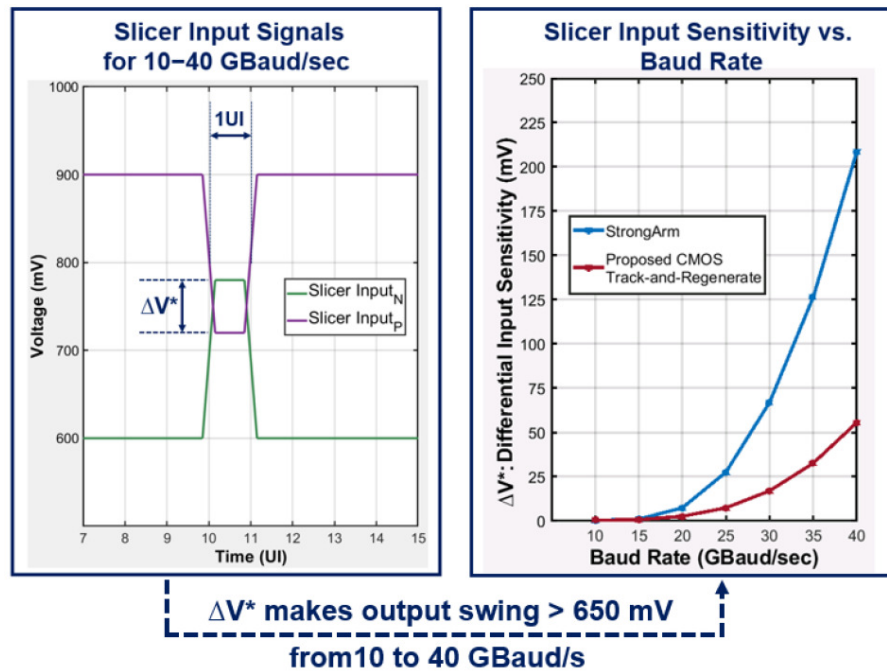
slicer, the proposed slicer offers rail-to-rail output swings and thus direct availability of

digital-level outputs. Meanwhile, in comparison with the StrongArm slicer, instead of resetting the latch, the proposed slicer tracks the input signals like how the CML slicer does, helping to reduce the required regeneration time. As a result, the proposed slicer improves the clock-to-Q delay as well as the output swing over the StrongArm. With the sizes of the input transistors and the output cross-coupled pairs designed to be identical, the worst-case clock-to-Q delay (with respect to the switching points defined as  $\pm 450$  mV) is simulated to be 30.96 ps for the StrongArm, whereas the delay reduces to 15.34 ps for the proposed slicer. As already been shown in [59], when not operating with low supply voltages or low input common-mode levels, the delay performance of the double-tail slicer [36] is similar to that of the StrongArm. This is also observed when a double-tail is tested with the same data pattern shown in Fig. 4.13(a). The double-tail slicer is designed to have the same input stage and output cross-coupled pairs as the StrongArm, achieving 31.2 ps with the input common-mode level set to 750 mV, and 29.58 ps with the input common-mode level reduced to 600 mV, for the worst-case clock-to-Q delay. Fig. 4.13(c) furthermore shows the immunity to the change in the power supply. A voltage drop of 50 mV from a 900-mV supply hinders the StrongArm from resolving the weak negative symbol to digital level with its output swing less than 450 mV, while the output swing of the proposed slicer is still approximately rail-to-rail and the penalty on resolving the weak symbol is 2.36 ps of increase in delay. Fig. 4.13(d) emphasizes another desirable feature offered by the proposed slicer; namely, the fast reaction to strong symbols. Since the strong symbols tend to cause relatively strong ISI for the next symbols, it is beneficial to have fast reaction and thus fast decision on them. The fast reaction makes sure the DFE summation is completely settled so as to minimize the negative impact of the residual ISI caused by the strong symbols.

In addition to the improved clock-to-Q delay performance, the proposed slicer also holds superior output swing and input sensitivity to the StrongArm, as can be seen below. Fig. 4.14(a) shows the input pattern under tests, which is similar to Fig. 4.13(a), but with the magnitude of  $\Delta V$  swept from 10 to 100 mV instead. The results in the right of Fig. 4.14(a) suggest that the proposed slicer outperforms the StrongArm in the output swings by



(a)



(b)

Fig. 4.14. (a) Slicer input signals, and the simulated slicer output swings with distinct input swings at 30 GBaud/s. (b) Slicer input signals, and the simulated slicer input sensitivity at different baud rates.

recovering the input signal to a stronger output. Next, to investigate the slicer's capability of resolving a relatively weak input to a level that can be identified and further easily processed as a digital signal, the input sensitivity is defined as the minimum required differential input swing,  $\Delta V^*$ , such that the output swing of the slicer is larger than the digital level of 650 mV. The input pattern is depicted in the left of Fig. 4.14(b), where the baud rates are swept from 10 to 40 GBaud/s, and the value of  $\Delta V^*$  at each baud rate is searched to fulfill the target output swing. In the right of Fig. 4.14(b), the simulated sensitivity performance at different baud rates is plotted. The proposed CMOS track-and-regenerate slicer achieves better input sensitivity than the StrongArm, especially for the higher data rates shown in Fig. 4.14(b). To investigate the effects of input common-mode level and supply voltage variations, simulations used for Fig. 4.13 are further extended with distinct settings. The results are shown in Fig. 4.15, where the output swings and delay performance for strong symbols and weak symbols are individually characterized. It would be worthwhile to reiterate that the input differential pairs and the output cross-coupled pairs in both slicers are designed to be identical for fair comparisons, and therefore the two slicers present similar area and loading to the summer circuitries.

Techniques to simulate and examine the noise performance of periodically clocked slicers have been well studied in [54], by identifying the periodically clocked slicers as linear periodically time-varying (LPTV) systems. As shown in [54], the procedures to obtain the dominant signal-to-noise ratio (SNR) involve both periodic steady-state (PSS) and periodic noise (PNOISE) simulations in time domain, which find out the large-signal response of the slicer, and the noise power at any specified observation point, respectively. After running the PSS and PNOISE simulations with respect to the differential output of the slicer under test, the output SNR in voltage can be derived from dividing the large-signal response by the root-mean-squared (rms) noise voltage at each observation time step. Fig. 4.16(a)–(c) respectively show the simulated differential output signal, differential rms noise voltage, and the resultant differential output SNR, of the proposed slicer. Since it is the SNR before rapid regeneration that dominates the probability of decision errors [54], the input-referred noise should be evaluated accordingly. As labeled in Fig. 4.16(c), the

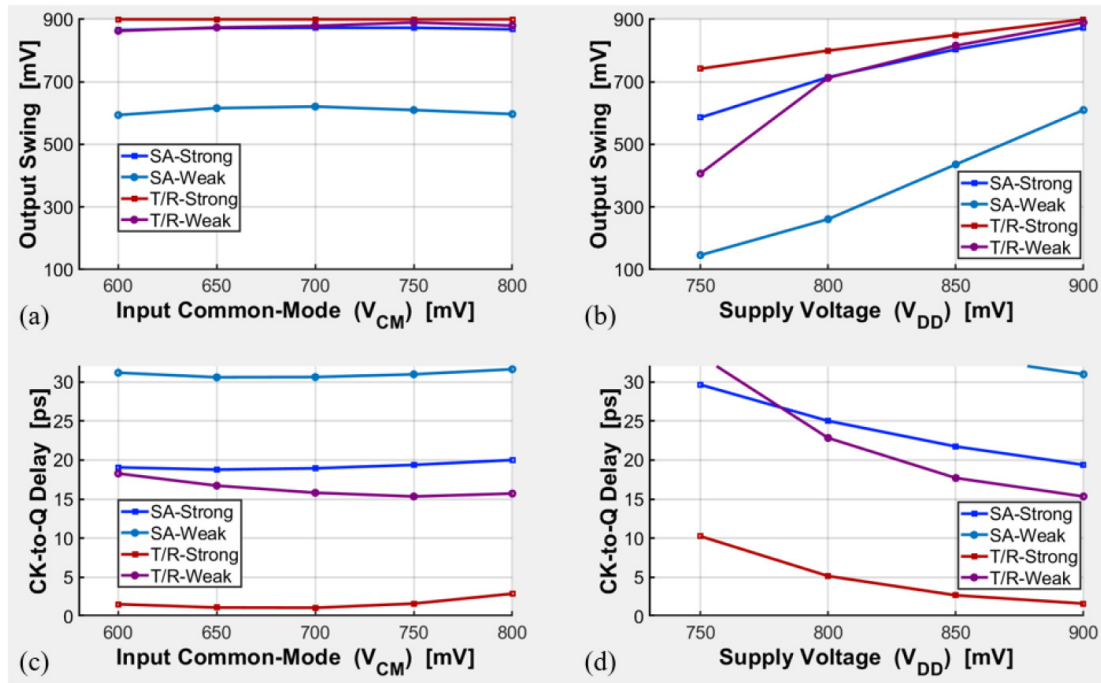


Fig. 4.15. Simulated output swing and clock-to-Q delay performance of the StrongArm (SA) and the proposed track-and-regenerate slicer (T/R) for typical strong symbols and weak symbols. (a) Output swing versus input common-mode level,  $V_{DD} = 0.9$  V. (b) Output swing versus supply voltage,  $V_{CM} = 0.75$  V. (c) Clock-to-Q delay versus input common-mode level,  $V_{DD} = 0.9$  V. (d) Clock-to-Q delay versus supply voltage,  $V_{CM} = 0.75$  V.

simulated differential output SNR before rapid regeneration is 31 dB (i.e., 35.5 V/V). Therefore, the differential input-referred noise is derived to be  $1.69 \text{ mV}_{\text{rms}}$ , given the differential input signal of 60 mV. The overall noise is investigated by referring the noise from other stages to the input of the slicer. The two-stage CTLE contributes  $0.80 \text{ mV}_{\text{rms}}$ , and the summer contributes  $0.58 \text{ mV}_{\text{rms}}$ , resulting in overall noise of  $1.96 \text{ mV}_{\text{rms}}$  at the slicer input.

In summary, the proposed CMOS track-and-regenerate slicer offers benefits of less delay and higher gain, thanks to its non-resetting mechanism when the allocated regeneration time becomes stringent. With the multistage architecture and the need for continuously conducting currents when performing the tracking function, the proposed slicer consumes more power. The noise, power, and offset comparisons of the slicers having identical input

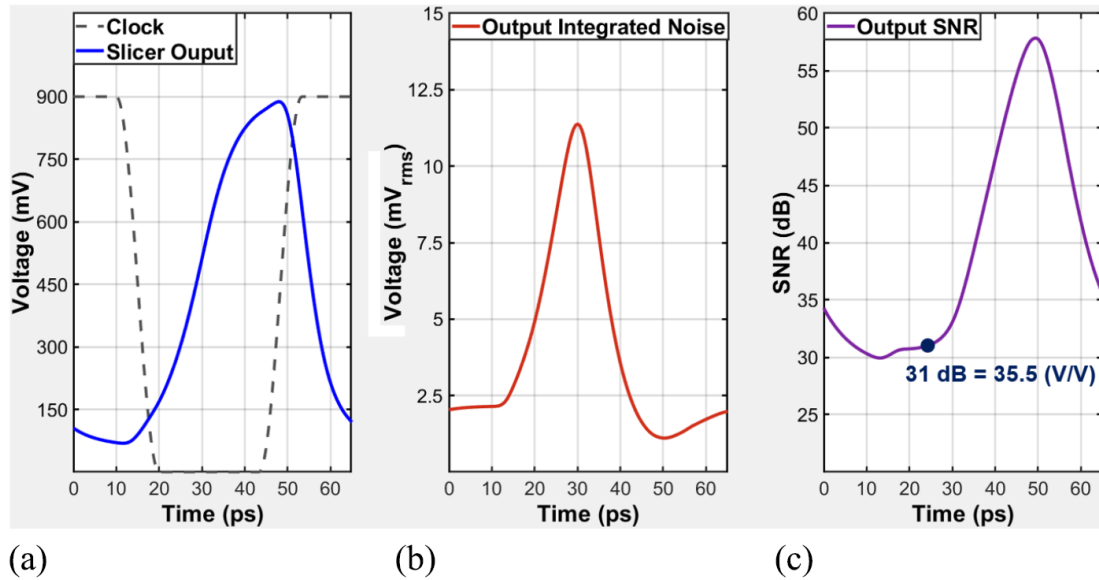


Fig. 4.16. Simulated results from the PSS and PNOISE simulations at 30-GBaud/s operation; the clock frequency is 15 GHz. (a) Clock signal and the proposed slicer's differential output signal from PSS simulation. (b) Simulated differential output integrated noise of the proposed slicer from PNOISE simulations in time domain. (c) Resultant differential output SNR of the proposed slicer.

	Proposed Slicer	StrongArm
<b>Input Capacitance (Loading to the summer)</b>	Identical (by design)	
<b>Output Capacitance</b>	Identical (by design)	
<b>Clock-to-Q Delay</b>	< ~16 ps	< ~31 ps
<b>Input-Referred Noise</b>	1.69 mV <sub>rms</sub>	3.37 mV <sub>rms</sub>
<b>Power Consumption (Clocked at 15 GHz)</b>	1.7 mW	0.43 mW
<b>Differential Offset Standard Deviation</b>	18.3 mV	20.9 mV
<b>Required Clock Phase(s)</b>	2	1

Table 4.1. Slicer comparisons (the proposed CMOS track-and-regenerate slicer vs. the conventional StrongArm slicer).

and output capacitances are summarized in Table 4.1.



#### 4.4 DFE Loops

In this section, the most stringent timing constraint for completing the DFE loops with the proposed slicer is examined. Referring back to the timing constraint diagram shown in Fig. 4.2, it can be inferred that for direct DFE loops, the tightest timing constraint lies in the loop of first tap, where

$$T_{CKQ} + T_{dh1} + T_{settle} + T_{setup} < 1 \text{ UI} \quad (4.1)$$

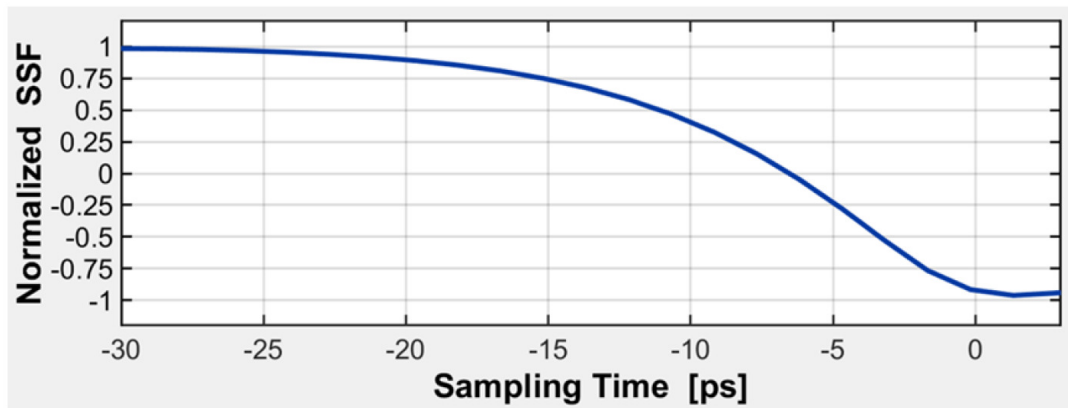
and 1 UI is 33.33 ps for 60-Gb/s PAM4 signaling, or 30-GBaud/s operation. With the StrongArm slicer presented in Section 4.3, aside from the undesirable smaller swing, it is nearly impossible to close the loop for the first tap of DFE at 30 GBaud/s, since its worst-case clock-to-Q delay ( $T_{CKQ}$ ) is 30.96 ps, leaving very little time for other parts to settle. By contrast, with the improved  $T_{CKQ}$ , the delay of the proposed slicer is not significant for strong symbols and is not more than 0.5 UI in the worst case shown in Fig. 4.13(b), allowing favorably more time for other operations to be finished. For  $T_{dh1}$  and  $T_{settle}$ , as mentioned previously, since the operations take place concurrently, it is more appropriate to view them as the additional delay with respect to  $T_{CKQ}$  during which the DFE tap currents and the summer have already started to settle toward their steady states. The setup time ( $T_{setup}$ ) is commonly used for digital gates or digital circuits to characterize the required time of arrival of digital inputs prior to the change of the state, for example, triggered by the rising/falling edge of the clock. In the context of analog-based DFE design, the idea of  $T_{setup}$  can be useful, whereas it is not directly associated with digital inputs anymore, but linked to the sampling aperture of the slicer. Specifically, the width of the sampling aperture of a slicer is characterized through an equivalent setup time. For instance, a wider sampling aperture suggests that the signal to have greater impact at the end of sampling phase needs to arrive earlier, and equivalently implies a larger value of  $T_{setup}$ . As the impulse response is used for describing a linear time invariant (LTI) system, the impulse sensitivity function (ISF) reveals the time-dependent sensitivity of the output at a certain observation time, to the impulse input with a specific arrival time. The ISF of a slicer can be interpreted as the weighted time-average sampling function, and the sampling aperture

corresponds with the shape of ISF. More fundamentals and details of ISF and LPTV systems can be found in [49], [54], and [55]. The approach to simulating the ISF of a clocked slicer has been developed and presented in [55]. First, a step function and a fixed offset are applied as inputs, where the height of the step function, that is, step-height, is self-adjusted by a feedback loop. The step sensitivity function (SSF) is obtained by searching the step-height that makes the slicer metastable at each time step. And then, the ISF is derived from taking the derivative of SSF.

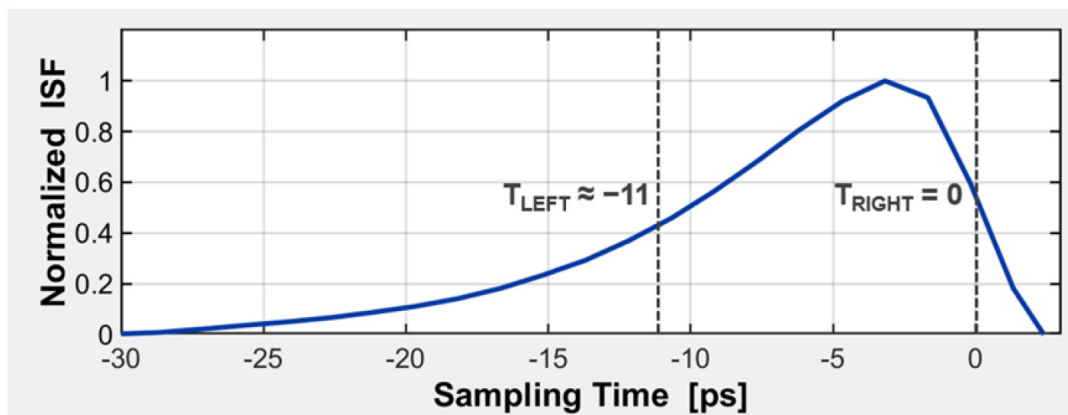
Fig. 4.17(a) shows the simulated SSF of the proposed track-and-regenerate slicer, and its normalized ISF is shown in Fig. 4.17(b), both at 30-GBaud/s operation. The sampling aperture can be defined as the time frame between  $T_{\text{LEFT}}$  and  $T_{\text{RIGHT}}$ , during which the integration of the area under the ISF from  $T_{\text{LEFT}}$  to  $T_{\text{RIGHT}}$  is 80% of the total area under ISF. The width of sampling aperture, that is,  $(T_{\text{RIGHT}} - T_{\text{LEFT}})$ , indicates the sampling bandwidth [55], and furthermore the values of  $T_{\text{LEFT}}$  and  $T_{\text{RIGHT}}$  specify an effective timing window for applying inputs so as to have their responses at the output influential. For the purpose of studying the DFE timing constraint, we conveniently set  $T_{\text{RIGHT}}$  to be 0, that is, aligned with the rising/falling edge of the clock signals, and define the analog-fashion  $T_{\text{setup}}$  as 90% of the sampling aperture width. Namely

$$T_{\text{setup}} = 0.9 \times (T_{\text{RIGHT}} - T_{\text{LEFT}}). \quad (4.2)$$

As labeled in Fig. 4.17(b),  $T_{\text{RIGHT}}$  is 0 and  $T_{\text{LEFT}}$  is about  $-11$  ps from simulations, resulting in the  $T_{\text{setup}}$  of 9.9 ps for the proposed slicer. With the simulated values of  $T_{\text{setup}}$  and  $T_{\text{CKQ}}$  along with the usage of (4.1), the desirable requirement of additional delays from  $T_{\text{dh1}}$  and  $T_{\text{settle}}$  can be calculated. For example, from the simulation shown in Fig. 4.13(b), the worst  $T_{\text{CKQ}}$  is 15.34 ps, and thus  $(T_{\text{dh1}} + T_{\text{settle}}) < (33.33 - 15.34 - 9.9) = 8.09$  ps guarantees the effectiveness of the first-tap DFE loop. However, it is noteworthy to point out that in the case of employing the track-and-regenerate slicers, even if the  $(T_{\text{dh1}} + T_{\text{settle}})$  does not completely satisfy the calculated specification from (4.1), the first-tap DFE loop can still be closed as long as the feedback signal is within the sampling aperture, which effectively leads to degraded accuracy of DFE summation. In other words, depending on the tolerable



(a)



(b)

Fig. 4.17. (a) Simulated SSF of the proposed track-and-regenerate slicer at 30 GBaud/s. (b) Simulated ISF of the proposed track-and-regenerate slicer at 30 GBaud/s.

accuracy of DFE summation, the factor of 0.9 appearing in (4.2) can be revised. The timing diagram for the first-tap DFE in a half-rate design is shown in Fig. 4.18, along with the simulated numbers for timing constraints, where the values of  $(T_{\text{dh1}} + T_{\text{settle}})$  are measured as the additional delay relative to the  $T_{\text{CKQ}}$ . To prove that the direct DFE loops can be closed with the proposed slicers and thus successfully expand the eye-opening, simulations with only the equalization offered by the direct 2-tap DFE loops have been carried out, excluding the benefits from CTLE. The differential pulse responses at the input and the output of the summer are shown in Fig. 4.19(a) and (b), respectively. These pulse responses

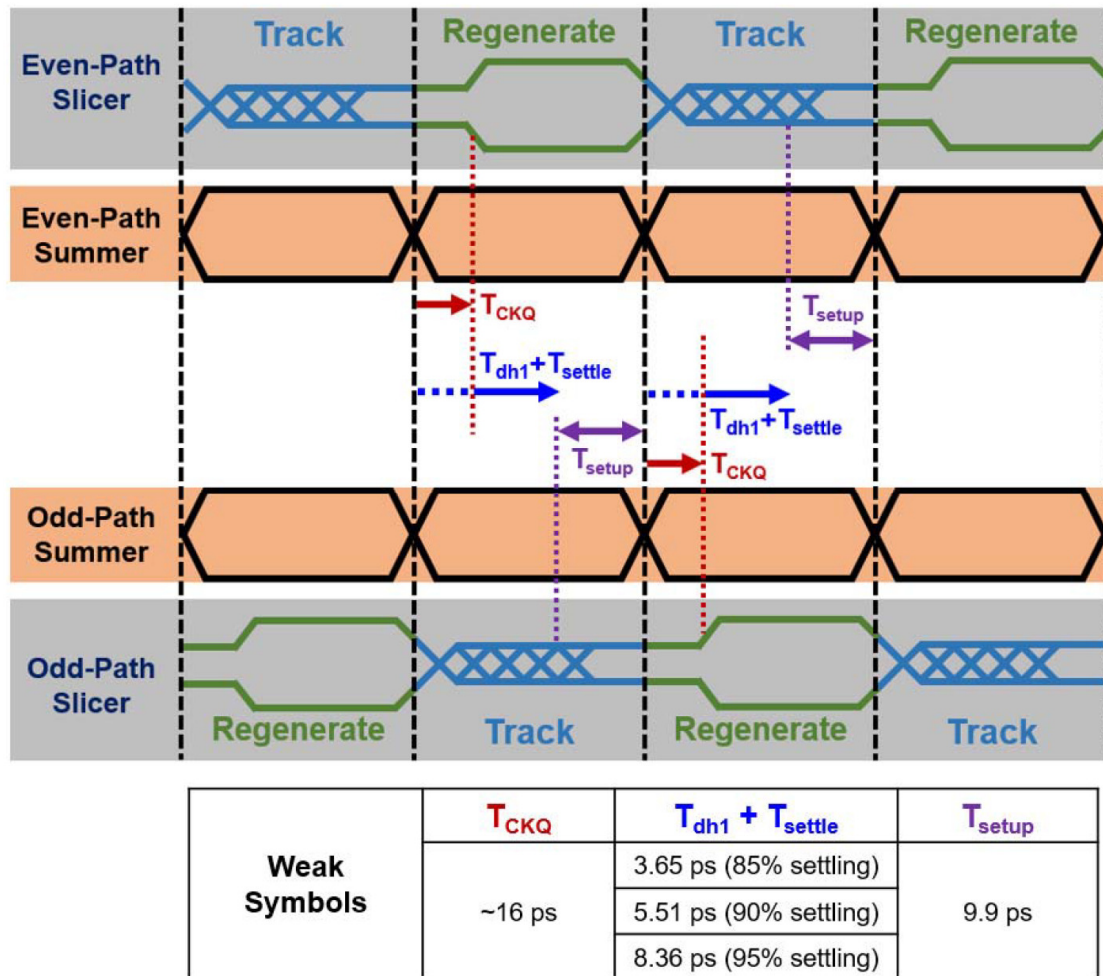


Fig. 4.18. Timing diagram for the first tap DFE in a half-rate design, and the simulated numbers for the timing constraints at 30 GBaud/s.

correspond to channel loss of ~6 dB at 15 GHz. When simulating the pulse responses and the DFE loops, in addition to the loading presented by the slicers, an additional capacitive load of 20 fF is added to each of the summer outputs for representing the input capacitance of clock and data recovery (CDR) circuits. Fig. 4.20(a)–(d) shows the simulated eye diagrams at 60-Gb/s PAM4 with distinct DFE settings. The input MSB and LSB patterns used in the simulations are both pseudorandom binary sequence-7 (PRBS-7), with the LSB pattern delayed by 5 bits relative to the MSB. The simulation results of the DFE match with those of the pulse responses. The first-tap DFE plays a critical role in opening the

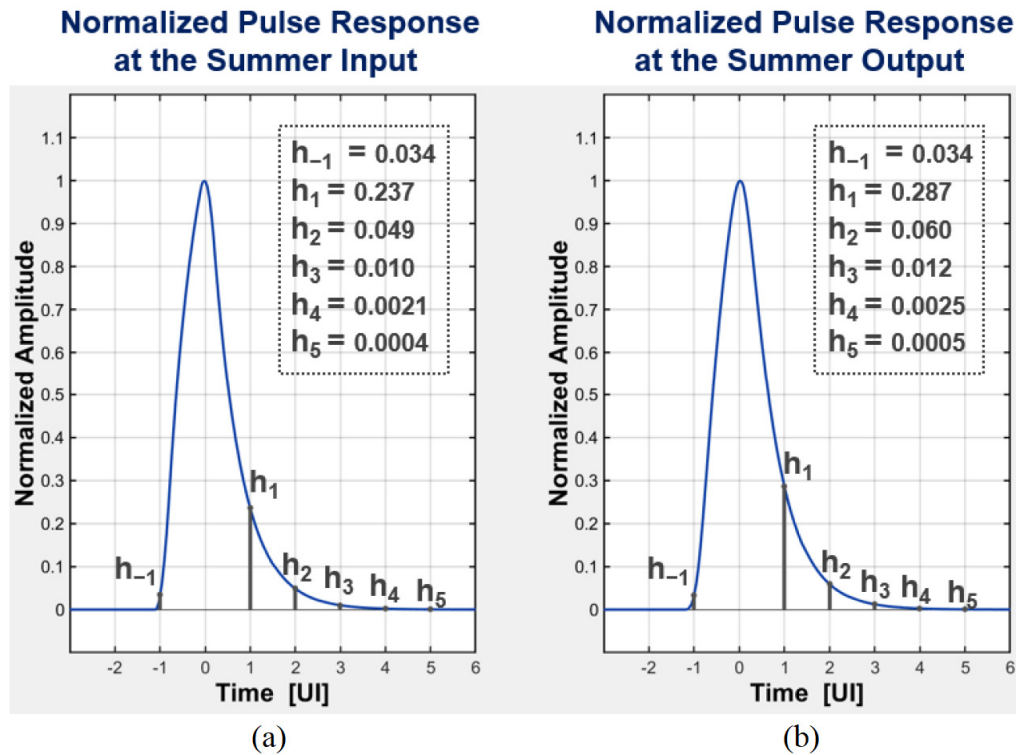


Fig. 4.19. Pulse responses with considerable post-cursor ISI. (a) Simulated normalized pulse response at the input of the summer. (b) Simulated normalized pulse response at the output of the summer.

eyes, and the simultaneous inclusion of the second-tap DFE further expands the eye-opening.

#### 4.5 Experimental Results

The PAM4 receiver chip was fabricated in 28-nm CMOS technology, and Fig. 4.21 shows the experiment setup. The receiver chip is wire-bonded to a printed circuit board (PCB). A high-speed pattern and clock generator transmits the PAM4 data and the half-rate differential clock signals to the chip via cables and PCB traces. The channel loss for the transmitted PAM4 data mainly consists of the loss of cables (48-in long) and PCB trace (~0.8 inch, FR4), which is measured to be 8.2 dB at 15 GHz excluding the bond wire. The associated 30-GBaud/s pulse response derived from S21 measurement, and the measured 60-Gb/s PAM4 eyes at the input of the receiver chip are shown in Fig. 4.22(a) and (b),

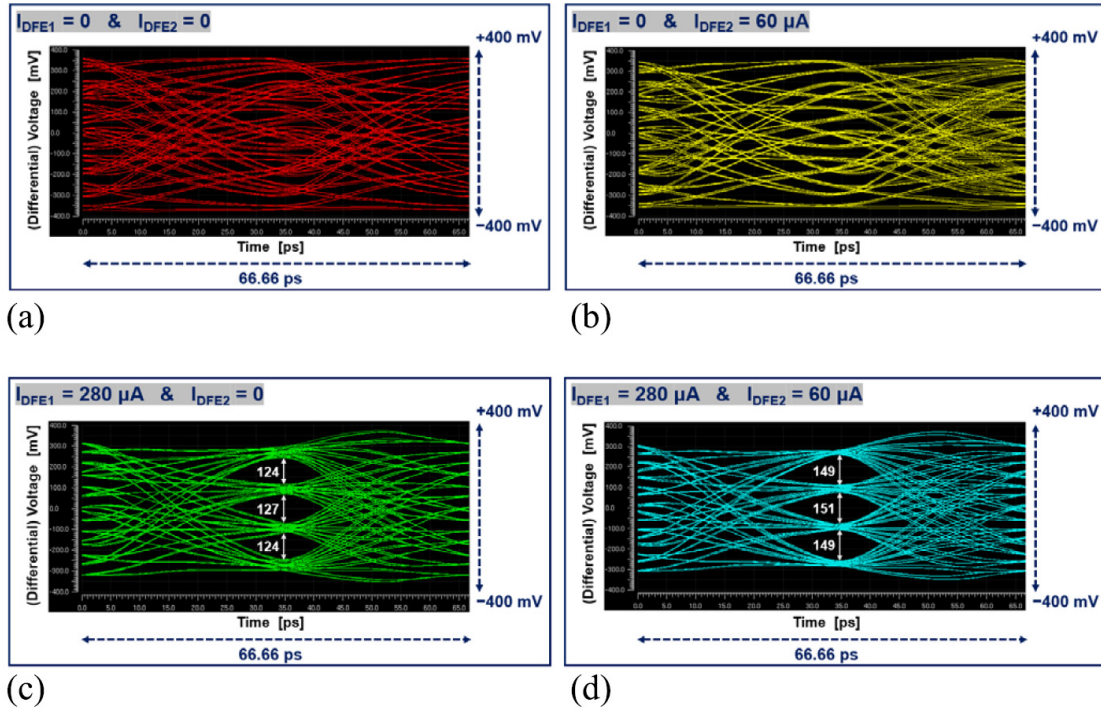


Fig. 4.20. Simulated differential output of the summer with distinct DFE settings. (a) First-tap DFE and second-tap DFE are both disabled. (b) First-tap DFE is disabled, while the second-tap DFE is enabled. (c) First-tap DFE is enabled, while the second-tap DFE is disabled. (d) First-tap DFE and the second-tap DFE are both enabled.

respectively. An oscilloscope is set up to measure the aforementioned input data eyes and for monitoring the recovered output data signals which are driven by on-chip CML drivers. In addition to the on-chip BEREC, an external commercial BER tester is connected to measure the BER and verify the function of the on-chip BEREC. To verify the effectiveness of the direct DFE loops implemented with the proposed slicer, PRBS-7, 9, 31 patterns have been fully tested and the bathtub curves with DFE loops disabled and enabled are measured at 60 Gb/s. As shown in Fig. 4.23(a), with DFE loops disabled, the measured BER is not better than  $1E-6$ , while with DFE loops enabled, the measured bathtub curve shows 0.15-UI horizontal opening for  $BER = 1E-12$ , when tested with PRBS-31 pattern. The eye contour map at 60 Gb/s, which is captured by the on-chip 2-D eye-monitoring circuits and shown in Fig. 4.23(b), confirms the open eyes after equalization. The 2-tap DFE coefficients are estimated to be  $(-0.212, -0.0311)$ , according to that the  $I_{DFE1}$  and  $I_{DFE2}$ ,

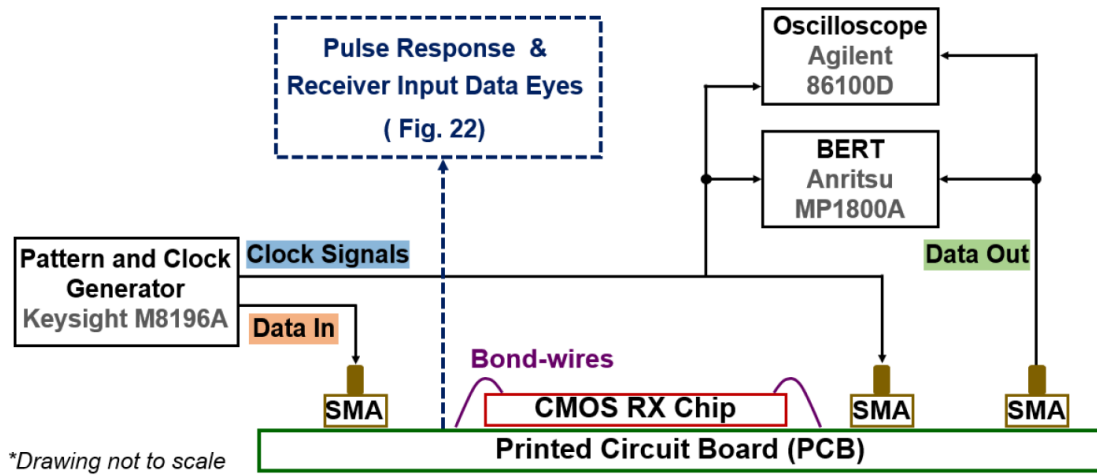


Fig. 4.21. Block diagram of the experiment setup.

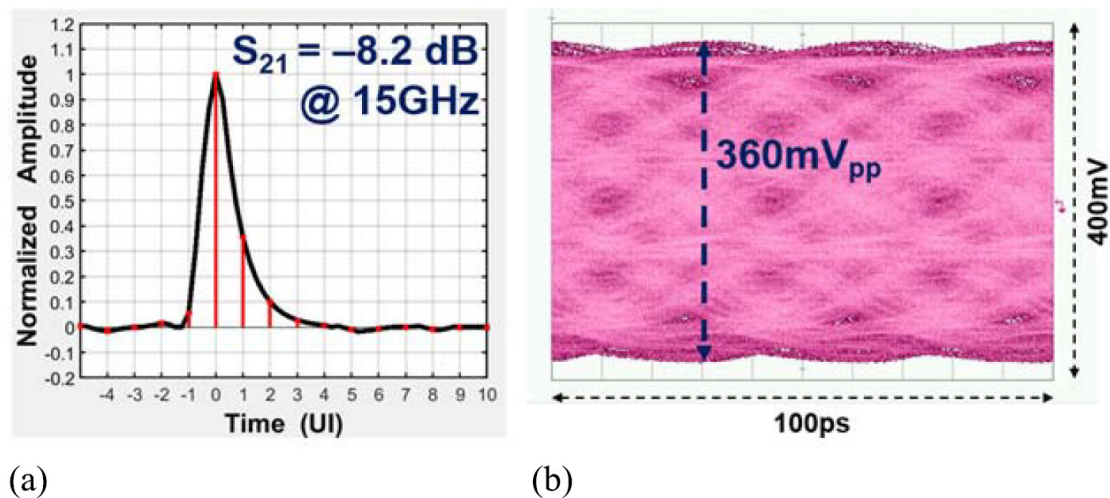


Fig. 4.22. (a) Measured 30-GBaud/s pulse response at the input of the receiver chip. (b) Measured (single-ended) 60-Gb/s PAM4 data eyes at the input of the receiver chip.

previously defined in Fig. 4.6(a), are set to 205 and 30  $\mu\text{A}$ , respectively.

The chip micrograph is shown in Fig. 4.24(a), with its key building blocks highlighted, including the CTLEs, the half-rate CML summers (Sum), the proposed slicers along with 2-tap DFE logics, the 1-to-32 data demultiplexer (DMUX), the synthesized BERF, the DCC circuits, CKBUFs, the digitally controlled DL, and the on-chip voltage DAC (VDAC)



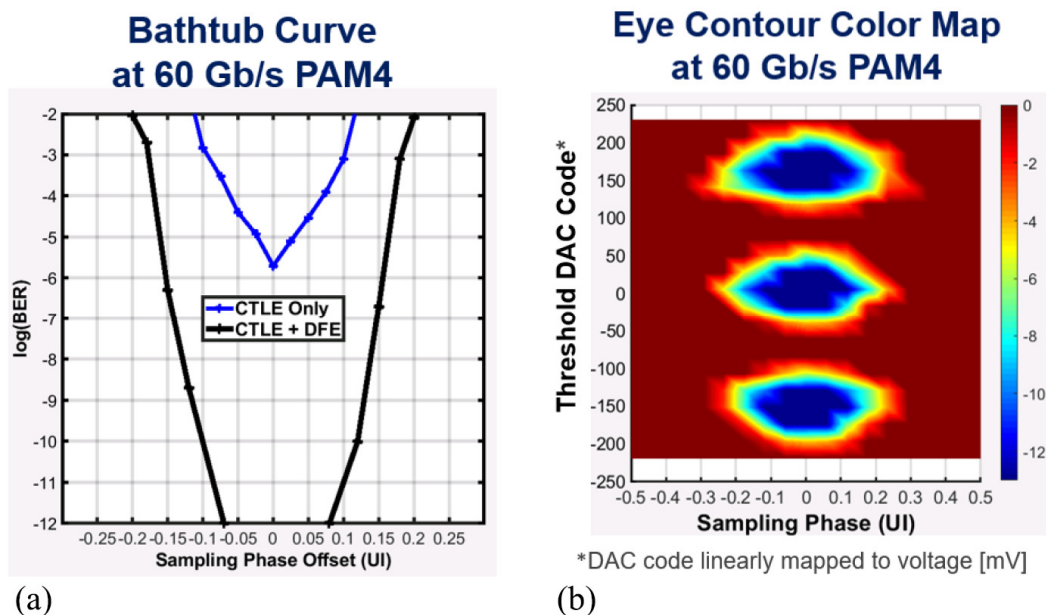


Fig. 4.23. (a) Measured bathtub curves at 60-Gb/s PAM4, with DFE loops disabled/enabled. (b) Measured eye contour color map at 60-Gb/s PAM4 after equalization.

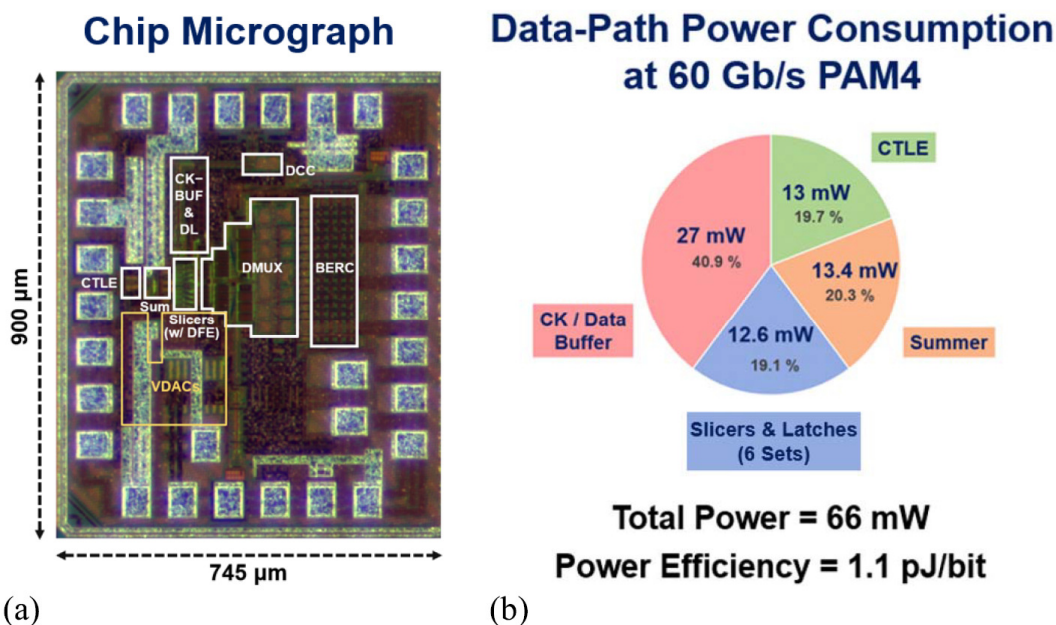


Fig. 4.24. (a) Chip micrograph with key building blocks highlighted. (b) Measured receiver data-path power consumption at 60-Gb/s PAM4.



	This Work	JSSC'17 [45]	JSSC'17 [56]	JSSC'17 [37]	ISSCC'18 [57]	VLSI'19 [58]	JSSC'19 [46]
Technology	28-nm CMOS	16-nm FinFET	65-nm CMOS	65-nm CMOS	16-nm FinFET	40-nm CMOS	65-nm CMOS
Data Rate	60 Gb/s	40-56 Gb/s	60 Gb/s	32 Gb/s	19-56 Gb/s	52 Gb/s	56 Gb/s
Modulation	PAM4	PAM4	NRZ	PAM4	PAM4	PAM4	PAM4
Equalization	CTLE 2-Tap DFE	CTLE 10-Tap DFE	CTLE 2-Tap FFE 3-Tap DFE	2-Tap TX FFE 1-Tap FIR and 2-Tap IIR DFE	CTLE	CTLE FFE	CTLE 1-Tap FIR and 1-Tap IIR DFE
Power	66 mW	230 mW @ 56G	136 mW	176.3 mW	360 mW @ 56G	48mW	259 mW
Power Efficiency	1.1 pJ/b*	4.11 pJ/b	2.26 pJ/b	5.5 pJ/b	6.4 pJ/b	0.92 pJ/b	4.63 pJ/b
Channel	8.2 dB @ 15G	10 dB @ 14G	21 dB @ 30G	13.5 dB @ 8G	7.4 dB @ 14G	7.3 dB @ 13G	20.8 dB @ 14G

\* Only data-path power consumption, without CDR or adaptation loops.

Table 4.2. Performance summary and comparison of wireline receivers.

banks. The total chip area measures  $900 \mu\text{m} \times 745 \mu\text{m}$ , or 0.67 mm squared. The power consumption of the receiver data-path circuitries together with its breakdown is shown in Fig. 4.24(b). At 60 Gb/s, the two stages of CTLE consume 13 mW, the two half-rate summers consume 13.4 mW, the slicers and latches consume 12.6 mW, and the CKBUF and data buffers consume 27 mW. In total, 66 mW is consumed by the receiver data-path, and 1.1-pJ/bit power efficiency is achieved at 60 Gb/s.

#### 4.6 Summary

A CMOS track-and-regenerate slicer is proposed and designed with the aims to improve the power efficiency, output swing, technological scalability over the conventional CML slicer, and to improve the clock-to-Q delay and output swing over the conventional StrongArm slicer as well. A PAM4 receiver, employing the proposed CMOS track-and-regenerate slicer, benefits from the relaxed settling time constraint thanks to the reduced slicer delay, and from the direct availability of rail-to-rail digital signals offered by the proposed slicer. The prototype fabricated in 28-nm CMOS technology achieves power efficiency of 1.1 pJ/bit at 60 Gb/s over a channel with 8.2-dB loss at Nyquist, demonstrating an energy-efficient PAM4-DFE design. The performance and comparisons with the state of the art are summarized in Table 4.2.

## ENERGY-EFFICIENT NEURAL-NETWORK-ENHANCED FFE FOR PAM4 ADC-BASED OPTICAL INTERCONNECTS

### 5.1 Overview

Optical interconnects promise great potential to support the rapidly growing data traffic in data centers, thanks to the relatively low data-rate-dependent losses of the fibers. In addition, it is feasible to increase the data rate by utilizing higher-order modulation formats, attributed to the augmented spectral efficiency. Four-level pulse amplitude modulation (PAM4) has become a particularly appealing option since its Nyquist frequency is only half of that of non-return-to-zero (NRZ) modulation. However, because of dividing the peak transmitted signal swings into multiple levels, the corresponding PAM4 receivers need to manage input data with a more stringent signal-to-noise ratio (SNR). Moreover, the presence of nonlinearities arising from optical modulators can cause mismatched level separations, which can effectively worsen the SNR and result in significant negative impact on the symbol-error-rate (SER) performance. In other words, nonlinearity equalization plays a critical role in realizing the benefits offered by the optical interconnects adopting higher-order modulation formats. Compared to analog equalizers, digital signal processing (DSP)-based equalizers hold superior immunity against process, voltage, and temperature (PVT) variations. Furthermore, considering on-chip implementations, the CMOS technology scaling favors digital circuits, and a DSP-based digital equalizer is considered as a versatile solution applicable to various CMOS processes through computer-aided syntheses. In light of these, this chapter aims to demonstrate the effectiveness of digital nonlinear equalizers, while presenting the efforts to reduce the hardware complexity and thus improve the area and energy efficiency of the nonlinear equalizers.

Volterra series empowers the characterization of a nonlinear system [61]. The detailed mathematical expressions for Volterra series will be given in the next section, while the concept of Volterra series can be concisely illustrated with Fig. 5.1. That is, the  $n$ -th output

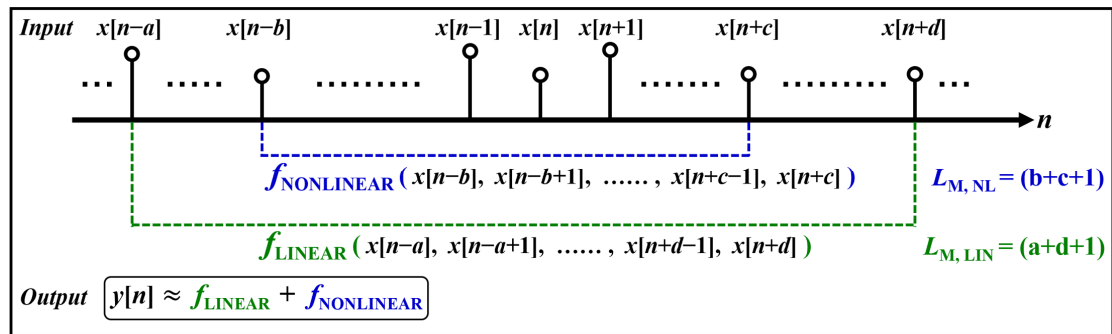


Fig. 5.1. Description of an output sample ( $y[n]$ ) with functions of input ( $x$ ) samples lying within certain time-spans.

sample,  $y[n]$ , can be (approximately) described with linear and nonlinear functions of input samples lying within certain time-spans. This time-span is commonly referred to as memory length ( $L_M$ ). In the example shown in Fig. 5.1, the memory lengths of the linear and nonlinear functions are denoted by  $L_{M,LIN}$ , and  $L_{M,NL}$ , which are equal to  $(a+d+1)$  and  $(b+c+1)$ , respectively. In general,  $a$ ,  $b$ ,  $c$ , and  $d$  assume nonnegative integers, and they do not necessarily match their relative positions drawn in Fig. 5.1 as an example. Based on the idea of Volterra series, Volterra equalizers have been prevalently employed to equalize the undesirable nonlinearity [61–64]. Nonetheless, since the operation relies on multiplicatively generating high-order terms along with further multiplications by their corresponding coefficients, the hardware and power overheads of Volterra equalizers are often of great concern, considering the increased number of required multipliers. Prior arts have therefore proposed techniques to simplify the implementation of a Volterra equalizer such that less multipliers are utilized. In [65], a pruning algorithm is applied to reduce the number of multipliers and the associated computational effort by dropping the terms with relatively insignificant coefficients. In [66, 67], architectural innovations are presented, in which piecewise linear (PWL) functions are leveraged in the nonlinearity compensation, making parts of the conventional Volterra equalizer as well as the corresponding multipliers dispensable. While these PWL equalizers can succeed in reducing the hardware complexity, their advantages diminish when a short memory length is of interest. In this work, we propose solutions for cases with short nonlinearity memory lengths and

demonstrate their practical applications in micro-ring modulator (MRM)-based optical interconnects. We first identify that a PWL equalizer can be viewed as a shallow nonlinear neural network, and hence open-source machine learning libraries are engaged to facilitate the design and adaptation of the (neural-network-alike) PWL equalizers. Afterward, we propose a series of custom neural-network-enhanced feedforward equalizer (NN-FFE) as area/energy-efficient alternatives to the Volterra equalizers with short memory lengths. The effectiveness and power overheads of the NN-FFEs in equalizing the nonlinearities originating from MRMs are examined together with their Volterra equalizer counterparts in different scenarios. Despite that MRM-based interconnects form the focus of the case studies, the design framework described in this chapter extends to different types of optical modulators, and the noise analysis techniques applied in this work lend themselves to exploring other architectures of nonlinear equalizers.

This paper is organized as follows. Section 5.2 describes the overview of MRM-based PAM4 interconnects and the quantitative characterization of the MRMs. Section 5.3 first explains the principle of nonlinear equalization employing PWL functions, and then a custom learnable PWL activation function is proposed. Detailed noise analysis and numerical comparisons to the Volterra counterpart are also included. Section 5.4 presents the complete custom neural-network-enhanced FFE (NN-FFE) and the extended techniques used in noise analysis and simulations. Section 5.5 compares the hardware complexity along with the effectiveness of different nonlinear equalizers, by carrying out SER simulations with a variety of MRM and equalizer designs substituted in the system. In Section 5.6, the overall design framework is presented and the power comparisons of the equalizers using standard-cell libraries of a commercial 28-nm CMOS technology, are reported. Finally, summary is drawn in Section 5.7.

## **5.2 MRM-Based PAM4 Interconnects**

MRM has been considered as a promising component to realize improvements in both power efficiency and aggregate channel capacity [68–73]. The relatively compact device size of an MRM results in small device capacitance, and its high modulation efficiency

allows achieving high extinction ratio with a low voltage drive [70]. These benefits enable lower power consumption of the driver circuitry and high-speed operations at the same time. Additionally, the wavelength selectivity of an MRM suggests outstanding compatibility to a wavelength-division multiplexing (WDM) system [68, 72]. In light of these advantages, MRM-based PAM4 optical interconnects are investigated in this work. The system overview is depicted in Fig. 5.2, in which the MRM in use is driven by a linear PAM4 driver and the resulting nonlinear distortion is then equalized by a PVT-tolerant DSP core. While nonlinearity mitigation schemes at the transmitter side have been proposed in prior arts [6, 74], these techniques rely on combining the outputs of a large number of segments in analog fashion. In [6], the electrical driver is segmented into 30 slices, essentially functioning as an electrical digital-to-analog converter (DAC). In [74], an optical DAC is realized by dividing an MRM device into 16 segments. A highly segmented electrical DAC would encounter challenges in maintaining the bandwidth, attributed to the increased number of electrical connections. On the other hand, a highly segmented optical DAC would cost increased overheads in the area and pin counts. Furthermore, mismatches among the electrical/optical DAC segments, introduced by PVT variations, can deteriorate the linearity and thus the performance of the DACs. Moreover, in order to calibrate and/or adaptively optimize the setting of the transmitter-side DAC, it would require a back-channel from the receiver or an extra signal path invested in the transmitter for monitoring the status of applied pre-distortion/pre-emphasis. Enabled by the receiver-side digital equalization, the robustness of equalization is improved, and the complexity of the transmitter is significantly reduced. In Fig. 5.2, each module is described in MATLAB with scripts or Simulink models, including a compact model for MRMs composed of PN-junction-based optical phase shifters reported in [75]. This MRM model takes the device characteristics (e.g., junction geometries, doping levels, intrinsic effective refractive index and absorption coefficient) along with the applied voltage as input parameters to quantify the varied effective refractive index and absorption coefficient. Accordingly, the static and dynamic mechanisms of an MRM can be modeled for given transmission and coupling coefficients. This MRM model not only offers satisfactory

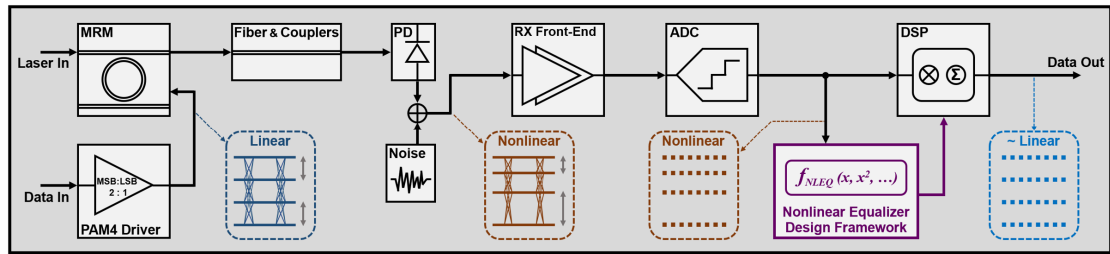


Fig. 5.2. System overview of MRM-based PAM4 optical interconnects.

predictions compared with the measured data from various silicon photonic platforms [75], but also facilitates the studies on the characteristics and applications of distinct MRMs constructed with various design parameters. In particular, the MRM nonlinearities and inter-symbol interference (ISI) are examined, as described in the next two subsections.

### 5.2.1 MRM Nonlinear Distortion and Bandwidth

The modulation of an optical MRM's output power originates from the changes in the refractive index and the absorption coefficient, introduced by electrical drivers varying the carrier densities within the waveguides. As described in [75, 76], this electro-optic modulation is generally nonlinear, and hence without any pre-distortion, the resulting MRM outputs are expected to bear mismatched levels. On the other hand, the speed of the modulation is also crucial to an MRM design, since pertinent trade-offs are involved and linked through the quality factor (Q) of an MRM. With a larger value of Q, an MRM offers greater optical modulation amplitude (OMA) at low modulation frequencies for a given driver voltage swing, whereas a high-Q MRM implies long photon lifetime, leading to low optical bandwidth. These key properties are well captured by large-signal transient simulations with the MRM model in [75]. As examples, Fig. 5.3(a) and Fig. 5.3(b) show the PAM4 eye diagrams of two different MRM designs of  $Q = 7820$ , and  $Q = 15640$ , respectively, where they are driven by an identical electrical driver with 2-V peak-to-peak swing and 50-GHz  $-3$ -dB bandwidth. This PAM4 driver assumes a simple linear architecture, in which no pre-distortion is applied, and the ratio of the most significant bit (MSB) to the least significant bit (LSB) is 2:1. Following the approach in [75], the laser wavelength or the MRM resonance wavelength is adjusted; the doping levels and the

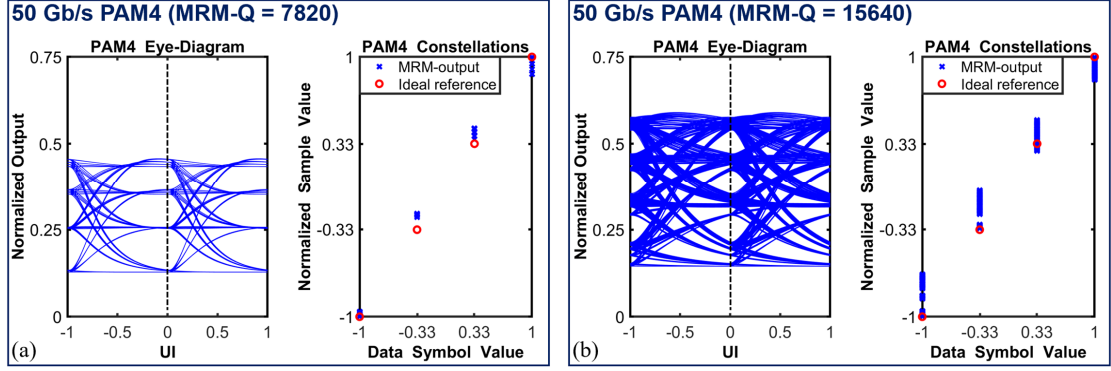


Fig. 5.3. (a) PAM4 eye-diagram and signal constellations simulated with MRM of  $Q = 7820$  at 50 Gb/s. (b) PAM4 eye-diagram and signal constellations simulated with MRM of  $Q = 15640$  at 50 Gb/s.

coupling coefficients are optimized such that the largest outer OMA is achieved for a specified value of  $Q$ . With fixed normalized input laser power, Fig. 5.3 confirms that increasing the  $Q$  of an MRM does enlarge the outer OMA at the price of bandwidth degradation. The corresponding constellation plots are shown next to the eye-diagrams, where the data points are obtained from the optimal discrete-time sampling that maximizes the vertical eye-openings and thus the signal margins. These constellation points display mismatched levels and the presence of ISI; that is to say, the information of nonlinearity and speed limitations are contained in the constellations. Aiming to extract and further quantify the underlying nonlinear distortion together with the ISI due to limited bandwidth, Volterra series fitting serves as a simple yet effective tool.

### 5.2.2 Volterra Series Fitting

A conventional discrete-time  $p$ -th order Volterra series can be expressed as [77]:

$$\begin{aligned}
 x(n) = & \sum_{k_1=0}^{m_1-1} h_1(k_1) D(n - k_1) + \sum_{k_1=0}^{m_2-1} \sum_{k_2=0}^{k_1} h_2(k_1, k_2) \prod_{j=1}^2 D(n - k_j) \\
 & + \sum_{k_1=0}^{m_3-1} \sum_{k_2=0}^{k_1} \sum_{k_3=0}^{k_2} h_3(k_1, k_2, k_3) \prod_{j=1}^3 D(n - k_j) + \dots \\
 & + \sum_{k_1=0}^{m_p-1} \sum_{k_2=0}^{k_1} \dots \sum_{k_p=0}^{k_{p-1}} h_p(k_1, k_2, \dots, k_p) \prod_{j=1}^p D(n - k_j)
 \end{aligned} \quad (5.1)$$

where  $x(n)$  and  $D(n)$  can be interpreted as the  $n$ -th sample of an output signal and the  $n$ -th sample of an input signal, respectively;  $m_p$  is the memory length of the  $p$ -th order terms, and  $h_p$  is commonly referred to as the  $p$ -th order Volterra kernel, consisting of the coefficients of  $p$ -th order terms. By treating  $D(n)$  in (5.1) as the  $n$ -th PAM4 data symbol;  $x(n)$  in (5.1) as the  $n$ -th sample at a channel output and allowing negative starting indices, the resulting fitted Volterra kernels reveal the channel characteristics. Specifically, the ISI and the bandwidth limitation can be examined by looking at the dependence of  $x(n)$  on the data symbols other than  $D(n)$ ; for example, a low-pass-filtered channel tends to cause higher causality with the previously transmitted symbols. Besides, nonzero high-order Volterra kernels point to the presence of nonlinearities, and the magnitudes of these coefficients reflect the severity of the nonlinear distortion. To illustrate these ideas, we consider the cases where an MRM of  $Q = 7820$ , which has a full-width-at-half-maximum (FWHM) bandwidth of 25 GHz, is used for 50-Gb/s and 100-Gb/s PAM4 data transmission. For better modeling the realistic link configuration, the MRM outputs are further filtered by cascaded first-order low-pass filters representing the photodiode (PD) and the receiver analog front-end (AFE), where their  $-3$ -dB bandwidths are set to 40 GHz, and 20 GHz, respectively. The resultant fitting coefficients using a third-order Volterra series are shown in Fig. 5.4, where only the ten most significant coefficients, excluding the main cursor, are shown for clarity and all coefficients are normalized with respect to the main-cursor value. For simplicity,  $D(n-k)$  is denoted by  $D_{-k}$ . Volterra series fitting offers two important capabilities. First, it eases the subsequent equalizer training in that it allows the use of a relatively small number of coefficients to reconstruct the output samples, or the channel response in a broader sense. Second, it helps to determine the design objectives of a nonlinear equalizer by indicating the most salient nonlinear terms. As shown in Fig. 5.4, the primary task of a nonlinear equalizer for MRM-based links is tackling the second-order products of the current data symbol ( $D_0$ ) and the previous data symbol ( $D_{-1}$ ), especially for the case that is not limited by the channel bandwidth, e.g., Fig. 5.4(a). In other words, compensation for the nonlinear  $D_0^2$ ,  $D_{-1}^2$ , and  $(D_0 D_{-1})$  terms should be the focus. In the next section, the principle and features of nonlinear compensation with a Volterra equalizer or an equalizer based on a PWL function are discussed in detail.



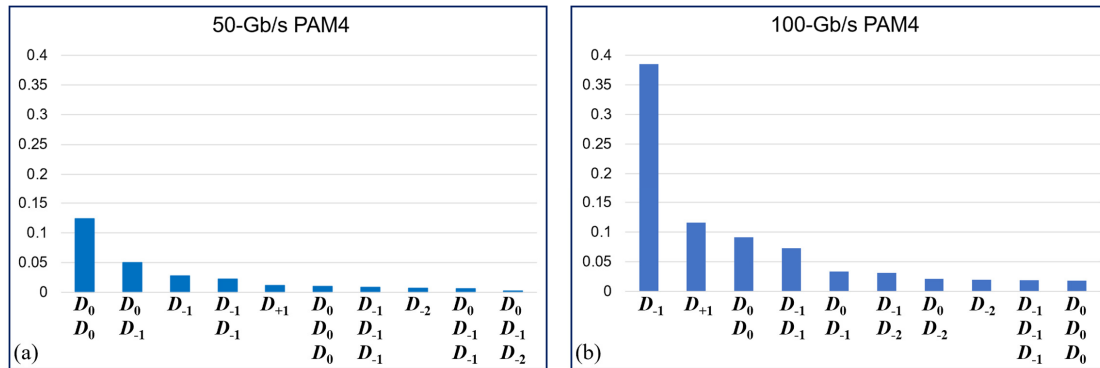


Fig. 5.4. (a) Volterra series fitting example at 50 Gb/s. (b) Volterra series fitting example at 100 Gb/s.

### 5.3 Principle and Noise Analysis

#### 5.3.1 Overview of Activation Functions

The general architectures of a conventional linear FFE and a feedforward neuron (i.e., one-layer neural network) are depicted in Fig. 5.5(a) and Fig. 5.5(b), respectively. A linear FFE can be viewed as a subset of feedforward neural networks, which has purely linear activation function and is therefore limited to performing linear operation. By contrast, with the inclusion of nonlinear activation functions such as the commonly used sigmoid function [78] and the hyperbolic tangent (tanh) function [79], more sophisticated tasks can be accomplished by a so-called neural network. In other words, the selection of the activation function in use is very critical to the neural network performance. However, activation functions like sigmoid and tanh considerably complicate the hardware implementations since they involve the computations or approximations of exponential functions. In the pursuit of better area/energy efficiency, PWL activation functions become appealing candidates to undertake nonlinear operations with reduced implementation complexity. For instance, the popular rectified linear unit (ReLU) [80] can be realized with simple logics and circuits.

#### 5.3.2 Custom Learnable PWL Activation Function

To visualize that PWL functions are capable of performing nonlinear operations, Fig. 5.6 is presented. It shows how the superposition of shifted-ReLU (SReLU) functions can

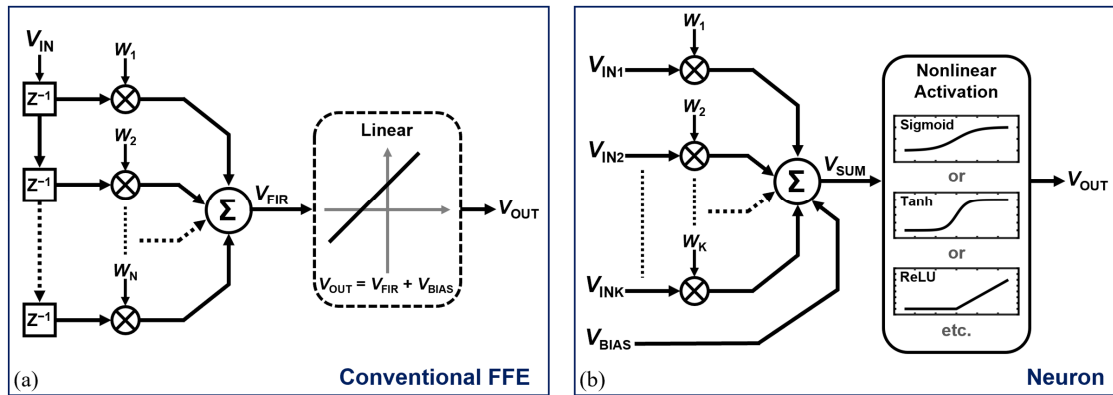


Fig. 5.5. (a) Conventional linear FFE. (b) Feedforward neuron (one-layer neural network).

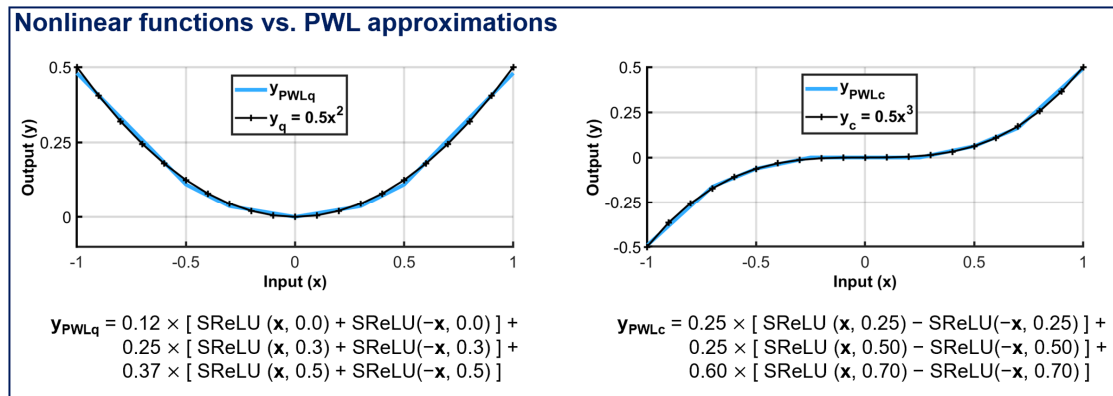


Fig. 5.6. Approximations of nonlinear functions using superpositions of PWL functions.

approximate high-order polynomial functions. The definition of a SReLU function is given below, where  $x$  denotes the input and  $k$  is referred to as the breakpoint. When  $k$  is set to 0, it corresponds to the well-known ReLU function.

$$\text{SReLU}(x, k) = \begin{cases} x - k & , \text{ if } x > k \\ 0 & , \text{ if } x \leq k \end{cases} \quad (5.2)$$

A few observations can be drawn. First, the overall activation function formed by the superposition has the shape which resembles that of the target polynomial function. Second,

the locations or values of the breakpoints have notable impact on the goodness of the approximation, and these breakpoints are not fixed in general. Third, it can be inferred that depending on the tolerance of discrepancy as well as the range of interest, more SReLU elements can be added to, or some of the existing SReLU elements can be removed from the superposition. The foregoing motivates the design of a custom activation function subject to two considerations. For one thing, the complexity of the overall activation function can be reduced as long as its input-output mapping characteristics are still satisfactory with respect to the application. More specifically, since PAM4 format is focused, it suggests that the most relevant mappings lie in the four constellations, relaxing the need of tight approximations spanning a wide range. For the other thing, in view of that the most prominent nonlinearities are associated with quadratic terms as shown in Section 5.2, the overall activation function should roughly act like a quadratic function. Accordingly, we design a custom low-complexity activation function, full-wave rectified linear unit (FReLU), which follows the naming of a full-wave electrical rectifier circuitry that maps both positive and negative inputs to positive outputs. The mathematic definition of FReLU is given as:

$$\text{FReLU}(x, p, q) = \begin{cases} x - p, & x > p \\ 0, & q \leq x \leq p \\ -x + q, & x < q \end{cases} \quad (5.3)$$

where  $x$  denotes the input to the FReLU,  $p$  and  $q$  are learnable parameters, and  $p$  is not smaller than  $q$ . To demonstrate the operation of a neural-network-alike PWL equalizer, the next two subsections are dedicated to illustrating the use of FReLU for PAM4 nonlinear equalization and comparing it with the case of using the quadratic function (i.e., a conventional second-order Volterra equalizer with memory length one). We start with the noise analysis and computation for these nonlinear functions.

### 5.3.3 Level-Dependent Noise Analysis

While a nonlinear equalizer makes efforts to improve the signal margin by compensating for the nonlinearities, one common side effect of nonlinear equalizers is the introduction of level-dependent noise [81]. Therefore, in order to accurately evaluate the performance of a nonlinear equalizer, noise analysis and simulations that address the nonlinear or level-

dependent noise distributions need to be developed. In particular, the computations of statistical cumulative distribution function (CDF) and complementary cumulative distribution function (CCDF) of the level-dependent noise are of greatest interest, since they are determining indications of the SER performance. The fundamental of level-dependent noise in a nonlinear equalizer is attributed to the interaction with the generated nonlinear (i.e., high-order) terms, which can be understood from (5.4) given below:

$$\text{Var} [X^2] = \text{Var} [(X_S + X_N)^2] = \text{Var} [X_S^2 + X_N^2 + 2X_S X_N] \quad (5.4)$$

where  $X$  denotes the sample value and its signal component and noise component are denoted by  $X_S$  and  $X_N$ , respectively. The rightmost term in (5.4),  $(2X_S X_N)$ , explains that the output variance (i.e., noise power) is dependent on the signal level, when the sample values are processed through a quadratic (i.e., nonlinear) function. Furthermore, while  $X_N$  in most cases assumes Gaussian distribution, the existence of  $(X_N^2)$  term in (5.4) makes the overall noise distribution deviate from a typical Gaussian distribution. The following expressions are thus presented in order to illustrate methods for addressing the nonlinearly filtered noise as in (5.4).

$$Y_N = X_N^2 + 2X_S X_N \quad (5.5)$$

$$I_S(x) = \begin{cases} 1, & \text{if } x \in S \\ 0, & \text{if } x \notin S \end{cases} \quad (5.6)$$

$$F_Y(y) = P_b \{ Y_N \leq y \} \quad (5.7)$$

$$F_Y(y) \equiv P_b \{ S \} \quad (5.8)$$

$$F_Y(y) = P_b \{ x_L \leq X_N \leq x_R \} \quad (5.9)$$

$$F_Y(y) = F_X(x_R) - F_X(x_L) \quad (5.10)$$

$$F_Y(y) = E [ I_S(X_N) ] \quad (5.11)$$

$$F_Y(y) = \int_{-\infty}^{+\infty} I_S(x) f_X(x) dx \quad (5.12)$$

The total noise component in (5.4) is denoted by  $Y_N$  as expressed in (5.5), which is a nonlinear function of Gaussian random variable  $X_N$ . The definition of an indicator function  $I_S$  with respect to an event/subset  $S$  is given as (5.6). If the outcome of a random variable, denoted by  $x$ , belongs to  $S$ , then  $I_S$  takes the value 1; otherwise,  $I_S$  takes the value 0. More intuitively speaking, the indicator function  $I_S$  indicates whether the event  $S$  happens or not. The probability of interest here is mathematically described in (5.7) and (5.8); that is, the CDF of the nonlinearly filtered noise,  $Y_N$ . This CDF of  $Y_N$  is denoted by  $F_Y$ ; the corresponding event,  $(Y_N \leq y)$ , is represented by  $S$ , and  $P_b\{ \}$  stands for the probability of the event in the brackets. In this relatively simple case defined by (5.5) where it assumes  $X_S$  (i.e., the signal level) is known and fixed, it can be solved that the event  $(Y_N \leq y)$  is equivalent to the event  $(x_L \leq X_N \leq x_R)$  by substituting (5.5) in  $(Y_N \leq y)$  and solving  $(X_N^2 + 2X_S X_N = y)$  with  $x_L$  and  $x_R$  respectively denoting the smaller and larger root. In other words, computing (5.7) is the same as computing (5.9), and accordingly the knowledge of a Gaussian random variable, whose CDF and probability density function (PDF) are well defined with its mean and variance, can be leveraged. By denoting the CDF of  $X_N$  by  $F_X$ , we derive and illustrate the analytical CDF computation of a nonlinearly filtered Gaussian random variable,  $F_Y$ , using the original Gaussian CDF,  $F_X$ , as shown in (5.10). On the other hand, from probability theory, the expectation (i.e., expected value) of an indicator is the probability of indicator's associated event. As a result, we have (5.11) as another equivalence of (5.7), in which  $E[ \ ]$  gives the expected value of the random variable in the square brackets. The computation of (5.11) involves the integral of  $(I_S f_X)$ , where  $f_X$  denotes the PDF of  $X_N$ , as shown in (5.12). This integral can be computed by analytically finding the range in which  $I_S$  equals to 1, and then accordingly revising the upper limit and lower limit of the integral. It comes as no surprise that the upper limit and lower limit of the integral in this case are simply the aforementioned  $x_R$  and  $x_L$ , respectively, and hence this integral becomes identical to (5.9) and (5.10) as well.

Nevertheless, it is noteworthy that with the introduction of the indicator function  $I_S$ , the

expression (5.12) can be numerically computed (e.g., using MATLAB) by integrating over a sufficiently large range without explicitly specifying the integral limits. As will be shown in the next subsection, the numerical integral results display strong agreement with the analytical solutions. This numerical-integral-based method is particularly valuable when the nonlinear characteristic appears to be much more complicated than (5.5). For example, when the overall noise arises from multiple correlated Gaussian random variables, it is not trivial to find the boundaries (i.e., all the integral limits) and derive the analytical CDF. By contrast, the CDF can be well approximated by carrying out numerical integrals with respect to a multivariate Gaussian PDF. The foregoing concepts also hold in the cases of PWL functions. By incorporating the PWL characteristics into the indicator function, the CDF of the noise filtered by a PWL function can also be computed using (5.12). In view of the convenience, this numerical-integral-based method is an efficient tool for computing CDFs and CCDFs, favorably assisting the design and evaluation of diverse nonlinear equalizers.

### 5.3.4 Numerical Examples and Comparisons

In this subsection, we demonstrate the utilization and features of the custom FReLU-based equalizer, compared with those of its Volterra equalizer counterpart. The main purpose is to offer useful insights with a numerically concise example.

A simple nonlinear response is considered and modeled as:

$$X_K = D_K + 0.2D_K^2 + n_K \quad (5.13)$$

where  $D_K$  is the  $K$ -th PAM4 data symbol value, taking a value in the set  $\{\pm 1/3, \pm 1\}$ ;  $X_K$  is the  $K$ -th received sample value, and  $n_K$  is the noise component assuming Gaussian distribution in  $X_K$ . Seeing that the nonlinearity in  $X_K$  is dominated by the current data symbol, it suggests the employment of either a second-order Volterra equalizer with memory length ( $L_M$ ) 1, or the designed FReLU-based PWL equalizer, to compensate the unwanted nonlinearity. The expressions for these nonlinear equalizer outputs are given as:

$$Y_K = a_1 X_K + a_2 X_K^2 + c \quad (5.14)$$

$$Y_K' = a_1' X_K + a_2' \times \text{FReLU}(X_K, p, q) + c' \quad (5.15)$$

where the  $Y_K$  in (5.14) is the  $K$ -th output sample of the Volterra equalizer, and the  $Y_K'$  in (5.15) is the  $K$ -th output sample equalized by FReLU. The equalizer coefficients, i.e.,  $a_1$ ,  $a_2$ ,  $c$ ,  $a_1'$ ,  $a_2'$ , and  $c'$  along with the FReLU parameters, i.e.,  $p$  and  $q$ , are trained by an optimizer which minimizes the mean squared error (MSE) between  $Y_K/Y_K'$  and the target data value  $D_K$ . PyTorch, an open-source development tool for machine learning, is equipped with built-in optimizers and libraries for creating neural-network-alike learnable PWL functions. Accordingly, the process of finding the optimal coefficients and parameters that minimize MSE, or the so-called equalizer training, can be efficiently accomplished utilizing PyTorch. The results after the nonlinear equalization are summarized in Table 5.1, where the trained equalizer coefficients corresponding to (5.14) and (5.15) are given as  $a_1 = 1.072$ ,  $a_2 = 0.1984$ ,  $c = 0.0052$ ,  $a_1' = 1.00$ ,  $a_2' = 0.3636$ ,  $p = 0.3111$ ,  $q = 0.7111$ , and  $c' = 0.0222$ . In the Volterra case, the equalized signal levels bear larger errors compared to those equalized by FReLU, resulting in one larger eye-opening and two smaller eye-openings. These larger errors in the Volterra case originate from the  $X_K^2 = (D_K + 0.2D_K^2)^2$  term, in which other nonlinear terms are produced and added to the Volterra equalizer output. Level-dependent noise can be observed by computing the root-mean-squared (RMS) noise power of the nonlinearly equalized noisy samples at each signal level (L1, L2, L3, L4) as shown in Table 5.1, where a total of 1E8 samples, whose SNR is set to 5 in voltage and independent of signal level prior to equalization, were simulated. To further characterize the level-dependent noise, the CDF at each level is computed for both Volterra and FReLU equalizers, using different methods. The first method is applying cumulative histogram, which records the noise values from transient simulation data and accordingly classifies those into corresponding bins. From the cumulative histogram, the CDF can be computed. Alternatively, as discussed in Section 5.3.3, the statistical methods based on either analytical or numerical integral are devoted

Architecture	2nd-order Volterra ( $L_M = 1$ )	PWL (FReLU)
Target nonlinear channel	$X_K = D_K - 0.2D_K^2 + n_K$	
SNR at equalizer input	$\text{SNR} = [(1/3) / \text{RMS}(n_K)] = [(1/3) / (0.0667)] = 5$	
Trained equalizer	$Y_K = 1.072X_K + 0.1984X_K^2 + 0.0052$	$Y_K = X_K + 0.3636 \times \text{FReLU}(X_K, 0.3111, 0.7111) + 0.0222$
L4: Signal level ( $D_K = +1, n_K = 0$ )	$Y_K = +0.9898$	$Y_K = +1.0000$
L3: Signal level ( $D_K = +1/3, n_K = 0$ )	$Y_K = +0.3579$	$Y_K = +0.3333$
L2: Signal level ( $D_K = -1/3, n_K = 0$ )	$Y_K = -0.3509$	$Y_K = -0.3334$
L1: Signal level ( $D_K = -1, n_K = 0$ )	$Y_K = -0.9955$	$Y_K = -1.0000$
Eye-openings [upper, middle, lower]	[0.6319, 0.7088, 0.6446]	[0.6667, 0.6667, 0.6666]
L4: RMS noise ( $D_K = +1$ )	0.0926	0.0909
L3: RMS noise ( $D_K = +1/3$ )	0.0797	0.0791
L2: RMS noise ( $D_K = -1/3$ )	0.0621	0.0667
L1: RMS noise ( $D_K = -1$ )	0.0398	0.0425

Table 5.1. Results of nonlinear equalization examples.

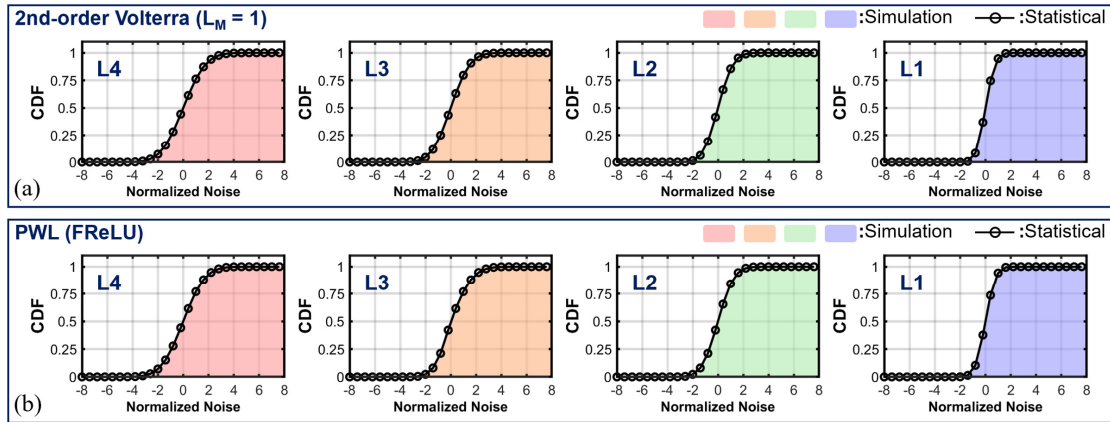


Fig. 5.7. Level-dependent noise CDFs associated with nonlinear equalization. (a) 2nd-order Volterra. (b) FReLU.

to CDF computations. Fig. 5.7(a) and Fig. 5.7(b) show the CDF plots of the nonlinearly filtered noise at the four signal levels in the case of Volterra, and FReLU, respectively. For every signal level, the CDF plots of both Volterra and FReLU cases are similar to each other. However, for distinct signal levels, the CDF plots possess different shapes. For example, the CDF plots of L1 are steeper at the point CDF = 0.5, compared to those of L4. These observations well reflect the level-dependent RMS-noise values in Table 5.1. Moreover, Fig. 5.7(a) and Fig. 5.7(b) both display the strong agreement between the CDF from transient simulation data and the statistical CDF. To further illustrate the reliability



of statistical CDF methods, the discrepancies at lower probabilities, between the statistical CDF and the CDF from transient simulations of  $1E8$  samples, are plotted in Fig. 5.8. Meanwhile, Fig. 5.9 confirms that by setting the relative error tolerance (Tol.) used in numerical integrals sufficiently small, the numerical-integral CDF and analytical CDF converge to an approximately identical statistical CDF, even for very low probabilities such as  $1E-15$ . While the transient-simulation-based method captures the realistic operation of equalizers by directly examining the noise components in all equalized output samples, it requires lengthy simulations to accurately characterize an event occurring with low probability. By contrast, with the requirements (e.g., large computer memory and long simulation time) for conducting very lengthy transient simulations significantly relaxed, the statistical methods are of great use.

This simple case study not only illustrates the principle of applying FReLU to equalize second-order nonlinearity, but also offers an informative takeaway. That is, the FReLU holds the potential to compensate for second-order nonlinearities with reduced hardware complexity. Specifically, unlike Volterra equalizer, which requires one multiplier for generating the  $X_K^2$  term, the FReLU costs two adders instead, taking advantage of the PWL approximations.

## 5.4 Neural-Network-Enhanced FFE

### 5.4.1 Architecture

In Section 5.3, we showed the effectiveness and benefits of employing an FReLU in equalizing the PAM4 nonlinearity proportional to the current data symbol squared (i.e.,  $D_0^2$ ). However, we also learned from Section 5.2 that in addition to mitigating the  $D_0^2$  term, the simultaneous compensation for both  $(D_0 D_{-1})$  and  $D_{-1}^2$  terms is necessary to further improve the SNR. Accordingly, this suggests extending the memory length of the nonlinear equalizers. One conventional solution is shown in Fig. 5.10(a), where a second-order Volterra equalizer of memory length two is implemented in parallel with a linear FFE. The linear FFE assumes five taps, which is sufficient to equalize the linear part of ISI in this work. Only one slice that is responsible for equalizing the sample  $X_K$  is drawn for better

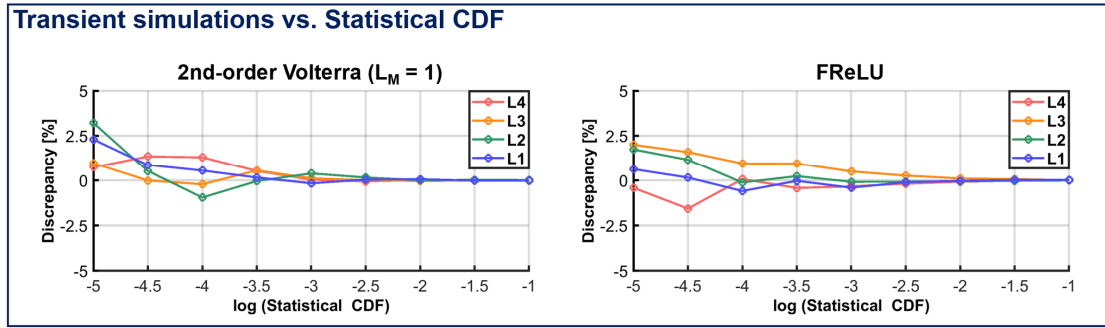


Fig. 5.8. Discrepancy between transient-simulation-based CDF and statistical CDF.

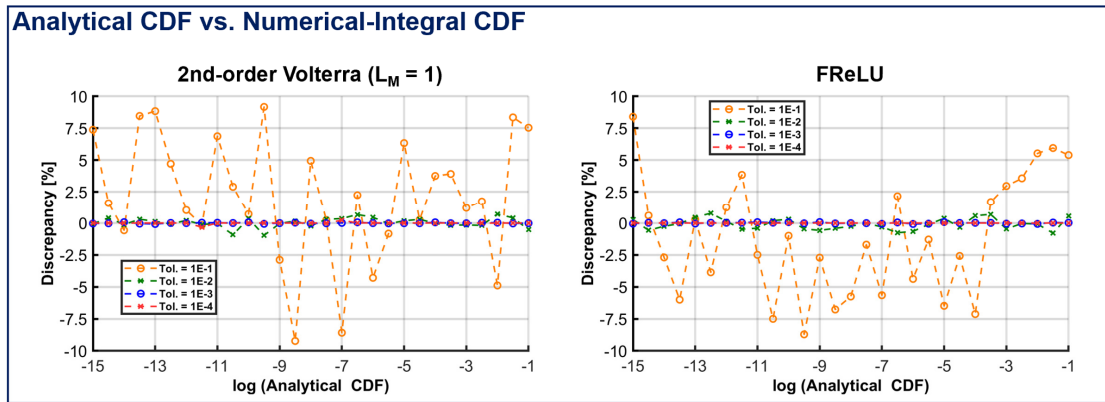


Fig. 5.9. Discrepancy between analytical CDF and numerical-integral CDF.

clarity. As a counterpart to the conventional Volterra equalizer in Fig. 5.10(a), the proposed alternative is presented in Fig. 5.10(b), extended upon the custom activation function, FReLU. The activation function outputs are scaled and added to the linear FFE output, forming a neural-network-enhanced FFE (NN-FFE) and empowering its capability of nonlinearity equalization. The design intuition can be perceived as what follows. The linear FFE output, labeled as  $L_K'$  in Fig. 5.10(b), extracts the data information of  $(D_K + \gamma D_{K-1})$  from the received samples (i.e.,  $X_{K+1}, X_K, \dots, X_{K-3}$ ), where  $D_i$  and  $X_i$  are the  $i$ -th data symbol value and the  $i$ -th received sample value, respectively. FReLU is then applied on  $L_K'$  to generate  $A_K'$ , which holds the information of  $(D_K + \gamma D_{K-1})^2$ . Furthermore, by taking advantage of reusing the two precedent computed  $L_{K-1}'$  and  $A_{K-1}'$  values, the  $(\gamma D_{K-1})$  and  $(\gamma D_{K-1})^2$  terms appearing in  $L_K'$  and  $A_K'$  can be individually adjusted/compensated through

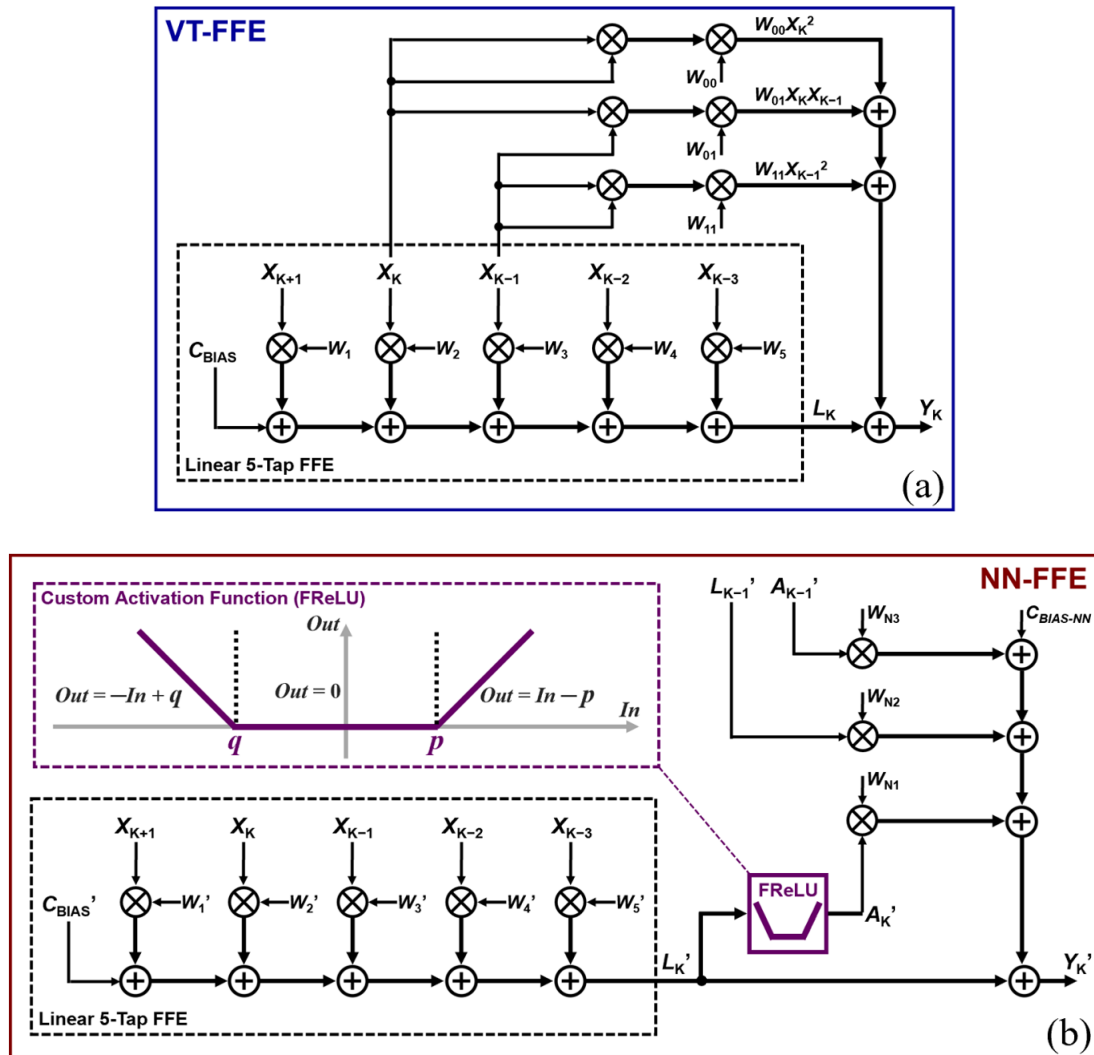


Fig. 5.10. (a) Volterra equalizer of memory length 2 with 5-tap FFE. (b) Custom neural-network-enhanced 5-tap FFE.

one extra scaling multiplier for each. The value of  $\gamma$  depends on the channel or device characteristics, and the breakpoints  $p$  and  $q$  in the custom FReLU are trained to best fit different scenarios. In the following, the equalizers in Fig. 5.10(a) and Fig. 5.10(b) are referred to as VT2-FFE and NN3-FFE, respectively, for the purpose of comparing with their simplified versions, i.e., VT1-FFE (Volterra equalizer with memory length 1, removing the paths associated with  $W_{01}$  and  $W_{11}$ ), NN2-FFE (removing the path associated with  $W_{N3}$ ), and NN1-FFE (removing the paths associated with  $W_{N2}$  and  $W_{N3}$ ).

### 5.4.2 Extended Noise Analysis Techniques

The primary objectives of noise analysis are still the statistical (level-dependent) CDFs and CCDFs, which are directly linked to the SER evaluations. The principles applied in computing the CDFs and CCDFs of a VT-FFE/NN-FFE remain the same as the simpler case presented in Section 5.3. Nonetheless, because of the multivariate nature (i.e., multiple noisy samples involved) in the context of a VT-FFE/NN-FFE, two additional pivotal properties are incorporated to improve the accuracy as well as to reduce the complexity and thus simulation time of the statistical models.

For one thing, the linear transformation of a multivariate Gaussian random vector is to be leveraged. Provided that  $G_0$  is a multivariate Gaussian random vector of size  $(m_0 \times 1)$ , characterized by a  $(m_0 \times 1)$  mean vector  $\mu_0$  and a  $(m_0 \times m_0)$  covariance matrix  $Z_0$ , then a linear transformation of  $G_0$ , defined as  $G_1 = LG_0 + C$ , where  $L$  is a  $(m_1 \times m_0)$  full-rank matrix and  $C$  is a  $(m_1 \times 1)$  vector, leads to another Gaussian random vector  $G_1$  of size  $(m_1 \times 1)$ . Moreover, the mean vector  $\mu_1$  and the covariance matrix  $Z_1$  of  $G_1$  are respectively given as:

$$\mu_1 = L\mu_0 + C \quad (5.16)$$

$$Z_1 = LZ_0L^T \quad (5.17)$$

where  $L^T$  is the transpose of  $L$ . The use of (5.16) and (5.17) directs to the powerful simplification in computing the CDF and CCDF integrals. By merging multiple linearly combined Gaussian random variables into one or fewer numbers of variable, it effectively reduces the dimension of the numerical integrals and hence favors the reduction in computational complexity.

For the other thing, related to the above technique as well, the quantification of the correlations among the noise components lying in different samples is addressed so as to be included in the covariance matrix. The correlations are quantified by the autocorrelation function of the filtered noise. In this work, the receiver AFE is modeled as a first-order

low-pass filter (LPF) with time constant ( $T_{RC}$ ), filtering the white Gaussian noise appearing at the receiver input. As a consequence, the covariance ( $R_V$ ) of noise components spaced by time difference  $\tau$  is given as [82]:

$$R_V(\tau) = \overline{n_T^2} \exp(-|\tau| / T_{RC}) \quad (5.18)$$

where  $\overline{n_T^2}$  is the total integrated noise power. To compute the covariance of noise components (i.e., noise random variables) spaced by  $m$  sampling period ( $T_S$ ) in time, the  $\tau$  in (5.18) is substituted with ( $mT_S$ ).

As can be seen from (5.18), it reflects the insights that the correlations between samples are affected by both the sampling period and AFE bandwidth. Specifically, faster sampling or an AFE of lower bandwidth (i.e., smaller  $\tau = mT_S$ , or larger  $T_{RC}$ , respectively) leads to higher correlation, and vice versa. In other words, when the noise is filtered by a first-order LPF, the common assumption/approximation that the sampled noise components are uncorrelated may no longer be satisfactory, depending on the sampling period and AFE bandwidth. This improvement on the accuracy is achieved by using (5.18), and the simultaneous application of (5.16) and (5.17) can further lead to faster computations of CDF and CCDF integrals.

### 5.5 Link Simulations and SER Performance

Provided that the PAM4 system assumes equiprobable data symbols among the four levels, respectively labeled as  $S_1$ ,  $S_2$ ,  $S_3$ , and  $S_4$ , its SER is given as in [83]:

$$\text{SER} = ( \sum_{a=1}^4 \sum_{b=1, b \neq a}^4 P_{ab} ) / 4 \quad (5.19)$$

where  $P_{ab}$  denotes the probability of detecting the received symbol as  $S_b$  while in reality  $S_a$  was transmitted. The probability  $P_{ab}$  is computed by replacing the complementary error functions (erfc) in [83] with the developed CDF/CCDF integrals for nonlinearly filtered Gaussian noise distributions.

For the purpose of characterizing the effectiveness of each equalizer under different scenarios, distinct combinations of MRM and equalizer designs are substituted in the optical link depicted in Fig. 5.2, and the SER performance of each configuration is evaluated. Simulations at 50-Gb/s PAM4 and 100-Gb/s PAM4 have been carried out, and the results are shown in Fig. 5.11(a), and Fig. 5.11(b), respectively. The former corresponds to the cases where relatively high-Q MRMs are operated at a lower data rate and the uncompensated MRM nonlinearities pose major deterioration in SNR. The latter, on the other hand, represents the cases in which the signal margin is already compromised, even with only the linear part of ISI, due to the bandwidth limitation. The Volterra series fitting examples previously presented in Fig. 5.4(a) and Fig. 5.4(b), aid with visualizing the nonlinearity and bandwidth as the main source of SNR impairment, respectively. In order to fairly study the effects of employing a varied MRM or equalizer, the laser power and the receiver input-referred noise power are kept constant at a given data rate. Besides, since the nonlinear equalizer design is of the greatest interest in this work, it is arranged such that the quantization noise from the ADC has negligible effects on the overall receiver input-referred noise power. This is in accordance with the common practice for optimizing the receiver sensitivity, whose input-referred noise power is dominated by the noise in the receiver front-end circuits and/or the shot noise of the PD. The  $-3$ -dB bandwidths of the MRM driver, PD, and receiver AFE, are fixed and set to 50 GHz, 40 GHz, and 20 GHz in the simulations, respectively. By replacing the MRM or equalizer design, one at a time, the simulated results of SER performance directly reflect the superiority or inferiority of the selection and thus suffice for comparing the equalizer capabilities.

Fig. 5.11(a) shows the 50-Gb/s simulations using MRMs with FWHM of 12.5 GHz and 25 GHz, along with different equalizer architectures. In this nonlinearity-limited scenario, nonlinear equalizers provide significant improvements, especially in the case of 12.5 GHz (i.e., with a higher value of  $Q$ ). The effects of  $\pm 0.01$ -nm resonance shifts are also included in the plots. Since the resonance shift of an MRM can change its OMA, bandwidth, and nonlinearity, all at the same time, the performance of the linear FFE and nonlinear equalizers can therefore be affected differently. As expected, the variations in SER by

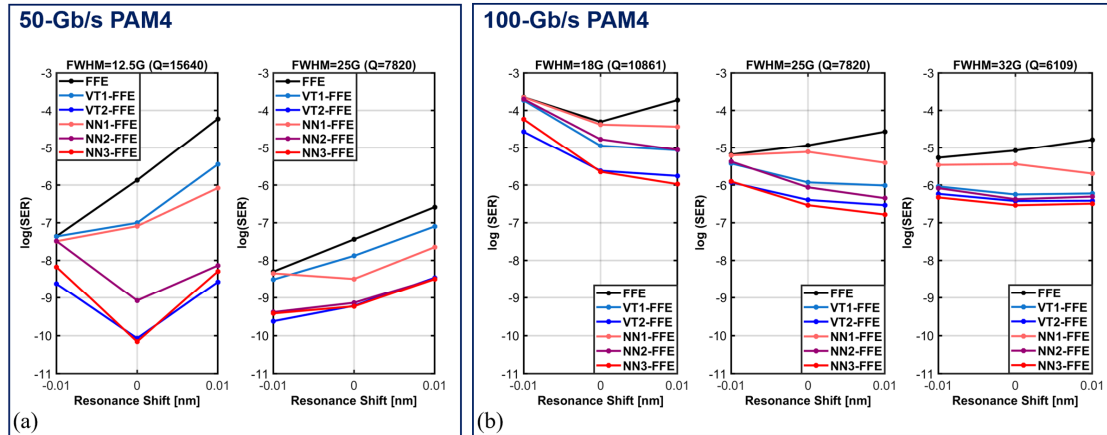


Fig. 5.11. SER simulations with distinct MRM and equalizer designs. (a) At 50-Gb/s PAM4. (b) At 100-Gb/s PAM4.

resonance shifts tend to be more dramatic for an MRM with a higher  $Q$ . It is noteworthy to point out NN2-FFE can outperform VT1-FFE, e.g., by  $\sim 100$  times improvement in SER for the 12.5-GHz-FWHM MRM without resonance shift, while they cost the same number of multipliers. Fig. 5.11(b) shows the 100-Gb/s simulations using MRMs with FWHM of 18 GHz, 25 GHz, and 32 GHz, also with different equalizer architectures. For the 100-Gb/s simulations, the laser power needs to be increased by  $\sim 30\%$  in order to have the SER on the order of  $1E-6$  after equalization. In this bandwidth-limited scenario, the nonlinear equalizers still lead to SER amelioration; however, the improvements are milder than those in the 50-Gb/s cases. While an MRM with higher  $Q$  gives rise to larger static OMA (i.e., low-frequency-modulated OMA), its lower bandwidth eventually degrades the SER. Among all the cases shown in Fig. 5.11, NN2-FFE achieves similar or better performance in comparison with VT1-FFE at the identical expense of multipliers. Moreover, NN3-FFE and VT2-FFE similarly enable the best SER performance, whereas NN3-FFE requires only half of the number of multipliers in VT2-FFE. In other words, either in the nonlinearity-limited or bandwidth-limited scenario, the proposed NN-FFEs serve as attractive alternatives to the Volterra counterparts, i.e., VT1-FFE and VT2-FFE, for area- and energy-efficient interconnects by reducing the number of multipliers. Table 5.2 summarizes the hardware overhead of each equalizer architecture, and the next section further includes the power overhead comparisons through hardware synthesis.

	VT1-FFE ( $L_M = 1$ )	VT2-FFE ( $L_M = 2$ )	NN1-FFE	NN2-FFE	NN3-FFE
Number of (12-bit) Multipliers	2	6	1	2	3
Number of (12-bit) Adders	1	3	4 <sup>a</sup>	5 <sup>a</sup>	6 <sup>a</sup>
Power Overhead / Slice (at 1 GHz, with 28-nm CMOS)	0.618 mW	1.616 mW	0.411 mW	0.760 mW	1.012 mW

<sup>a</sup> including the 2 adders to implement the custom activation function

Table 5.2. Hardware and power overhead comparisons (the proposed NN-FFEs vs. the conventional VT-FFEs).

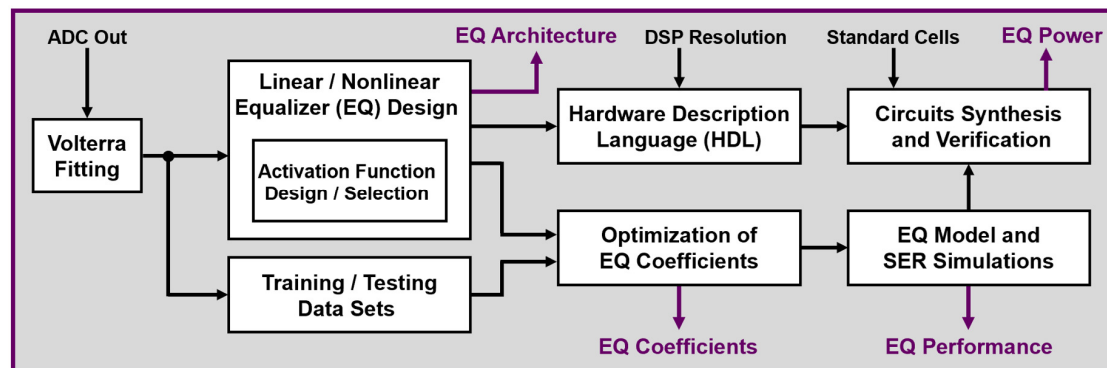


Fig. 5.12. Design framework summary.

## 5.6 Design Framework Summary and Hardware Synthesis

The materials elaborated in the preceding sections suggest a general framework for custom equalizer designs, instead of simply experimenting with various nonlinear equalizers or investing unnecessarily excessive hardware. The design framework used in this work is summarized in Fig. 5.12. Based on the results of Volterra series fitting, we customize the architecture of the equalizer and optimize the associated coefficients and parameters with the help of PyTorch. These trained equalizer settings together with the designed equalizer architecture fulfill the construction of the corresponding equalizer model in MATLAB. The efficacy of the designed equalizer is then examined by evaluating the resultant link SER performance. In this work, the hardware implementations are also investigated. As shown in Fig. 5.12, once the equalizer architecture is determined, its structural properties, behavioral operations, and data flow can be defined with a hardware description language (HDL). In the HDL implementations, the DSP resolution, i.e., the number of bits used in



the arithmetic modules such as multipliers or adders, is a design parameter that is often chosen so as to discard round-off errors [84]. Commercial computer-aided design (CAD) tools have been developed to automate the creation and optimization of digital-gate-level circuit implementations. The CAD tools translate the HDL designs, and then perform automatic synthesis using the imported circuit libraries with the power consumption and timing information of the synthesized circuits reported. Finally, the function of the synthesized circuits is verified with commercial circuit simulators, by making sure these circuits generate identical outputs compared to those in MATLAB simulations.

In this work, Verilog is used to create the HDL abstraction. When the DSP resolution is set to 12 bits, including 1 bit for the sign, 2 bits for the integer part, and 9 bits for the decimal part, the digital equalizers introduce negligible round-off effects on the computed SER in Fig. 5.11. A commercial 28-nm CMOS technology and its standard cell libraries are imported to the CAD tool for the digital equalizer syntheses. While the aggregate operating data rates of the state-of-the-art DSP cores have exceeded 100 Gb/s, e.g., [85, 86], highly parallelized architectures are adopted in these DSP designs such that each slice is operated at roughly 1 GHz (875 MHz in [85], and 778 MHz in [86]). Accordingly, each slice of the NN-FFEs and VT-FFEs is synthesized for 1-GHz operations, and the resulting power overheads consumed for nonlinear equalization are summarized in Table 5.2. As can be seen, VT2-FFE and NN3-FFE offer strongest capability of nonlinear equalization, whereas the proposed NN3-FFE saves  $\sim 37\%$  of the power overhead consumed by VT2-FFE. Moreover, more power saving can be accomplished with the employment of NN2-FFE, which can already lead to significant improvement in the SER performance.

## 5.7 Summary

The proposed NN-FFEs achieve significant power saving at superior or similar SER performance compared to the conventional Volterra equalizer counterparts, by alleviating the explicit multiplicative computations of high-order terms. The presented design framework facilitates the adaptation of both Volterra equalizers and the ones employing a learnable PWL activation function, thereby allowing nonlinear equalizers of various

architectures to be explored. Meanwhile, the impact of a varied link component including the MRM on the overall link SER performance can be studied, favoring further link-level optimizations for the target SER specification.

*Chapter 6*

## CONCLUSION

Improving the energy efficiency of high-speed interconnects is the way to support the ever-increasing data traffic. That is, it is of great value and significance to achieve higher data bandwidths with a constant power budget. In this dissertation, efforts leading to energy-efficient high-speed receivers and interconnects are presented.

In an optical interconnect, the improvement in receiver sensitivity can substantially save the laser power consumption and thus benefit the overall energy efficiency. This suggests the employment of an APD, for its capability of enlarging SNR at the receiver front-end with the optimized gain. The ongoing advancement of the gain-bandwidth product of an APD has made it even more suitable for high-speed data communication. On the other hand, equalization plays a critical role in effectively ameliorating the receiver sensitivity by compensating for the ISI and hence expanding the data eye-openings. Especially for high-speed interconnects, where any slow dynamics within the signal path can induce considerable ISI, the inclusion of equalization in the receivers becomes more obligatory. In equalizer design, the conventional resistively loaded summer circuits can be replaced with current-integrating summer circuits to reduce power consumption. In addition, the double-sampling technique serves as an energy-efficient option to implement a two-tap FFE in discrete-time fashion, with the capability of cancelling the long-tail ISI stemming from a channel that well resembles a first-order  $RC$  low-pass system. Besides the pursuit of high receiver sensitivity, the necessity of incorporating adaptability in a burst-mode optical receiver is recognized, when this receiver is expected to take on the data bursts with distinct dc and amplitude characteristics from multiple transmitters. Rapid reconfiguration accommodating different dc and amplitude levels is highly desirable in order to benefit the overall link latency and bandwidth. Integrating dc and amplitude comparators are therefore proposed to eliminate the settling time constraints faced in the conventional designs based on  $RC$  low-pass filters, and to significantly accelerate the burst-mode reconfiguration. Designed with the foregoing

consideration and techniques, an APD-based burst-mode optical receiver, which employs current-integrating equalization along with the proposed integrating dc and amplitude comparators, is demonstrated. At 25 Gb/s, this NRZ receiver achieves  $-16$ -dBm OMA sensitivity with  $1.37$ -pJ/b energy efficiency for BER better than  $1E-12$  and accomplishes the reconfiguration in  $2.24$  ns.

The adoption of PAM4 signaling offers advantages over the NRZ modulation, by virtue of its superior spectral efficiency. That is, for a given target data rate, transceivers using PAM4 as the substitution for NRZ allow the clock signals to be generated and distributed at halved frequency, meanwhile permitting the channel to have lower bandwidth. In other words, replacing NRZ modulation with PAM4 signaling can potentially give rise to higher data rates and/or lower power consumptions. However, provided that the peak swing of the transmitter is fixed, the reduced eye-height as the result of PAM4 signaling sets a more demanding sensitivity requirement for the decision circuits. Moreover, compared to the NRZ cases, the smaller eye-openings in PAM4 systems appear to be more vulnerable to the same proportion of ISI level, raising the need for the use of equalization. DFE has been a particularly attractive option, in view of its capability of compensating for post-cursor ISI and reflections without enhancing the noise or crosstalk. Building a PAM4-DFE at high data rates can be challenging. The loop-unrolling technique requires prohibitively excessive hardware and area costs, even if only a few taps are unrolled. Implementing a PAM4-DFE in the direct feedback fashion, however, poses demanding speed requirement for the decision circuits to satisfy the tight DFE timing constraints. In light of these demanding sensitivity and speed requirements, a CMOS track-and-regenerate slicer is proposed as one solution that advances the performance of the prevalent CML and StrongArm slicers in several aspects. Designed in CMOS dynamic latch fashion, the proposed CMOS track-and-regenerate slicer leverages the cross-coupled pairs to regenerate the signals with strong positive feedback, improving the technological scalability, power efficiency, and output swing over the conventional CML slicer. In contrast to the reset-and-regenerate StrongArm slicer, the non-resetting mechanism of the proposed CMOS track-and-regenerate slicer leads to significantly improved clock-to-Q delay, higher gain, and thus better input sensitivity, when the allocated regeneration time

becomes stringent. Serving as a crucial building block, the proposed slicer enables an energy-efficient direct PAM4-DFE for high-speed operations, thanks to the direct availability of rail-to-rail digital feedback signals with shortened delay. A PAM4 receiver which incorporates CTLEs and a two-tap direct DFE employing the proposed CMOS track-and-regenerate slicer circuits is fabricated in 28-nm CMOS technology. At 60 Gb/s, this PAM4 receiver achieves BER better than  $1E-12$  and 1.1-pJ/b energy efficiency, measured over a channel with 8.2-dB loss at the Nyquist frequency.

An optical interconnect adopting PAM4 signaling, which combines the benefits from low modulation-frequency-dependent losses of optical fibers as well as the improved spectral efficiency over NRZ modulation, promises the accomplishment of higher data bandwidths. Nevertheless, nonlinearities in the forms of mismatched level-separations and dynamics, arising from the nonlinear responses of optical modulators, can further detrimentally compromise the relatively stringent SNR. Therefore, nonlinear equalization, aiming to counteract the undesirable nonlinearities, holds the key to realizing high-performance optical interconnects where high-order modulation formats are adopted. Unlike transmitter-side equalizers that would demand an extra back-channel to capture the overall signal characteristics, receiver-side equalizers are capable of compensating for the accumulated signal impairments including those attributed to the channel or the receiver front-end circuits without the need of a back-channel. Furthermore, with an ADC employed in the receiver, the receiver-side equalization can be implemented in the digital domain, which brings in the benefits from CMOS technology scaling and the strong immunity against PVT variations. Conventional digital Volterra equalizer-based FFEs (VT-FFEs) have proved effective in compensating for the nonlinearities by means of generating high-order terms with multipliers. In order to improve the power efficiency, neural-network-enhanced FFEs (NN-FFEs) employing a learnable custom PWL activation function are proposed, allowing the replacement of the relatively power-hungry multipliers with adders. MRM-based optical interconnects incorporating different nonlinear equalizer architectures are studied with simulation results of 50-Gb/s and 100-Gb/s PAM4 interconnects. Greater than 37% reduction in the power overhead can be achieved, by having an NN-FFE as the substitution for the VT-

FFE counterpart leading to similar SER improvement, with all equalizers synthesized in the same 28-nm CMOS technology.

In summary, this dissertation demonstrates energy-efficient receiver design for high-speed interconnects. The presented techniques enabling superior energy efficiencies count on the advancement of both optical device and electronic circuit design. An APD is employed along with the current-integrating equalizer for improved optical receiver sensitivity with relatively low power overhead. The proposed integrating dc and amplitude comparators empower substantial acceleration for the burst-mode reconfigurations, improving the overall link latency and bandwidth. The proposed CMOS track-and-regenerate slicer accomplishes the implementation of an energy-efficient direct PAM4-DFE at high data rates. The proposed NN-FFEs offer significant reduction in power and area consumptions, by cutting down the explicit multiplicative computations. These circuit techniques and the associated concepts can well extend to the development of other high-speed optical or electrical interconnects. The future high-speed interconnects place greater emphasis on the energy efficiency, in light of a limited power budget for supporting the ever-increasing data traffic. The fulfillment of the envisioned energy-efficient high-speed interconnects will have to rely even more on the innovations, co-design, and co-optimization of the transceiver architectures along with the electronic circuits as well as the optical components.

## BIBLIOGRAPHY

- [1] Cisco, "Cisco Annual Internet Report (2018-2023)," *White Paper*, 2020.
- [2] D. C. Daly, L. C. Fujino and K. C. Smith, "Through the Looking Glass-2020 Edition: Trends in Solid-State Circuits From ISSCC," in *IEEE Solid-State Circuits Magazine*, vol. 12, no. 1, pp. 8-24, winter 2020, doi: 10.1109/MSSC.2019.2952282.
- [3] 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," in *IEEE Solid-State Circuits Magazine*, vol. 10, no. 1, pp. 30-46, winter 2018, doi: 10.1109/MSSC.2017.2771103.
- [4] M. Raj *et al.*, "Design of a 50-Gb/s Hybrid Integrated Si-Photonic Optical Link in 16-nm FinFET," in *IEEE Journal of Solid-State Circuits*, vol. 55, no. 4, pp. 1086-1095, April 2020, doi: 10.1109/JSSC.2019.2960487.
- [5] M. Raj, M. Monge and A. Emami, "A Modelling and Nonlinear Equalization Technique for a 20 Gb/s 0.77 pJ/b VCSEL Transmitter in 32 nm SOI CMOS," in *IEEE Journal of Solid-State Circuits*, vol. 51, no. 8, pp. 1734-1743, Aug. 2016, doi: 10.1109/JSSC.2016.2553040.
- [6] H. Li *et al.*, "A 3-D-Integrated Silicon Photonic Microring-Based 112-Gb/s PAM-4 Transmitter With Nonlinear Equalization and Thermal Control," in *IEEE Journal of Solid-State Circuits*, vol. 56, no. 1, pp. 19-29, Jan. 2021, doi: 10.1109/JSSC.2020.3022851.
- [7] A. Roshan-Zamir *et al.*, "A two-segment optical DAC 40 Gb/s PAM4 silicon microring resonator modulator transmitter in 65nm CMOS," *2017 IEEE Optical Interconnects Conference (OI)*, 2017, pp. 5-6, doi: 10.1109/OIC.2017.7965503.
- [8] A. H. Talkhooncheh, A. Zilkie, G. Yu, R. Shafiiha, and A. Emami, "A 100 Gb/s PAM-4 silicon photonic transmitter with two binary-driven EAMs in MZI structure," *accepted to 2021 IEEE Photonics Conference (IPC)*, 2021.
- [9] J. Im *et al.*, "6.1 A 112Gb/s PAM-4 Long-Reach Wireline Transceiver Using a 36-Way Time-Interleaved SAR-ADC and Inverter-Based RX Analog Front-End in 7nm FinFET," *2020 IEEE International Solid-State Circuits Conference - (ISSCC)*, 2020, pp. 116-118, doi: 10.1109/ISSCC19947.2020.9063081.
- [10] M. Erett *et al.*, "A 2.25pJ/bit Multi-lane Transceiver for Short Reach Intra-package and Inter-package Communication in 16nm FinFET," *2019 IEEE Custom Integrated Circuits Conference (CICC)*, 2019, pp. 1-8, doi: 10.1109/CICC.2019.8780221.

- [11] J. Buckwalter and A. Hajimiri, "An active analog delay and the delay reference loop," *2004 IEEE Radio Frequency Integrated Circuits (RFIC) Systems. Digest of Papers*, 2004, pp. 17-20, doi: 10.1109/RFIC.2004.1320512.
- [12] S. Kiran, S. Cai, Y. Zhu, S. Hoyos and S. Palermo, "Digital Equalization With ADC-Based Receivers: Two Important Roles Played by Digital Signal Processing in Designing Analog-to-Digital-Converter-Based Wireline Communication Receivers," in *IEEE Microwave Magazine*, vol. 20, no. 5, pp. 62-79, May 2019, doi: 10.1109/MMM.2019.2898025.
- [13] Y. Lu and E. Alon, "Design Techniques for a 66 Gb/s 46 mW 3-Tap Decision Feedback Equalizer in 65 nm CMOS," in *IEEE Journal of Solid-State Circuits*, vol. 48, no. 12, pp. 3243-3257, Dec. 2013, doi: 10.1109/JSSC.2013.2278804.
- [14] K. K. Parhi, "Pipelining of parallel multiplexer loops and decision feedback equalizers," *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2004, pp. V-21, doi: 10.1109/ICASSP.2004.1327037.
- [15] S. Shahramian and A. Chan Carusone, "A 0.41 pJ/Bit 10 Gb/s Hybrid 2 IIR and 1 Discrete-Time DFE Tap in 28 nm-LP CMOS," in *IEEE Journal of Solid-State Circuits*, vol. 50, no. 7, pp. 1722-1735, July 2015, doi: 10.1109/JSSC.2015.2402218.
- [16] C. DeCusatis, "Optical Interconnect Networks for Data Communications," in *Journal of Lightwave Technology*, vol. 32, no. 4, pp. 544-552, Feb. 15, 2014, doi: 10.1109/JLT.2013.2279203.
- [17] J. C. Campbell, "Recent Advances in Avalanche Photodiodes," in *Journal of Lightwave Technology*, vol. 34, no. 2, pp. 278-285, 15 Jan. 15, 2016, doi: 10.1109/JLT.2015.2453092.
- [18] X. Chen *et al.*, "The Emergence of Silicon Photonics as a Flexible Technology Platform," in *Proceedings of the IEEE*, vol. 106, no. 12, pp. 2101-2116, Dec. 2018, doi: 10.1109/JPROC.2018.2854372.
- [19] A. Rylyakov *et al.*, "A 25 Gb/s Burst-Mode Receiver for Low Latency Photonic Switch Networks," in *IEEE Journal of Solid-State Circuits*, vol. 50, no. 12, pp. 3120-3132, Dec. 2015, doi: 10.1109/JSSC.2015.2478837.
- [20] M. G. Ahmed *et al.*, "A 12-Gb/s -16.8-dBm OMA Sensitivity 23-mW Optical Receiver in 65-nm CMOS," in *IEEE Journal of Solid-State Circuits*, vol. 53, no. 2, pp. 445-457, Feb. 2018, doi: 10.1109/JSSC.2017.2757008.
- [21] A. Tyagi *et al.*, "A 50 Gb/s PAM-4 VCSEL Transmitter With 2.5-Tap Nonlinear Equalization in 65-nm CMOS," in *IEEE Photonics Technology Letters*, vol. 30, no. 13, pp. 1246-1249, 1 July 1, 2018, doi: 10.1109/LPT.2018.2841841.



- [22] C. Liao and S. Liu, "40 Gb/s Transimpedance-AGC Amplifier and CDR Circuit for Broadband Data Receivers in 90 nm CMOS," in *IEEE Journal of Solid-State Circuits*, vol. 43, no. 3, pp. 642-655, March 2008, doi: 10.1109/JSSC.2007.916626.
- [23] T. H. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*, Cambridge, U.K.:Cambridge Univ. Press, 1998.
- [24] Y. Kang *et al.*, "Monolithic germanium/silicon avalanche photodiodes with 340 GHz gain–bandwidth product", *Nature Photon.*, vol. 3, pp. 59-63, Jan. 2009.
- [25] M. Huang *et al.*, "25Gb/s normal incident Ge/Si avalanche photodiode," *2014 The European Conference on Optical Communication (ECOC)*, 2014, pp. 1-3, doi: 10.1109/ECOC.2014.6964088.
- [26] M. Nada, Y. Yamada and H. Matsuzaki, "Responsivity-Bandwidth Limit of Avalanche Photodiodes: Toward Future Ethernet Systems," in *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 24, no. 2, pp. 1-11, March-April 2018, Art no. 3800811, doi: 10.1109/JSTQE.2017.2754361.
- [27] G. P. Agrawal, *Lightwave Technology: Telecommunication Systems*, New York, NY, USA: Wiley, 2005.
- [28] R. J. McIntyre, "Multiplication noise in uniform avalanche diodes," in *IEEE Transactions on Electron Devices*, vol. ED-13, no. 1, pp. 164-168, Jan. 1966, doi: 10.1109/T-ED.1966.15651.
- [29] M. H. Nazari and A. Emami-Neyestanak, "An 18.6Gb/s double-sampling receiver in 65nm CMOS for ultra-low-power optical communication," *2012 IEEE International Solid-State Circuits Conference*, 2012, pp. 130-131, doi: 10.1109/ISSCC.2012.6176949.
- [30] M. H. Nazari and A. Emami-Neyestanak, "A 24-Gb/s Double-Sampling Receiver for Ultra-Low-Power Optical Communication," in *IEEE Journal of Solid-State Circuits*, vol. 48, no. 2, pp. 344-357, Feb. 2013, doi: 10.1109/JSSC.2012.2227612.
- [31] Seuk Son, Hanseok Kim, Myeong-Jae Park, Kyung Hoon Kim and Jaeha Kim, "A 2.3-mW, 5-Gb/s decision-feedback equalizing receiver front-end with static-power-free signal summation and CDR-based precursor ISI reduction," *2012 IEEE Asian Solid State Circuits Conference (A-SSCC)*, 2012, pp. 133-136, doi: 10.1109/IPEC.2012.6522643.
- [32] S. Saeedi and A. Emami, "A 25Gb/s 170 $\mu$ W/Gb/s optical receiver in 28nm CMOS for chip-to-chip optical communication," *2014 IEEE Radio Frequency Integrated Circuits Symposium*, 2014, pp. 283-286, doi: 10.1109/RFIC.2014.6851720.
- [33] A. Cevrero *et al.*, "29.1 A 64Gb/s 1.4pJ/b NRZ optical-receiver data-path in 14nm CMOS FinFET," *2017 IEEE International Solid-State Circuits Conference (ISSCC)*, 2017, pp. 482-483, doi: 10.1109/ISSCC.2017.7870471.

- [34] J. Proesel *et al.*, "A 32Gb/s, 4.7pJ/bit optical link with  $-11.7$ dBm sensitivity in 14nm FinFET CMOS," *2017 Symposium on VLSI Circuits*, 2017, pp. C318-C319, doi: 10.23919/VLSIC.2017.8008523.
- [35] I. Ozkaya *et al.*, "A 64-Gb/s 1.4-pJ/b NRZ Optical Receiver Data-Path in 14-nm CMOS FinFET," in *IEEE Journal of Solid-State Circuits*, vol. 52, no. 12, pp. 3458-3473, Dec. 2017, doi: 10.1109/JSSC.2017.2734913.
- [36] D. Schinkel, E. Mensink, E. Klumperink, E. van Tuijl and B. Nauta, "A Double-Tail Latch-Type Voltage Sense Amplifier with 18ps Setup+Hold Time," *2007 IEEE International Solid-State Circuits Conference. Digest of Technical Papers*, 2007, pp. 314-605, doi: 10.1109/ISSCC.2007.373420.
- [37] A. Roshan-Zamir, O. Elhadidy, H. Yang and S. Palermo, "A Reconfigurable 16/32 Gb/s Dual-Mode NRZ/PAM4 SerDes in 65-nm CMOS," in *IEEE Journal of Solid-State Circuits*, vol. 52, no. 9, pp. 2430-2447, Sept. 2017, doi: 10.1109/JSSC.2017.2705070.
- [38] J. Han, Y. Lu, N. Sutardja, K. Jung and E. Alon, "Design Techniques for a 60 Gb/s 173 mW Wireline Receiver Frontend in 65 nm CMOS Technology," in *IEEE Journal of Solid-State Circuits*, vol. 51, no. 4, pp. 871-880, April 2016, doi: 10.1109/JSSC.2016.2519389.
- [39] X. Yin *et al.*, "A 10Gb/s burst-mode TIA with on-chip reset/lock CM signaling detection and limiting amplifier with a 75ns settling time," *2012 IEEE International Solid-State Circuits Conference*, 2012, pp. 416-418, doi: 10.1109/ISSCC.2012.6177071.
- [40] T. D. Ridder *et al.*, "10 Gbit/s burst-mode post-amplifier with automatic reset," *Electron. Lett.*, vol. 44, no. 23, pp. 1371-1373, Nov. 2008.
- [41] Y. Frans *et al.*, "A 56-Gb/s PAM4 Wireline Transceiver Using a 32-Way Time-Interleaved SAR ADC in 16-nm FinFET," in *IEEE Journal of Solid-State Circuits*, vol. 52, no. 4, pp. 1101-1110, April 2017, doi: 10.1109/JSSC.2016.2632300.
- [42] T. Ali *et al.*, "6.4 A 180mW 56Gb/s DSP-Based Transceiver for High Density IOs in Data Center Switches in 7nm FinFET Technology," *2019 IEEE International Solid-State Circuits Conference - (ISSCC)*, 2019, pp. 118-120, doi: 10.1109/ISSCC.2019.8662523.
- [43] D. Pfaff *et al.*, "A 56Gb/s Long Reach Fully Adaptive Wireline PAM-4 Transceiver in 7nm FinFET," *2019 Symposium on VLSI Circuits*, 2019, pp. C270-C271, doi: 10.23919/VLSIC.2019.8778051.
- [44] M. Pisati *et al.*, "A 243-mW 1.25–56-Gb/s Continuous Range PAM-4 42.5-dB IL ADC/DAC-Based Transceiver in 7-nm FinFET," in *IEEE Journal of Solid-State Circuits*, vol. 55, no. 1, pp. 6-18, Jan. 2020, doi: 10.1109/JSSC.2019.2936307.

- [45] J. Im *et al.*, "A 40-to-56 Gb/s PAM-4 Receiver With Ten-Tap Direct Decision-Feedback Equalization in 16-nm FinFET," in *IEEE Journal of Solid-State Circuits*, vol. 52, no. 12, pp. 3486-3502, Dec. 2017, doi: 10.1109/JSSC.2017.2749432.
- [46] A. Roshan-Zamir *et al.*, "A 56-Gb/s PAM4 Receiver With Low-Overhead Techniques for Threshold and Edge-Based DFE FIR- and IIR-Tap Adaptation in 65-nm CMOS," in *IEEE Journal of Solid-State Circuits*, vol. 54, no. 3, pp. 672-684, March 2019, doi: 10.1109/JSSC.2018.2881278.
- [47] P. Peng, J. Li, L. Chen and J. Lee, "6.1 A 56Gb/s PAM-4/NRZ transceiver in 40nm CMOS," *2017 IEEE International Solid-State Circuits Conference (ISSCC)*, 2017, pp. 110-111, doi: 10.1109/ISSCC.2017.7870285.
- [48] M. Choi and A. A. Abidi, "A 6 b 1.3 GSamples/s A/D converter in 0.35  $\mu\text{m}$  CMOS," *2001 IEEE International Solid-State Circuits Conference. Digest of Technical Papers. ISSCC (Cat. No.01CH37177)*, 2001, pp. 126-127, doi: 10.1109/ISSCC.2001.912571.
- [49] T. Toifl *et al.*, "A 22-gb/s PAM-4 receiver in 90-nm CMOS SOI technology," in *IEEE Journal of Solid-State Circuits*, vol. 41, no. 4, pp. 954-965, April 2006, doi: 10.1109/JSSC.2006.870898.
- [50] K. J. Wong, A. Rylyakov and C. K. Yang, "A 5-mW 6-Gb/s Quarter-Rate Sampling Receiver With a 2-Tap DFE Using Soft Decisions," in *IEEE Journal of Solid-State Circuits*, vol. 42, no. 4, pp. 881-888, April 2007, doi: 10.1109/JSSC.2007.892189.
- [51] T. Kobayashi, K. Nogami, T. Shirotori, Y. Fujimoto and O. Watanabe, "A current-mode latch sense amplifier and a static power saving input buffer for low-power architecture," *1992 Symposium on VLSI Circuits Digest of Technical Papers*, 1992, pp. 28-29, doi: 10.1109/VLSIC.1992.229252.
- [52] P. A. Francese *et al.*, "23.6 A 30Gb/s 0.8pJ/b 14nm FinFET receiver data-path," *2016 IEEE International Solid-State Circuits Conference (ISSCC)*, 2016, pp. 408-409, doi: 10.1109/ISSCC.2016.7418080.
- [53] K. C. Chen and A. Emami, "A 25-Gb/s Avalanche Photodetector-Based Burst-Mode Optical Receiver With 2.24-ns Reconfiguration Time in 28-nm CMOS," in *IEEE Journal of Solid-State Circuits*, vol. 54, no. 6, pp. 1682-1693, June 2019, doi: 10.1109/JSSC.2019.2902471.
- [54] J. Kim, B. S. Leibowitz, J. Ren and C. J. Madden, "Simulation and Analysis of Random Decision Errors in Clocked Comparators," in *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 56, no. 8, pp. 1844-1857, Aug. 2009, doi: 10.1109/TCSI.2009.2028449.

- [55] M. Jeeradit *et al.*, "Characterizing sampling aperture of clocked comparators," *2008 IEEE Symposium on VLSI Circuits*, 2008, pp. 68-69, doi: 10.1109/VLSIC.2008.4585955.
- [56] J. Han, N. Sutardja, Y. Lu and E. Alon, "Design Techniques for a 60-Gb/s 288-mW NRZ Transceiver With Adaptive Equalization and Baud-Rate Clock and Data Recovery in 65-nm CMOS Technology," in *IEEE Journal of Solid-State Circuits*, vol. 52, no. 12, pp. 3474-3485, Dec. 2017, doi: 10.1109/JSSC.2017.2740268.
- [57] P. Upadhyaya *et al.*, "A fully adaptive 19-to-56Gb/s PAM-4 wireline transceiver with a configurable ADC in 16nm FinFET," *2018 IEEE International Solid - State Circuits Conference - (ISSCC)*, 2018, pp. 108-110, doi: 10.1109/ISSCC.2018.8310207.
- [58] C. Wang, G. Zhu, Z. Zhang and C. P. Yue, "A 52-Gb/s Sub-1pJ/bit PAM4 Receiver in 40-nm CMOS for Low-Power Interconnects," *2019 Symposium on VLSI Circuits*, 2019, pp. C274-C275, doi: 10.23919/VLSIC.2019.8778159.
- [59] S. Babayan-Mashhadi and R. Lotfi, "Analysis and Design of a Low-Voltage Low-Power Double-Tail Comparator," in *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 22, no. 2, pp. 343-352, Feb. 2014, doi: 10.1109/TVLSI.2013.2241799.
- [60] K. Chen, W. W. Kuo and A. Emami, "A 60-Gb/s PAM4 Wireline Receiver with 2-Tap Direct Decision Feedback Equalization Employing Track-and-Regenerate Slicers in 28-nm CMOS," *2020 IEEE Custom Integrated Circuits Conference (CICC)*, 2020, pp. 1-4, doi: 10.1109/CICC48029.2020.9075948.
- [61] W. Huang *et al.*, "93% Complexity Reduction of Volterra Nonlinear Equalizer by  $\ell_1$ -Regularization for 112-Gbps PAM-4 850-nm VCSEL Optical Interconnect," *2018 Optical Fiber Communications Conference and Exposition (OFC)*, 2018, pp. 1-3.
- [62] N. Stojanovic, F. Karinou, Z. Qiang and C. Prodaniuc, "Volterra and Wiener Equalizers for Short-Reach 100G PAM-4 Applications," in *Journal of Lightwave Technology*, vol. 35, no. 21, pp. 4583-4594, 1 Nov.1, 2017, doi: 10.1109/JLT.2017.2752363.
- [63] A. Masuda, S. Yamamoto, H. Taniguchi and M. Fukutoku, "Achievement of 90-Gbaud PAM-4 with MLSE Based on 2nd Order Volterra Filter and 2.88-Tb/s O-band Transmission Using 4- $\lambda$  LAN-WDM and 4-Core Fiber SDM," *2018 Optical Fiber Communications Conference and Exposition (OFC)*, 2018, pp. 1-3.
- [64] L. Zhang *et al.*, "Nonlinearity Tolerant High-Speed DMT Transmission With 1.5- $\mu\text{m}$  Single-Mode VCSEL and Multi-Core Fibers for Optical Interconnects," in *Journal of Lightwave Technology*, vol. 37, no. 2, pp. 380-388, 15 Jan.15, 2019, doi: 10.1109/JLT.2018.2851746.

- [65] C. Chuang *et al.*, "Sparse Volterra Nonlinear Equalizer by Employing Pruning Algorithm for High-Speed PAM-4 850-nm VCSEL Optical Interconnect," *2019 Optical Fiber Communications Conference and Exhibition (OFC)*, 2019, pp. 1-3.
- [66] Y. Fu *et al.*, "Piecewise Linear Equalizer for DML Based PAM-4 Signal Transmission Over a Dispersion Uncompensated Link," in *Journal of Lightwave Technology*, vol. 38, no. 3, pp. 654-660, 1 Feb.1, 2020, doi: 10.1109/JLT.2019.2948096.
- [67] Y. Yu, T. Bo, Y. Che, D. Kim and H. Kim, "Low-Complexity Equalizer Based on Volterra Series and Piecewise Linear Function for DML-Based IM/DD System," *2020 Optical Fiber Communications Conference and Exhibition (OFC)*, 2020, pp. 1-3.
- [68] Q. Xu, B. Schmidt, J. Shakya, and M. Lipson, "Cascaded silicon micro-ring modulators for WDM optical interconnection," *Opt. Express* 14, 9431-9436 (2006).
- [69] J. C. Rosenberg *et al.*, "High-speed and low-power microring modulators for silicon photonics," *IEEE Photonic Society 24th Annual Meeting*, 2011, pp. 256-257, doi: 10.1109/PHO.2011.6110518.
- [70] M. Pantouvaki *et al.*, "Active Components for 50 Gb/s NRZ-OOK Optical Interconnects in a Silicon Photonics Platform," in *Journal of Lightwave Technology*, vol. 35, no. 4, pp. 631-638, 15 Feb.15, 2017, doi: 10.1109/JLT.2016.2604839.
- [71] H. Ramon *et al.*, "Low-Power 56Gb/s NRZ Microring Modulator Driver in 28nm FDSOI CMOS," in *IEEE Photonics Technology Letters*, vol. 30, no. 5, pp. 467-470, 1 March1, 2018, doi: 10.1109/LPT.2018.2799004.
- [72] M. Moralis-Pegios *et al.*, "A 160 Gb/s (4x40) WDM O-band Tx subassembly using a 4-Ch array of silicon rings co-packaged with a SiGe BiCMOS IC driver," *45th European Conference on Optical Communication (ECOC 2019)*, 2019, pp. 1-4, doi: 10.1049/cp.2019.0914.
- [73] S. Fatholouloumi *et al.*, "1.6 Tbps Silicon Photonics Integrated Circuit and 800 Gbps Photonic Engine for Switch Co-Packaging Demonstration," in *Journal of Lightwave Technology*, vol. 39, no. 4, pp. 1155-1161, 15 Feb.15, 2021, doi: 10.1109/JLT.2020.3039218.
- [74] S. Moazeni *et al.*, "A 40-Gb/s PAM-4 Transmitter Based on a Ring-Resonator Optical DAC in 45-nm SOI CMOS," in *IEEE Journal of Solid-State Circuits*, vol. 52, no. 12, pp. 3503-3516, Dec. 2017, doi: 10.1109/JSSC.2017.2748620.
- [75] S. Lin, S. Moazeni, K. T. Settaluri and V. Stojanović, "Electronic-Photonic Co-Optimization of High-Speed Silicon Photonic Transmitters," in *Journal of Lightwave Technology*, vol. 35, no. 21, pp. 4766-4780, 1 Nov.1, 2017, doi: 10.1109/JLT.2017.2757945.

- [76] M. Song, L. Zhang, R. G. Beausoleil and A. E. Willner, "Nonlinear Distortion in a Silicon Microring-Based Electro-Optic Modulator for Analog Optical Links," in *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 16, no. 1, pp. 185-191, Jan.-feb. 2010, doi: 10.1109/JSTQE.2009.2030154.
- [77] Z. Wan *et al.*, "64-Gb/s SSB-PAM4 Transmission Over 120-km Dispersion-Uncompensated SSMF With Blind Nonlinear Equalization, Adaptive Noise-Whitening Postfilter and MLSD," in *Journal of Lightwave Technology*, vol. 35, no. 23, pp. 5193-5200, 1 Dec.1, 2017, doi: 10.1109/JLT.2017.2768431.
- [78] F. Ortega-Zamorano, J. M. Jerez, G. Juárez, J. O. Pérez and L. Franco, "High precision FPGA implementation of neural network activation functions," *2014 IEEE Symposium on Intelligent Embedded Systems (IES)*, 2014, pp. 55-60, doi: 10.1109/INTELES.2014.7008986.
- [79] A. H. Namin, K. Leboeuf, R. Muscedere, H. Wu and M. Ahmadi, "Efficient hardware implementation of the hyperbolic tangent sigmoid function," *2009 IEEE International Symposium on Circuits and Systems*, 2009, pp. 2117-2120, doi: 10.1109/ISCAS.2009.5118213.
- [80] G. E. Dahl, T. N. Sainath and G. E. Hinton, "Improving deep neural networks for LVCSR using rectified linear units and dropout," *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp. 8609-8613, doi: 10.1109/ICASSP.2013.6639346.
- [81] D. Li *et al.*, "180 Gb/s PAM8 Signal Transmission in Bandwidth-Limited IMDD System Enabled by Tap Coefficient Decision Directed Volterra Equalizer," in *IEEE Access*, vol. 8, pp. 19890-19899, 2020, doi: 10.1109/ACCESS.2020.2968128.
- [82] H. Wey and W. Guggenbuhl, "Noise transfer characteristics of a correlated double sampling circuit," in *IEEE Transactions on Circuits and Systems*, vol. 33, no. 10, pp. 1028-1030, October 1986, doi: 10.1109/TCS.1986.1085840.
- [83] K. Szczerba *et al.*, "4-PAM for high-speed short-range optical communications," in *IEEE/OSA Journal of Optical Communications and Networking*, vol. 4, no. 11, pp. 885-894, Nov. 2012, doi: 10.1364/JOCN.4.000885.
- [84] S. Kiran *et al.*, "Modeling of ADC-Based Serial Link Receivers With Embedded and Digital Equalization," in *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 9, no. 3, pp. 536-548, March 2019, doi: 10.1109/TCPMT.2018.2853080.
- [85] Y. Krupnik *et al.*, "112-Gb/s PAM4 ADC-Based SERDES Receiver With Resonant AFE for Long-Reach Channels," in *IEEE Journal of Solid-State Circuits*, vol. 55, no. 4, pp. 1077-1085, April 2020, doi: 10.1109/JSSC.2019.2959511.

- [86] J. Im *et al.*, "A 112-Gb/s PAM-4 Long-Reach Wireline Transceiver Using a 36-Way Time-Interleaved SAR ADC and Inverter-Based RX Analog Front-End in 7-nm FinFET," in *IEEE Journal of Solid-State Circuits*, vol. 56, no. 1, pp. 7-18, Jan. 2021, doi: 10.1109/JSSC.2020.3024261.