Adaptive Receiver Design for High Speed Optical Communication

THÈSE N^{O} 8684 (2018)

PRÉSENTÉE LE 15 JUIN 2018

À LA FACULTÉ DES SCIENCES ET TECHNIQUES DE L'INGÉNIEUR
LABORATOIRE DE SYSTÈMES MICROÉLECTRONIQUES
PROGRAMME DOCTORAL EN MICROSYSTÈMES ET MICROÉLECTRONIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Ilter ÖZKAYA

acceptée sur proposition du jury:

Dr G. Boero, président du jury
Prof. Y. Leblebici, Dr T. Toifl, directeurs de thèse
Prof. P. Hanumolu, rapporteur
Prof. S. Palermo, rapporteur
Prof. A. Burg, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE
Suisse
2018

Acknowledgements

This thesis would not have been possible without the help of all the people supporting me in the process. First of all, I would like to thank my advisors Prof. Yusuf Leblebici and Dr. Thomas Toifl first for accepting me for the position and then for their continuing support both in technical and non-technical problems I faced throughout my PhD. The insightful meetings and discussions have always led to creative solutions that produced outstanding results achieved in this work.

Of course, it would not have been possible to accomplish all that if my colleague and friend Dr. Alessandro Cevrero did not work with me on the project literally day and night. Thank you Ale!

I also would like to thank my colleague and friend Dr. Pier Andrea Francese for sharing his deep expertise on analog circuit design and IO links as well as providing me with his personal opinion on almost any subject. Furthermore, our mutual interest in history have always led to curious conversations during coffee breaks.

There are many other people from High Speed IO team in IBM Zurich Research Labs that I would like to thank: Dr. Lukas Kull who has thought me a lot about Swiss way of thinking; Dr. Christian Menolfi who was always willing to sacrifice his own time to solve my CAD related problems; Dr. Thomas Morf, Dr. Marcel Kossel, Dr. Mattheas Braendli, Danny Luu, and Joerg-Eric Sagmeister.

Last but not the least, I would like to thank my family: My wife for loving me through thick and thin, and supporting me in all my decisions, my 2 sons Bartu and Sarp for giving me the motivation to go on, and my parents for bringing me up to be who I am!

Ilter Ozkaya
Rueschlikon, 14 March 2018

Abstract

Conventional input/output (IO) links consume power, independent of changes in the bandwidth demand by the system they are deployed in. As the system is designed to satisfy the peak bandwidth demand, most of the time the IO links are idle but still consuming power. In big data centers, the overall utilization ratio of IO links is less than 10%, corresponding to a large amount of energy wasted for idle operation.

This work demonstrates a 60 Gb/s high sensitivity non-return-to-zero (NRZ) optical receiver in 14 nm FinFET technology with less than 7 ns power-on time. The power on time includes the data detection, analog bias settling, photo-diode DC current cancellation, and phase locking by the clock and data recovery circuit (CDR). The receiver autonomously detects the data demand on the link via a proposed link protocol and does not require any external enable or disable signals. The proposed link protocol is designed to minimize the off-state power consumption and power-on time of the link.

In order to achieve high data-rate and high-sensitivity while maintaining the power budget, a 1-tap decision feedback equalization method is applied in digital domain. The sensitivity is measured to be -8 dBm, -11 dBm, and -13 dBm OMA (optical modulation amplitude) at 60 Gb/s, 48 Gb/s, and 32 Gb/s data rates, respectively. The energy efficiency in always-on mode is around 2.2 pJ/bit for all data-rates with the help of supply and bias scaling.

The receiver incorporates a phase interpolator based clock-and-data recovery circuit with approximately 80 MHz jitter-tolerance corner frequency, thanks to the low-latency full custom CDR logic design.

This work demonstrates the fastest ever reported CMOS optical receiver and runs almost at twice the data-rate of the state-of-the-art CMOS optical receiver by the time of the publication. The data-rate is comparable to BiCMOS optical receivers but at a fraction of the power consumption.

Zusammenfassung

Konventionelle Datenlinks haben einen Leistungsverbrauch, der unabhängig vom augenblicklichen Bandbreitenbedarf des Systems ist, weil diese Systeme so konzipiert sind, dass sie die Spitzenbandbreite erbringen können. Meistens haben diese Links einen unveränderlichen Leistungsverbrauch auch wenn keine Daten übertragen werden. In grossen Datenzentren ist der Auslastungsgrad aller Links unter 10%. Dies resultiert in eine grosse Menge von verschwendeter Energie, während diese Links im Leerlauf sind.

Diese Dissertation präsentiert einen 60 Gb/s Non-Return-to-Zero (NRZ) optischen Empfänger mit hoher Empfindlichkeit in einer 14nm FinFET Technologie, der eine Einschaltdauer von unter 7 ns erreicht. Die Einschaltdauer beinhaltet das Detektieren eines neuen Datenpakets, die Einschwingzeit der analogen Bias-Spannungen, die Kompensation des Photodiodengleichstroms, und die Synchronisation des Symboltakts (CDR). Der Empfänger detektiert selbständig den Übertragungsbedarf durch einen eigens dafür entwickeltes Protokoll und benötigt daher kein externes Einschaltsignal. Das aufgezeigte Protokoll ermöglicht die Minimierung des Energieverbrauchs im Leerlauf und der Einschaltzeit.

Um eine hohe Datenübertragungsrate und Empfindlichkeit bei tiefem Leistungsverbrauch zu erreichen, wird ein 1-Tap Decision-Feedback-Equalization (DFE) verwendet. Die gemessene Empfindlichkeit des Empfängers ist -8 dBm, -11 dBm oder -13 dBm optische Modulationsamplitude (OMA) bei einer entsprechenden Datenübertragungsrate von 60 Gb/s, 48 Gb/s oder 32 Gb/s. Die Energieeffizienz im Dauerbetrieb ist 2.2 pJ/bit über alle Datenraten mithilfe der Skalierung der Versorgungsspannung und Bias-Spannungen.

Der Empfänger enthält eine CDR-Schaltung mit Phaseninterpolator, der dank der tiefen Latenz der massgefertigten CDR-Logik eine Jittertoleranzgrenze von ungefähr 80 MHz erreicht.

Die präsentierte Implementation ist der schnellste optische CMOS-Empfänger mit einer Datenübertragungsrate, die beinahe doppelt so hoch ist im Vergleich mit den neusten optischen CMOS-Empfängern bis zum Zeitpunkt der Publikation. Die erreichte Datenrate ist vergleichbar mit optischen BiCMOS-Empfängern, jedoch bei einem Bruchteil des Leistungsverbrauchs

Contents

Chapter 1

Introduction

Continuous developments in the semiconductor industry allowed the exponential increase in transistor density per chip, keeping up with the Moore's law, up to date. Following up Moore's prediction in transistor scaling, Dennard [1] calculated that the power saving introduced by smaller transistors and lower supply voltages will be compensated by the increase in transistor count and clock speed resulting in constant power density over the chip area. The power density being constant, the computation power scaled with transistor density without affecting the package or power source of the product for many generations.

However, around 2003 Dennard scaling has ended [2] leading to an increase in power density in the silicon whose pace can not be matched by improvements in the ability to dissipate heat off the chip. The confluence of these trends has led to a phenomenon referred to as the utilization wall or dark silicon: a chip sized for the economic manufacturability sweet-spot will have far more transistors than can be used on a sustained basis [3]. Thus, in normal operation, a large part of the chip is put in a low-power mode to maintain a constant thermal envelope. The full performance of the chip can only be sustained for very short bursts with architectural solutions such as computational sprinting presented in [3].

Thermal design packaging can be considered as the chip-scale aspect of the power management problem faced by the information and communication technology (ICT) industry. Regarding the size and continuous growth of this industry, there also exists a global-scale aspect of power management problem. The green house gases caused by the ICT industry is becoming a major problem. Global data centers used roughly 416 terawatts (4.16×10^{14} watts), which corresponds to about 3% of the total electricity used globally in 2016, nearly 40% more than the entire United Kingdom and more than the total power consumed by the whole aviation business worldwide. Moreover, the consumption is expected to double every four years [4] in the future.

The power consumption inside a data center is roughly distributed as follows: around 40% of the total IT power is consumed by servers, up to 37% by storage, and 23% by the network devices [5]. On the other hand, this distribution does not include the networking power due to the local communication between the server and the local memory. For example, the I/O links included in the POWER9, which

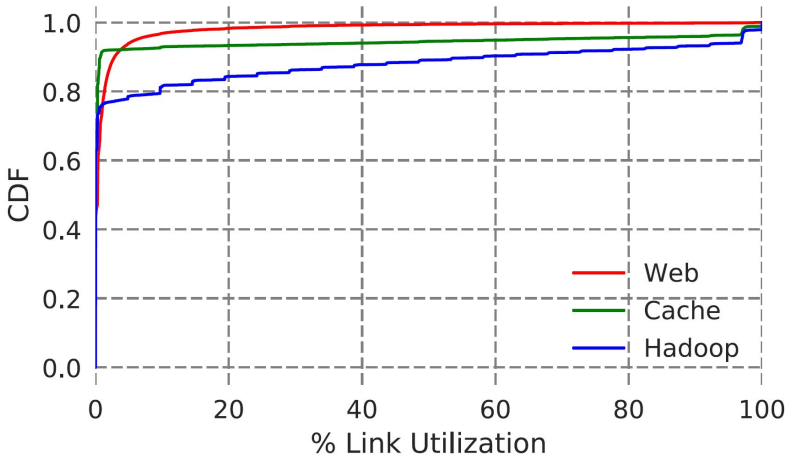

Figure 1.1: CDF of link utilization [8]

is the latest microprocessor of IBM designed to be used in data centers for high performance computing, consume approximately 10% of the total chip power [6]. The same can be assumed for local storage devices, potentially increasing the share of networking up to around 1/3 of the whole data center power consumption. Thus, a power reduction in networking would have a significant effect on the the reduction of green house gases emitted by the data centers, as well as the costs of the services provided by the ICT industry.

Latest publications report very low link utilization ratios inside data centers. “Even the most loaded links are lightly loaded over 1-minute time scales: 99% of all links are typically less than 10% loaded.” is reported for Facebook’s data centers in [7]. A more detailed study with higher timing resolution in analyzing data packets declares similar results as given in Fig. 1.1. The plot clearly shows that the average link utilization is on the order of only 1-2% for all three types of web servers included in the study: web, cache, and hadoop. Since the conventional links are kept powered on independent of link utilization, a large amount of energy is wasted during the idle periods.

Some of the system level techniques proposed to improve energy efficiency of low utilization networking interface can be listed as follows [9]:

- Sleep: In this scheme, the network components such as switches, routers are put to sleep or are switched off in the idle period in between of workload arrivals [10], [11]
- Aggregation: Modifying the network topology so as to consolidate the network flow on minimum active set of network devices [10]
- Rate adaptation: Use of rate adaptation technique is demonstrated by adjusting the workload rate such that traffic is serviced within the required

time constraints [11]

- Traffic shaping: In ElasticTree scheme [12], the traffic is split into bursts, such that traffic to same destinations is buffered before it is routed. This scheme increased the idle periods between the traffic bursts used to transition the network devices into low power states.

Although improving the energy efficiency, those techniques are slow reacting and cumbersome to implement. Most of the time they require digital processing such as aggregation and traffic shaping, both of which may result in significant increase in latency that may not be tolerated in some applications. Moreover, those techniques are applicable to stand alone devices such as routers and switches, and are very difficult to apply to IO links of the servers for example.

In this thesis, a transistor level solution is proposed as a means of saving energy by powering off the idle IO links. The advantage of the proposed solution over the techniques mentioned above is that the custom design will reduce the power-on and -off times orders of magnitude compared to system level solutions. This reduction is expected to have a significant impact on networking applications both in data centers and chip design. For example, the power saving in networking of a data center -around 1/3 of the total power- can be reallocated for computation and storing devices, making the data centers more efficient and reducing the service costs significantly. Moreover, a rapid power-on and -off IO link may change the server architecture significantly. The IO links that could not be turned-off previously, eating-up around 10% of the power, can be powered down allowing higher thermal budget for computation. Furthermore, since the IO links would be energy proportional, the bandwidth of the server can be extended significantly without compromising the thermal sustainability of the chip.

Specifically, in this work, an adaptive optical receiver with rapid power-on functionality was implemented using 14 nm FinFET technology. The receiver can operate at a data rate of 60 Gb/s with approximately 7 ns wake-up time, enabling burst-mode operation in packet level. As a result, the bursty data does not have to be heavily aggregated for high efficiency, reducing the complexity and latency of the network. The receiver is autonomous and can power itself up and down depending on the bandwidth demand without the need for an enable or disable signal. Thus the power on and off modes of the receiver do not have to be controlled by a controller, further simplifying the system design.

The burst mode operation of the presented RX relies on a proposed data protocol. The data protocol is used to determine the beginning and end of a data burst as well as to reduce the clock and data recovery phase locking time.

The receiver was designed to provide state of the art performance in terms of data-rate and energy-efficiency in order to meet the standards expected from

the next generation products. In other words, neither speed nor energy-efficiency was compromised by the rapid power-on functionality. As a result, the proposed architecture can be used without the need to significantly improve performance. Moreover, the implementation is done in a CMOS technology enabling the receivers direct integration into larger chips. The proposed RX architecture can also be migrated to smaller technology nodes, such as 10 nm and 7 nm for higher integration capability. It is also possible to convert it into an electrical receiver with similar power-on performance as the CDR architecture can be reused as it is.

1.1 Thesis Goal

The goal of this thesis is to propose a suitable architecture, design and implement an adaptive high speed receiver for optical communication in 14 nm FinFET process. At the beginning of the thesis project the targeted specifications were as follows:

- Data rate (target is > 56 Gb/s): The higher the data rate the more cost efficient an optical link is. To the best of the authors knowledge, the speed target was double the data rate of the state of the art CMOS optical receivers ([13], [14]) at the time the project was specified.

- Power-on time (target is < 10 ns): In order for the receiver to be efficient even with small packets, the power-on time needs to be minimized. The power-on time includes the settling of analog biases, the DC offset cancellation of the photo diode current and phase locking by CDR. Since the adaptivity itself is a relatively new concept there are not many publications on high speed applications to compare against. In 2015, Anand et al. presented an adaptive transceiver with a power-on time of less than 20 ns in [15]. However, it is not straightforward to compare performance of the receiver presented in [15] with the targeted power-on time due to the following reasons:

  - Electrical interface: Unlike optical receivers, differential electrical links do not need to find a reference for the incoming signal. Thus, during power-on, electrical links do not need to cancel dc offset, which may require extra time.

  - Source Synchronous: The transceiver presented in [15] is source synchronous, whereas this work covers non-source synchronous implementations as well.

  - Data rate: The target data rate is 8 times faster than the data-rate of [15], reducing available timing margin significantly in the eye-diagram.

- Sensitivity (target is < -9 dBm OMA): The losses on the optical path and connections may reduce the input signal amplitude significantly. Thus, sensitivity is an important specification in an optical receiver. The target sensitivity for the receiver is set to match state of the art sensitivity of -9 dBm at 56 Gb/s reported in [16].

- Energy efficiency (target is < 2 pJ/b): The proposed solution should be competitive in terms of power consumption for high utilization ratios as well. Hence, the always on power target is set as < 2 pJ/b to be comparable to state of the art optical receivers.

- Off-power (target is minimum): The power consumption in off mode limits the power savings of the receiver for low utilization links.

Adaptive transmitter design was not included in the scope of this thesis not to lose focus especially on the circuit design details of the receiver. Moreover, the modifications required on the transmitter to enable power-on and off functionality are relatively straightforward, apart from the challenge of designing a high speed transmitter itself. Since it is physically located very close to the source of the data it does not need to sense a power-on or off event, an enable signal can be connected to it directly. Furthermore, it is already source-synchronous in most applications and does not require a CDR that has to find the phase of the incoming data.

1.2 Thesis Organization

The thesis is organized as follows: In Chapter 2 the definition of the basic terms that are widely used in the thesis are given and the basics of equalization techniques are introduced. In Chapter 3, the high speed optical receiver architecture for always on mode is presented and the details of the circuit design of each block are explained. Also the techniques used to improve the speed, sensitivity and CDR performance are discussed thoroughly. In Chapter 4, the link protocol required to run the rapid power on algorithm is described. Also, the circuit modifications done on the high speed optical receiver to enable rapid power-on and -off functionality are clarified. In Chapter 5, the measurement results of the implemented optical RX are provided. Finally, Chapter 6 concludes the study.

Chapter 2

Theory Review

The aim of this chapter is to introduce the basic techniques that are used to enhance the IO link efficiency and technical terms used for characterization of the IO links.

2.1 Technical Terms

A number of technical terms are used throughout this thesis to characterize the link performance (optical receiver performance to be more specific), which are defined here.

Data-rate, is the number of bits transmitted through the link in one second. It is given in terms of bits/s (bps).

Baud-rate, is the number of symbols transmitted through the link in one second. One signal may contain more than 1-bit of information.

Unit-interval, UI, is the time required to transmit a symbol (1/Baud-rate).

Non-return-to-zero, NRZ, is the form of transmission where binary states are represented by 2 specific levels, with no other neutral or rest condition.

Bit-error-rate, BER, is the ratio of the number of erroneous bits to the total number of bits transmitted.

Pseudo-random-bit-sequence, PRBS, is a deterministic bit sequence that exhibits similar statistical behavior to a truly random sequence. PRBS is extensively used to measure the link performance, as it is easily generated by very simple circuitry such as linear feedback shift registers. The general notation for PRBS sequences is PRBSk (or PRBS-k) where k is the word length of the PRBS sequence, such as PRBS7, PRBS15, and PRBS31. In general, the link performance drops as the PRBS length used in the measurement increases because the bandwidth of the test signal increases.

Inter-symbol-interference, ISI, is a form of distortion where one or more symbols interfere with other symbols reducing the signal integrity. The

main reasons for ISI are multi-path propagation and bandwidth limitation of the link.

Eye diagram, is a figure that is generated by overlapping segments of the signal with a period of a certain number of UIs. It gives indications on the ISI and channel noise.

Jitter, is the deviation of a periodic signal from its true periodicity in time domain.

Sensitivity, is defined as the smallest optical signal power which can be detected by the receiver while meeting the required BER specifications.

Extinction ratio, ER, is the ratio of the high level optical signal power to the low level optical signal power generated by an optical source, such as a laser diode. Its formula is given by the simple equation:

$$ER = \frac{P_1}{P_0} \tag{2.1}$$

where P_1 is the high level optical power and P_0 is the low level optical power.

Vertical-cavity surface-emitting laser, VCSEL, is a semiconductor laser diode with beam emission perpendicular to wafer surface. They are widely used in fiber optic communications.

Optical modulation amplitude, OMA, is the difference between the two optical power levels generated by an optical source:

$$OMA = P_1 - P_0 \tag{2.2}$$

solving the two equation Eq. (2.1) and Eq. (2.2) the relation between OMA and ER can be found as:

$$OMA = 2P_{AVG}\frac{ER-1}{ER+1} \tag{2.3}$$

where P_{AVG} is the average received optical power. This is a rather useful formula for sensitivity measurements. Once the ER is measured, the OMA can be calculated by measuring P_{AVG} only for different attenuation values as the attenuation does not change the ER. Thus it is not necessary to measure the optical signal lengths for both levels.

Responsivity, is the gain of a photo detector (or photo diode) from the optical input power to the electrical output signal (usually photo-current). It is expressed in terms of Amps/Watt.

2.2 Equalization Techniques

At high speed communication the signal degradation due to bandwidth limitation of the IO links lead to various equalization techniques to be introduced to cancel ISI and improve the transmitted signal. Although those techniques are mostly proposed to compensate for the effects of the channel response of the electrical links, they can be implemented for optical links to increase the performance such as the data-rate and sensitivity.

This section will explain the 3 of the most common equalization techniques and analyze their effects on the transmitted data.

2.2.1 Continuous Time Linear Equalization (CTLE)

CTLE is an analog equalization technique in which the bandwidth of the data-path is recovered by a high-pass filter.

The equalization is illustrated on an example in Fig. 2.1. Fig. 2.1a shows the frequency responses of the bandwidth limited signal, the CTLE filter, and the equalized signal. The signal has a first order response with a single pole at 1 Grad/s, and the high-pass CTLE has a zero at 1Grad/s and a pole at 10 Grad/s resulting in an equalized signal of 10 Grad/s bandwidth. The time domain pulse responses (pulse width = 1ns) of the bandwidth limited signal and equalized signal are given in Fig. 2.1b. And the corresponding 2-UI eye-diagrams of those two signals are given in Fig. 2.1c and Fig. 2.1d, respectively.

There are various circuits used to implement the CTLE filter. Two circuit implementations are given in Fig. 2.2a and Fig. 2.2b as examples for passive and active CTLE filters, respectively. The transfer functions of the passive circuit (Fig. 2.2a) is written as:

$$TF_{PAS} = \frac{OUT}{IN} = \frac{R_2}{R_1+R_2}\frac{1+sC_1R_1}{1+s(C_1+C_2)\frac{R_1R_2}{R_1+R_2}} \tag{2.4}$$

Assuming $C_2 << C_1$ and $R_2 << R_1$ the transfer function simplifies to:

$$TF_{PAS} = \frac{OUT}{IN} = \frac{R_2}{R_1}\frac{1+sC_1R_1}{1+sC_1R_2} \tag{2.5}$$

resulting in a dc gain of less than R_2/R_1 and a maximum gain of 1 at high frequencies.

The transfer function of the active circuit is (assuming $g_{ds} = 0$):

$$TF_{ACT} = \frac{O_P - O_N}{IN_P - IN_N} = \frac{2g_mR_L}{2+g_mR_S}\frac{1+sC_SR_S}{(1+sC_S\frac{R_S}{2+g_mR_S})(1+sR_LC_L)} \tag{2.6}$$

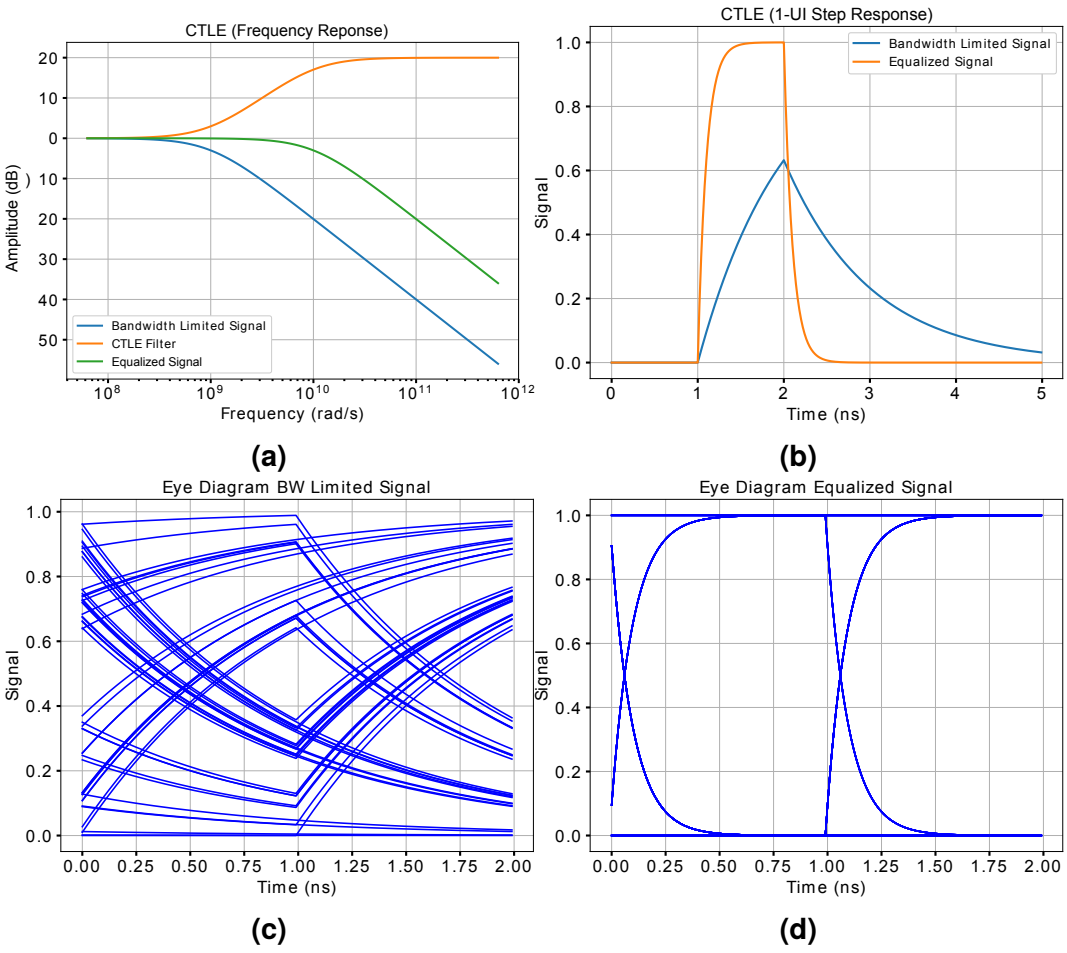

Figure 2.1: CTLE example: (a)Frequency response (b) Pulse response (c)Eye diagram of the bandwidth limited signal (d)Eye diagram of the equalized signal.

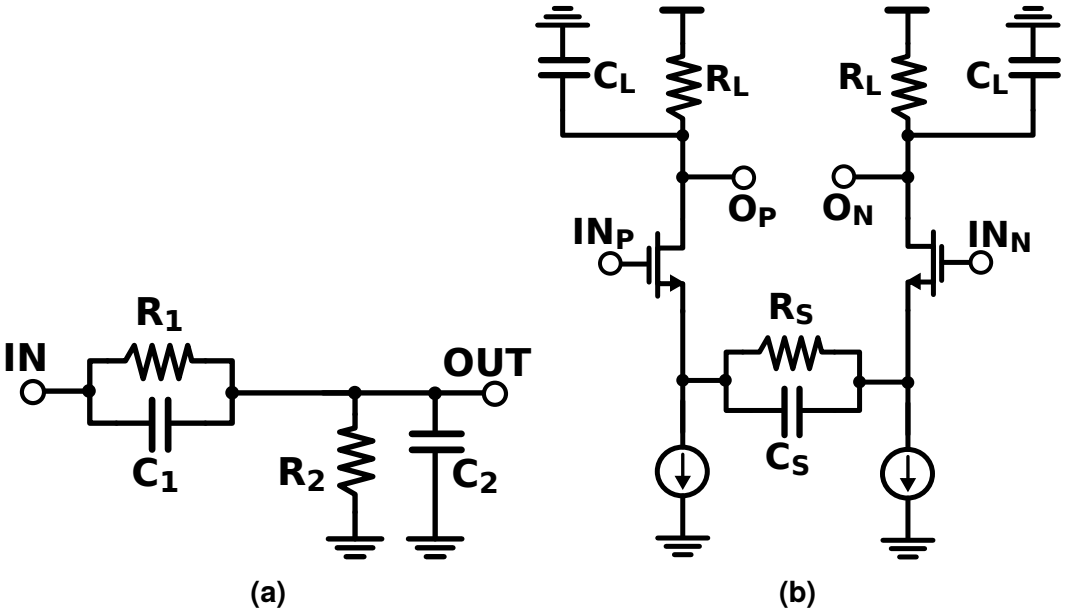

Figure 2.2: CTLE circuit example: (a)Passive (b)Active

where g_m is the transconductance of the input transistors. Assuming $g_m R_S >> 2$, the transfer function simplifies to:

$$TF_{ACT} = \frac{O_P - O_N}{IN_P - IN_N} = \frac{2R_L}{R_S} \frac{1 + sC_S R_S}{(1 + sC_S \frac{1}{g_m})(1 + sR_L C_L)} \tag{2.7}$$

resulting in a dc gain of $2R_L/R_S$, and a maximum gain of $R_L g_m$. Various other circuit topologies and a more detailed analysis on CTLE can be found in [17].

As found in Eq. (2.5) and Eq. (2.7), the peaking characteristics of the CTLE are, most of the time, defined by the values of resistances and capacitors placed in the circuit. In order to equalize channels with various attenuation characteristics, the value of those elements has to be controllable. Usually this control is achieved by putting many capacitors or resistors in parallel and connecting as many of them as required via switches. However those switches come with their own parasitic elements that generate unwanted poles and zeros in the actual transfer function and degrade the CTLE performance. The switches connected to "flying" elements such as C_S and R_S in Fig. 2.2b are specifically difficult to realize with small parasitics.

Another well-known characteristic of the CTLE is that it boosts the high frequency noise as well as the signal. Thus while improving the signal by canceling ISI, it may result in increased the noise.

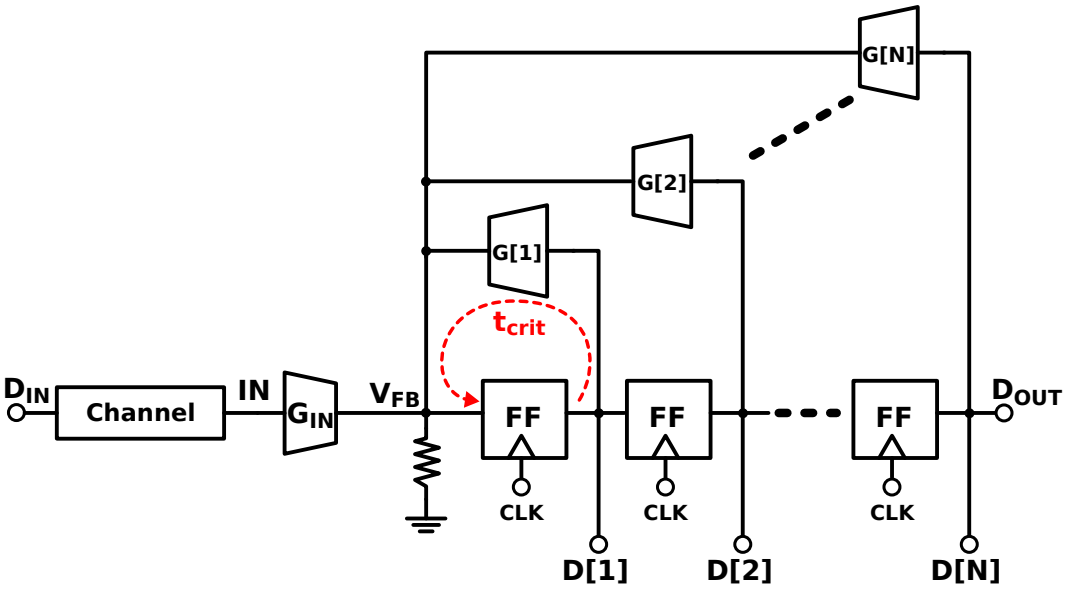

Figure 2.3: N-tap DFE conceptual block diagram

2.2.2 Decision Feedback Equalization (DFE)

The action of the DFE is to feed back a weighted sum of past decision(s) to cancel the ISI they cause in the present signal ([18]). The conceptual architecture for N-tap DFE is given in Fig. 2.3. The digital data D_{IN} is transmitted through a channel creating an analog input voltage of IN. The analog signal V_{FB} is sliced by the flip-flop to generate a digital decision $D[1]$. N previous decisions are stored in the pipeline and they are fed back into the summing node V_{FB} with their respective weights $G[k]$. As a result the voltage at V_{FB} is expressed as:

$$V_{FB} = R(V_{IN}G_{IN} + \sum_{k=1}^{N} G[k]D[k]) \tag{2.8}$$

where N is the total number of DFE taps used for equalization.

The eye diagrams of the analog signal in node V_{FB} are given in Fig. 2.4 for $0-3$ taps DFE where the channel is modeled by a first order low-pass filter. Without DFE, the eye diagram is marginally open and as the number of DFE taps increases the eye opening increases. However, the contribution from the later DFE taps gets smaller and smaller. As the circuit complexity and power consumption increases with each additional DFE tap, the minimum number of taps that satisfies the channel specifications should be chosen for the design.

At high data rates DFE implementation becomes challenging because the feedback loop latency must be less than 1-UI for the first post cursor ISI cancellation. This critical timing path is shown in Fig. 2.3 with the red dashed line,

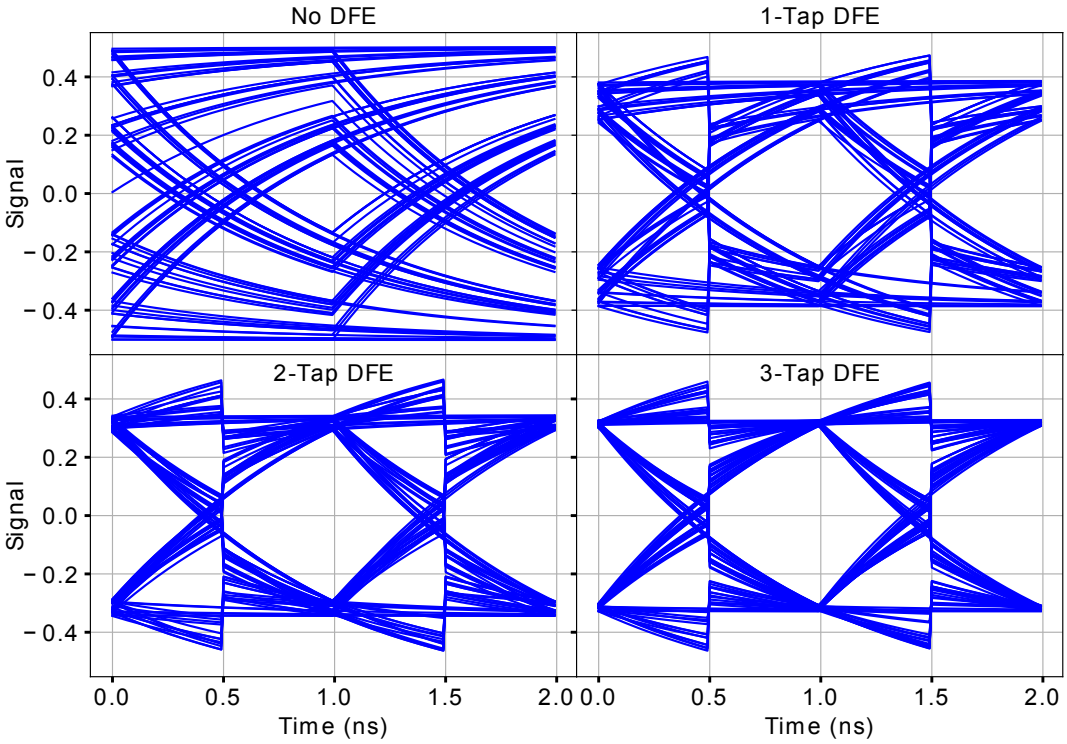

Figure 2.4: Eye diagrams after DFE

t_{crit}. And the timing should satisfy the condition:

$$t_{cq} + t_{setup} + t_{G[1]} < UI \tag{2.9}$$

where t_{cq} and t_{cq} are the clock-to-Q delay and setup time of the slicer (FF), respectively, and $t_G[1]$ is the delay of the feedback element $G[1]$. The higher DFE taps ($G[2] - G[N]$) are less timing critical since the clocks of the flip-flops that create the digital signals ($D[2] - D[N]$) for those taps can be shifted by a delay t_{shift} to relax their timing constraints to:

$$t_{cq} + t_{setup} + t_{G[k]} < UI + t_{shift} \tag{2.10}$$

Speculative DFE architectures (also called loop-unrolled) such as the example given in Fig. 2.5 slightly relaxes the timing constraint of the first DFE tap to:

$$t_{cq} + t_{setup} + t_{sq,MUX} < UI \tag{2.11}$$

where $t_{sq,MUX}$ is the select-to-Q delay of the multiplexer (MUX), and is usually smaller than $t_{G[1]}$ in Eq. (2.9) ([19]).

Unlike CTLE, DFE does not increase the noise. However, the errors in the previous decisions tend to propagate because the ISI components that depend on

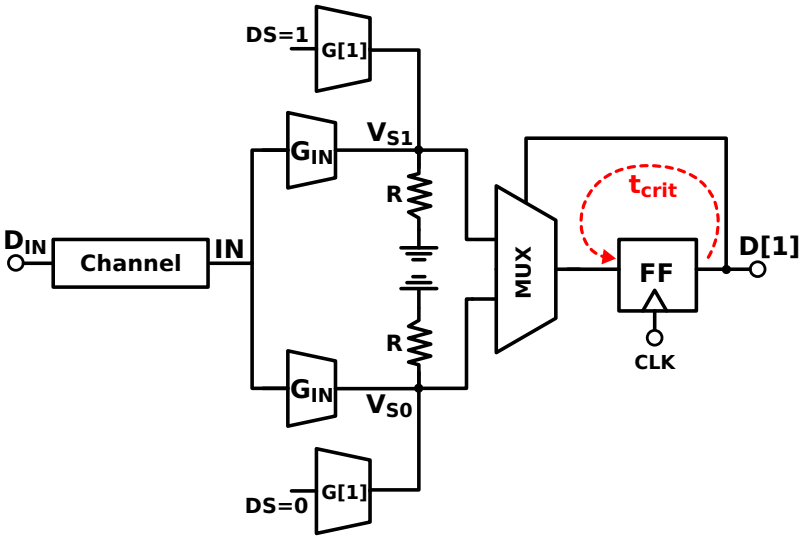

Figure 2.5: 1-tap speculative DFE

the erroneous decision(s) are calculated wrong resulting in reduced noise margin for the future decisions ([20]). This phenomenon is called 'error propagation'. In general, this effect increases as the total number of DFE taps and the weights of the taps increases.

2.2.3 Feed-forward Equalization (FFE)

This technique employs a finite impulse response (FIR) filter with a series of tap weights programmed to adjust the impulse and, by duality, frequency response. The application of this technique on RX side requires analog samplers and analog addition or subtraction operations which are cumbersome to design and vulnerable to noise or analog-to-digital converters which are power hungry for high-speed operations. On the other hand it is relatively easy to apply this technique on transmitter side.

The block diagram for an N-tap FFE on transmitter is given in Fig. 2.6. And its transfer function can be written as:

$$V_{OUT} = R \sum_{k=0}^{N} G[k]D[k] \tag{2.12}$$

Generally the coefficients $G[k]$ are selected so as to de-emphasize low frequency components reducing the low frequency signal envelope level in proportion to the attenuation experienced at high frequency [21].

The effect of a 1-tap FFE on the pulse response of a 1^{st} order low-pass filter is given in Fig. 2.7, and the corresponding eye diagrams for the cases with and without FFE is given in Fig. 2.8. From the figures it is straightforward to deduce

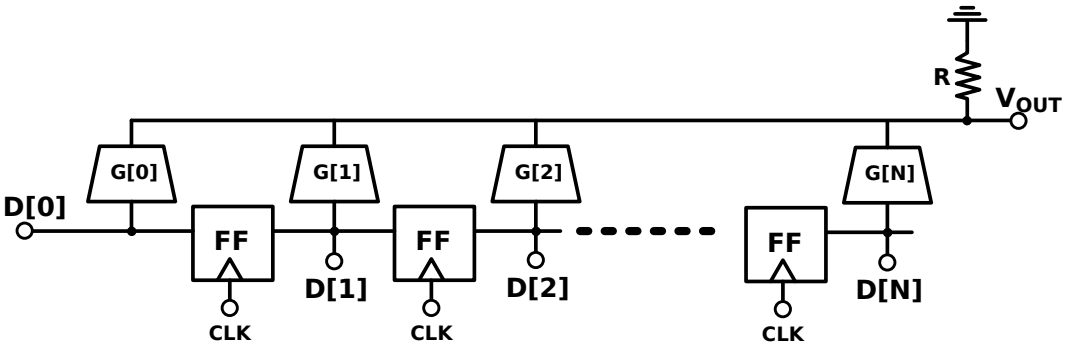

Figure 2.6: N-tap FFE block diagram

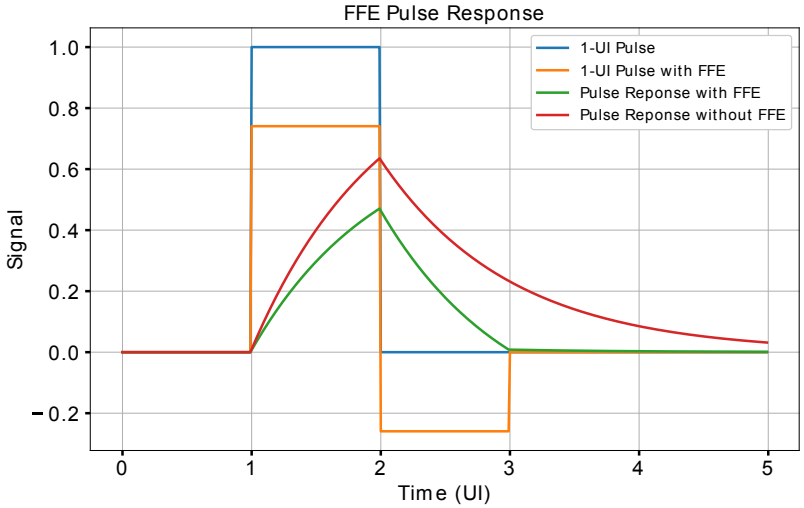

Figure 2.7: FFE pulse response

that, unlike DFE, a 1-tap FFE can cancel multiple ISI components.

Another advantage of FFE over DFE is that it can also cancel the precursors which appear with higher order channel responses. This is possible since the digital data to be transmitted is already known in the TX whereas only the past decisions are known on the receiver side.

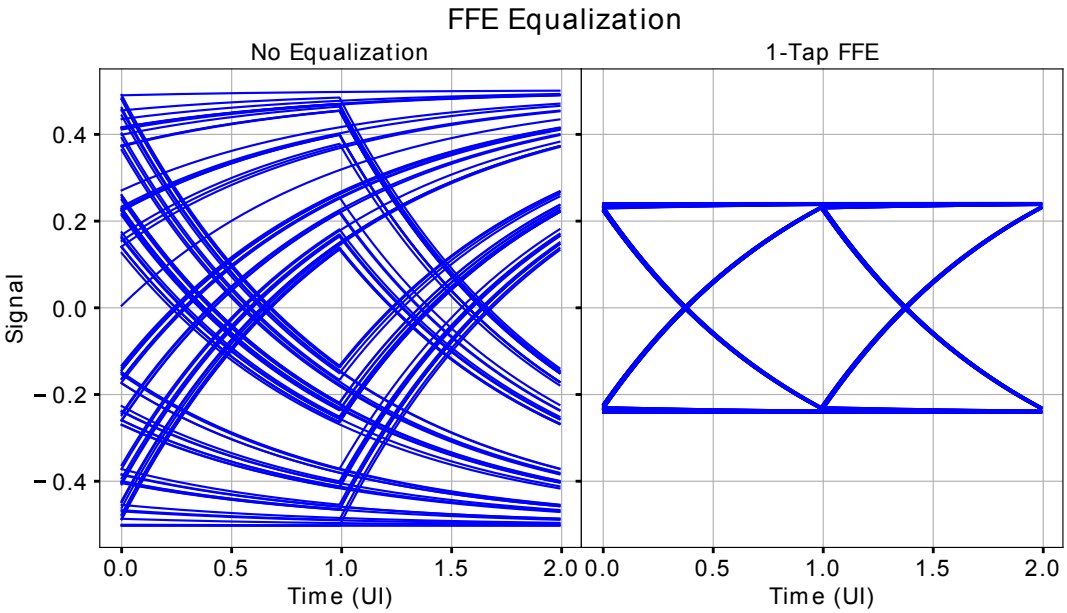

Figure 2.8: FFE Eye diagram

Chapter 3

High Speed Optical RX

This chapter presents a 60 Gb/s optical receiver including the digital clock and data recovery circuit in 14 nm FinFET CMOS technology, achieving approximately −8 dBm optical OMA with an energy efficiency of < 2 pJ/b. It uses an active silicon area of around 150 μm by 150 μm enabling it to be used for multi-channel applications.

The chapter is organized as follows. In Section 3.1 the top level block diagram of the optical receiver is introduced and the functions of the implemented building blocks are described briefly to provide the reader with a general overview of the RX. Then, in Section 3.2 the blocks on the data path are analyzed and their circuit implementations are shown. The design choices to maximize speed and sensitivity while maintaining low power operation are explained. Finally, in Section 3.3 the clock and data recovery (CDR) circuit is described in detail. Moreover, the parameters defining the CDR performance are analyzed using a top level CDR model implemented in Python software.

3.1 Receiver Architecture

The top level block diagram of the RX is shown in Fig. 3.1. The dc portion of the input current signal from the photo diode is canceled by two blocks: a 12-bit resolution current output digital-to-analog converter to cancel the coarse DC component and an operational trans-admittance amplifier (OTA) connected on the feedback path of the AFE to cancel the residual DC current. The ac component is converted into a voltage signal via the transimpedance amplifier (TIA) to be processed by the rest of the circuit. High coupling losses on the optical path may reduce the input optical signal to such levels where only few 10s of μA current can be generated on the photo-diode as input to the RX. Concerning the bandwidth requirements for high speed design, the TIA gain is unlikely to provide enough voltage output to be properly sampled by the comparators. Thus a variable gain amplifier (VGA) is used to amplify the input signal and drive the comparators.

At the target data-rate of 60 Gb/s, it is extremely difficult and power hungry to do the sampling at full-speed even in the cutting edge 14nm technology. Hence a time-interleaving strategy is implemented to reduce the maximum clock-speed

and relax the timing requirements for the samplers. In the given technology, a time-interleaving factor of 4X gives the optimal balance between power consumption and circuit complexity. As a result, the analog signal at the output of analog front-end (AFE) is sampled by 4-way time-interleaved data and edge comparators to provide the signal and phase information for 2X oversampled CDR. Each data sample consists of two comparators to generate speculative decisions for 1-tap DFE. After that, all the signals are aligned to a single quarter rate clock. And, the speculative decisions (D_H and D_L) are resolved in the look ahead DFE generating the output data signals D. Edge signals (E') are delayed by the same amount as look ahead DFE to keep D and E signals synchronized. Then, the output data (D) is sent into an on-chip pseudo random bit sequence (PRBS) checker via a 4-to-32 demultiplexer (DEMUX) to measure the bit error rate (BER).

The CDR logic block receives the data (D) and the edge (E) signals to detect the phase information and it generates the gated clock ($c4_G$) and up-down (U_D) signals that drive the phase rotator control block (PR_C). The PR_C generates the control signals for the 128-step phase rotator (PR) itself. The PR also receives in-phase (I) and quadrature (Q) clocks from a frequency divider and generates an output clock consisting of differential signals C_p and C_n depending on the digital control signals. Then, a CML based IQ generator generates 8 signals with nominal phase apertures of 45° that correspond to data and edge phases in quarter rate sampling. Finally the phases of the sampling clocks are adjusted in the IQ Calibration block and the clocks are converted into CMOS levels via CML-to-CMOS converters (CML2CMOS).

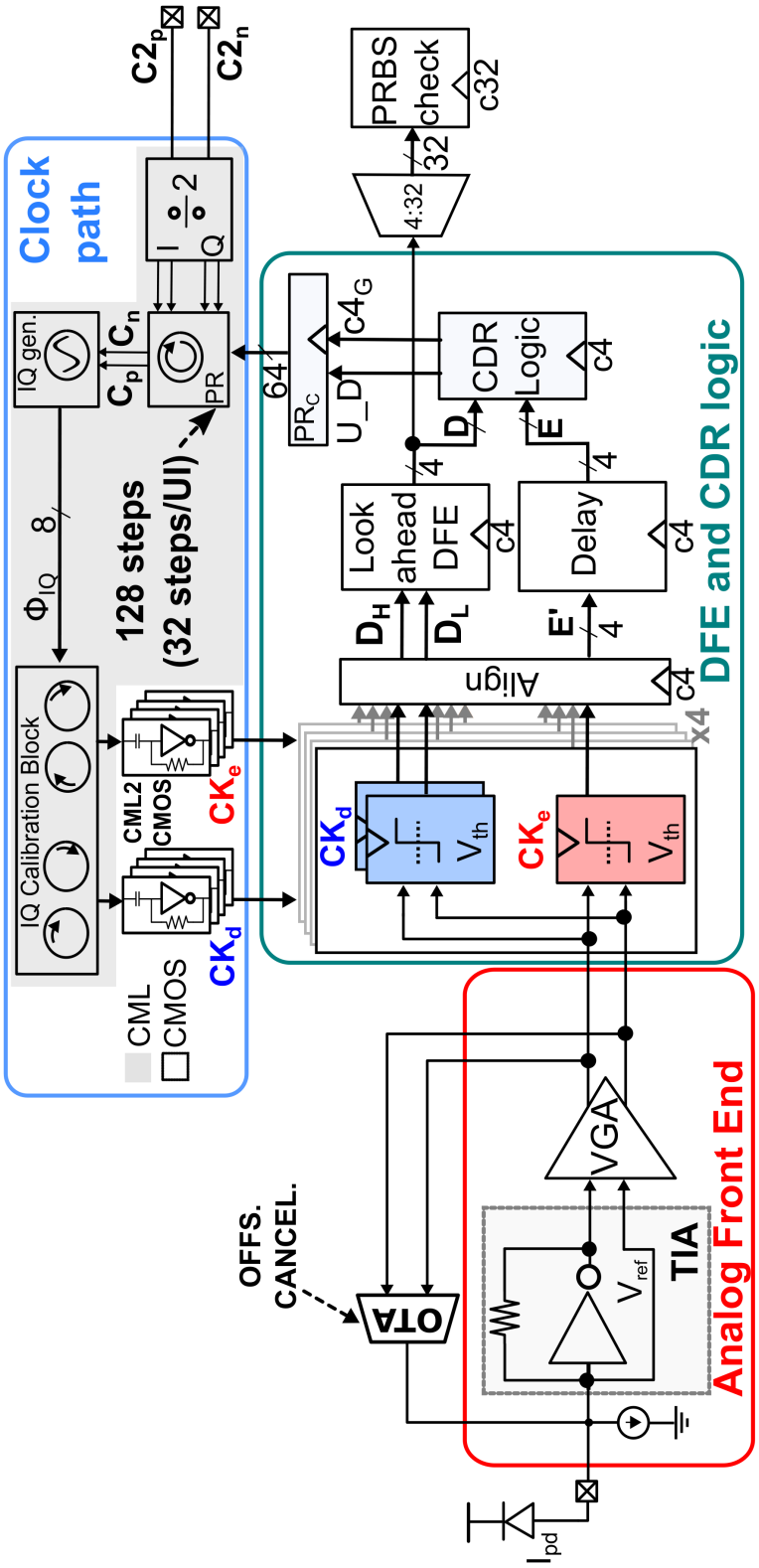

Figure 3.1: RX top level block diagram

3.2 Data Path [1]

Data path consists of the AFE, comparators, aligner and DFE blocks in this implementation, as illustrated in Fig. 3.1. In Section 3.2.1 and Section 3.2.2 the design techniques implemented in this work to maximize the data path performance will be discussed in detail.

3.2.1 AFE Design

The overall sensitivity of an optical link is determined by the RX AFE. The transimpedence amplifier (TIA) that converts the input current coming from the photo-diode into voltage limits the SNR performance. Thus, an in depth signal and noise analysis of the TIA will be provided in this section. First, the choice of TIA architecture will be presented. Then, the optimum SNR in the absence of equalization will be provided for that TIA arcitecture. Finally, the SNR improvement by optimizing TIA bandwidth according to the number of taps to be used in DFE will be explained.

In principle, a simple resistor connected to ground or a common-mode voltage can be used to convert the current of a PD to a voltage. However, the large capacitance associated with pad and PD severely limits the bandwidth. Hence, to improve gain-bandwidth product (GBW), active structures are commonly employed. Regulated cascode (RGC) [22] and shunt feedback resistor (SFR) are the most common TIA circuit topologies [23] (Fig. 3.2). Both RGC and SFR architectures have been compared in 14 nm FinFET technology. The two designs were optimized for maximum SNR at 60 Gb/s for a given DC gain (50 dB in this case). Our investigations showed that while RGC and SFR TIAs have comparable GBW products and power dissipation, SFR TIA has better SNR since the output integrated noise of the RGC is approximately 20% higher. The main reason stems from the noise generated by the bias current source, as described in [23]. It must be noted that the DC current source which is used to cancel the average PD current is much smaller (a few hundred μA) than the bias current required for a high bandwidth RGC-TIA (at least 2-3 mA), resulting in much smaller noise contribution. Moreover, the average PD current must be subtracted from both SFR and RGC TIAs, which means the current that biases the RGC-TIA is an extra noise source. Furthermore, in advanced technology nodes operating at low supply voltages (below 1 V) RGC design is challenging due to limited voltage headroom. Thus, the SFR topology was used for the optical RX.

[1] This section is based on: I. Ozkaya, A. Cevrero, P. A. Francese, C. Menolfi, T. Morf, M. Brandli, D. M. Kuchta, L. Kull, C. W. Baks, J. E. Proesel, M. Kossel, D. Luu, B. G. Lee, F. E. Doany, M. Meghelli, Y. Leblebici, and T. Toifl. "A 64-Gb/s 1.4-pJ/b NRZ Optical Receiver Data-Path in 14-nm CMOS FinFET". IEEE Journal of Solid-State Circuits 52.12 (Dec. 2017), pp. 3458-3473.

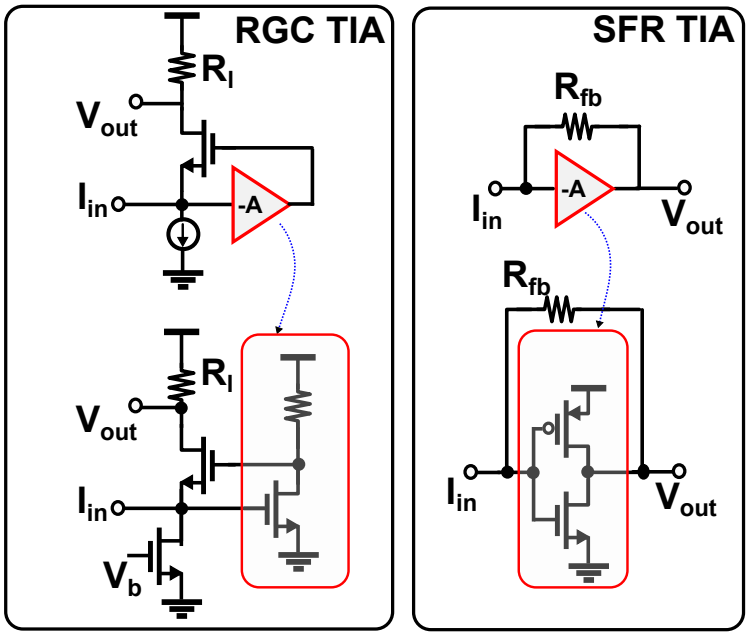

Figure 3.2: RGC and SFR TIA topology and circuit implementation.

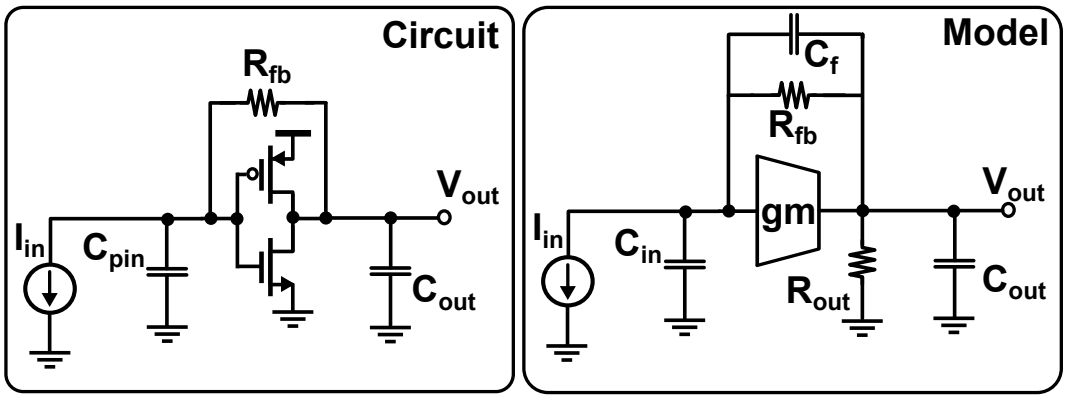

Figure 3.3: SFR TIA circuit diagram and small signal model.

Signal Analysis

An SFR TIA together with its model is shown in Fig. 3.3. In the circuit C_{pin} and C_{out} are the sum of PD capacitance and pad capacitance, and the load capacitance driven by TIA, respectively. In the model, C_{in} includes the gate-to-source capacitances (C_{gs}) of the transistors in addition to C_{pin}. C_f consists of the drain-to-gate capacitance (C_{gd}), whereas R_{out} is the equivalent output resistance of the inverter. The total transconductance of the CMOS inverter is denoted as gm. The transimpedance of the TIA can be expressed in terms of the given parameters as:

$$Z_t(s) = \frac{R_{out}(gmR_{fb} - 1 - sC_fR_{fb})}{1 + gmR_{out} + sD_{t1} + s^2D_{t2}} \tag{3.1}$$

where

$$D_{t1} = C_{in}(R_{out} + R_{fb}) + C_f R_{fb}(1 + gmR_{out}) \tag{3.2}$$

$$D_{t2} = R_{fb}R_{out}C_{in}(C_{out} + C_f) \tag{3.3}$$

The number of parameters in this equation can be reduced since some of them are coupled together by the technology node. We can write:

$$A = gmR_{out} \tag{3.4}$$

$$f_t = \frac{gm}{2\pi C_{gate}} \tag{3.5}$$

where A is the intrinsic gain and f_t is the transit frequency of the transistors. For a given biasing condition ($V_{DD}/2$ in this case) A and f_t are known. Note that C_{gate} is the total gate capacitance and transistor level simulations show that it can be distributed between C_{gs} and C_{gd} with a ratio of 2/3 and 1/3. C_{gd} corresponds to C_f, whereas C_{gs} contributes to the input capacitance, C_{in}. It must be emphasized that the inclusion of C_f is extremely important due to Miller effect. Omitting this capacitance would result in oversimplified models which may invalidate the further analysis.

We can further reduce the number of parameters by fixing the value of C_{pin}, which is determined by the pad and PD capacitances. In the implemented design, the PD and pad capacitance were approximately 60 fF and 40 fF, respectively. Hence, C_{pin} = 100fF. In the given 2-pole system a numerical analysis shows that a high C_{out}/C_{pin} ratio results in peaking in the transfer function. For a maximally flat response, a ratio of less than 0.25 must be satisfied. Therefore, C_{out} is taken as 25 fF and it defines the input capacitance of the following stage.

As a result, the full design space can be defined by two parameters: R_{fb} and gm. In Fig. 3.4, R_{fb} is swept for three different gm values to find the 3 dB bandwidth of the TIA. Since the curves are monotonic, any given bandwidth and gm pair corresponds to a unique R_{fb} value. Figure 3.5 depicts gm versus R_{fb} curves with 4 different TIA bandwidths. The plot clearly shows that for constant bandwidth, R_{fb} needs to be reduced towards large gm. This is due to self loading. The parasitic feedback capacitor C_f increases due to large transistor size and $R_{fb}C_f$ becomes the dominant pole.

Figure 3.6(a-e) show main cursor ($V_{Tap(0)}$) together with pre- and post-cursors at 64 Gb/s across gm, R_{fb} pairs on a constant bandwidth curve. The main cursor and inter-symbol-interference (ISI) components are derived from the pulse response (1 μA normalized current pulse applied) of the system described by Eq. (3.1). As expected, lowering the bandwidth increases the main cursor due to

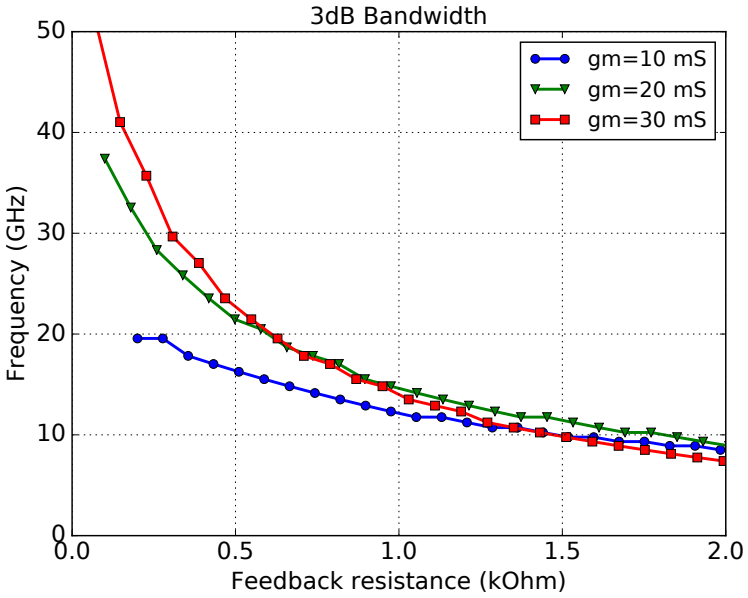

Figure 3.4: TIA 3-dB bandwidth for various gm.

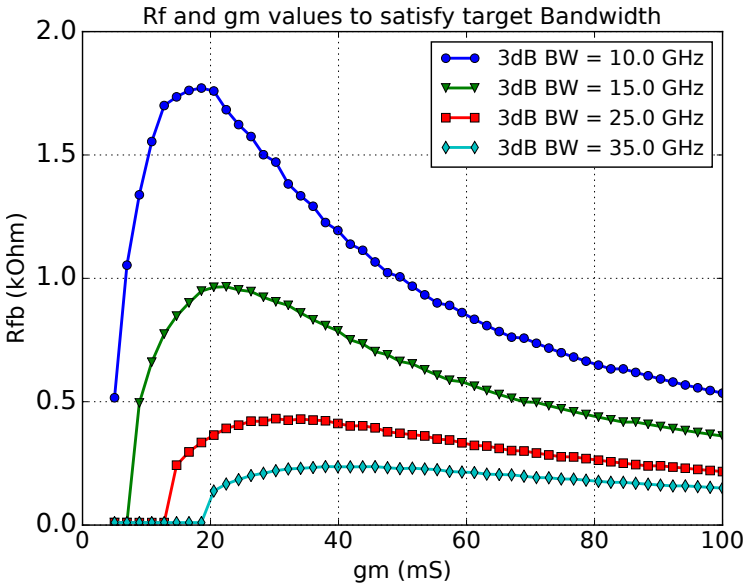

Figure 3.5: gm and R_{fb} values for a target TIA bandwidth.

larger transimpedance gain, which is approximately equal to R_{fb}. At the same time, bandwidth limitation increases ISI terms. The worst case signal without equalization can be calculated as follows:

$$V_{Signal} = V_{Tap(0)} - |V_{Tap(-1)}| - \sum_{n=1}^{\infty} |V_{Tap(n)}| \tag{3.6}$$

as shown in Fig. 3.6f for 5 different TIA bandwidths. Without equalization the optimal bandwidth is approximately 0.4 times the data-rate(25 GHz at 64 Gb/s). Interestingly for a given bandwidth the point with the largest V_{signal} doesn't correspond to the gm with the largest R_{fb} due to peaking in the 2nd order system.

Noise Analysis

The analysis above gives a perspective on the signal point of view only. In order to find the TIA specifications that provides optimum SNR, the noise characteristic must also be analyzed.

The main noise sources of the implemented TIA are the transconductance stage (CMOS inverter) and R_{fb} thermal noises, which are shown on the TIA model in Fig. 3.7a as I^2_{ngm} and I^2_{nR}, respectively. The power spectral densities (PSD) can be expressed as:

$$I^2_{ngm} = 4kT\gamma gm \tag{3.7}$$

$$I^2_{nR} = \frac{4kT}{R_{fb}} \tag{3.8}$$

where γ is the noise excess factor of the transistor. It must be noted that the noise generated by the output resistance (R_o) of the inverter can be omitted since its PSD is a factor of gmR_{out} smaller than the I^2_{ngm}.

In order to simplify the equivalent output noise calculation, we can split the I^2_{nR} noise source as proposed in [24] (Fig. 3.7b). The output squared noise can be expressed as:

$$V^2_{nR} = \int_0^{\infty} I^2_{nR} |Z_t + Z_o|^2 \delta f \tag{3.9}$$

$$V^2_{ngm} = \int_0^{\infty} I^2_{ngm} |Z_o|^2 \delta f \tag{3.10}$$

$$V^2_{nout} = V^2_{nR} + V^2_{ngm} \tag{3.11}$$

where Z_t is the transimpedance of the TIA expressed in Eq. (3.1), and Z_o is the output impedance of the TIA, which can be calculated as:

$$Z_o(s) = \frac{R_{out}(1 + sR_{fb}(C_f + C_{in}))}{1 + gmR_{out} + sD_{o1} + s^2D_{o2}} \tag{3.12}$$

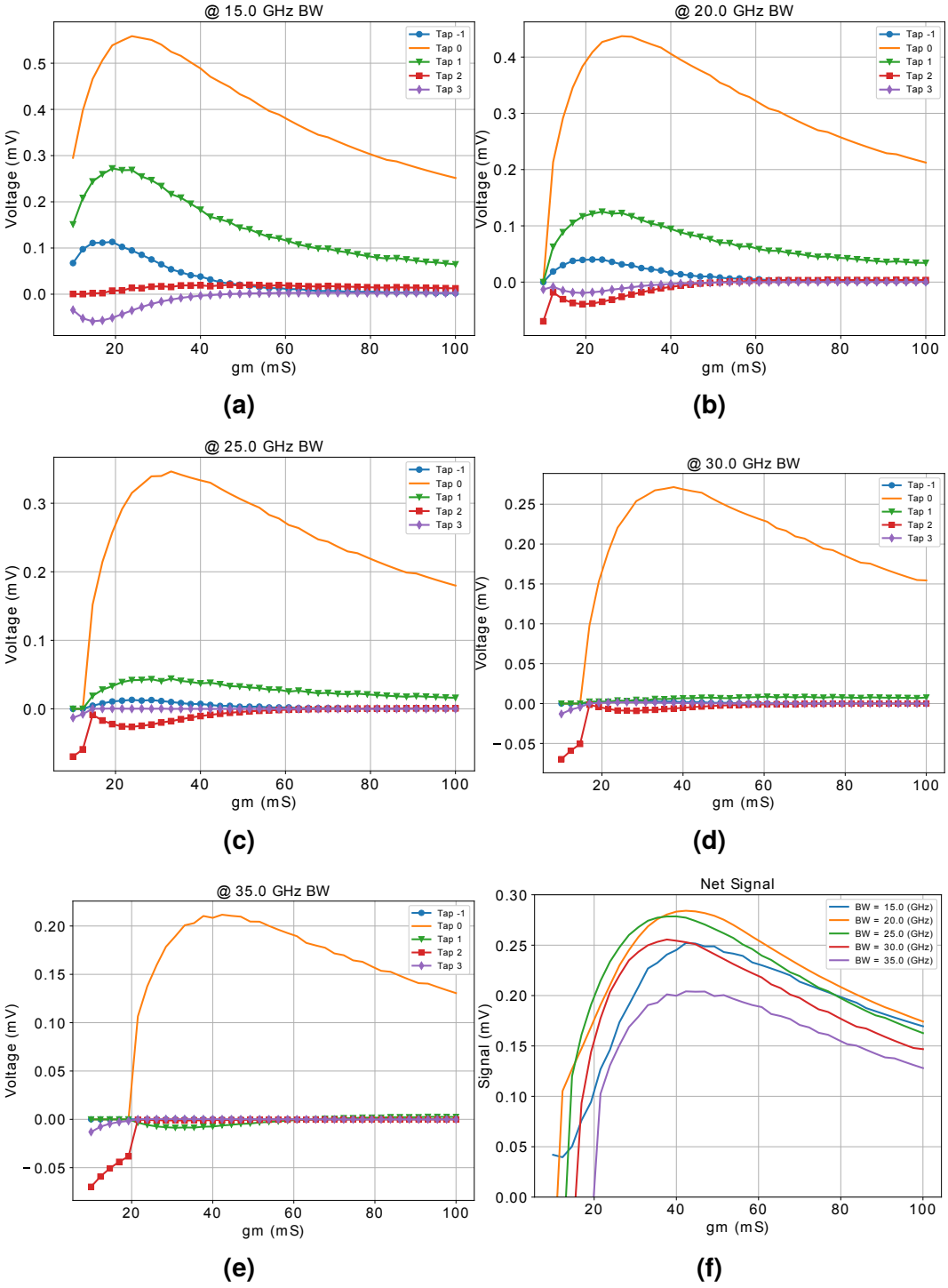

Figure 3.6: Tap sizes for different TIA bandwidths (a-e) and net signal (f).

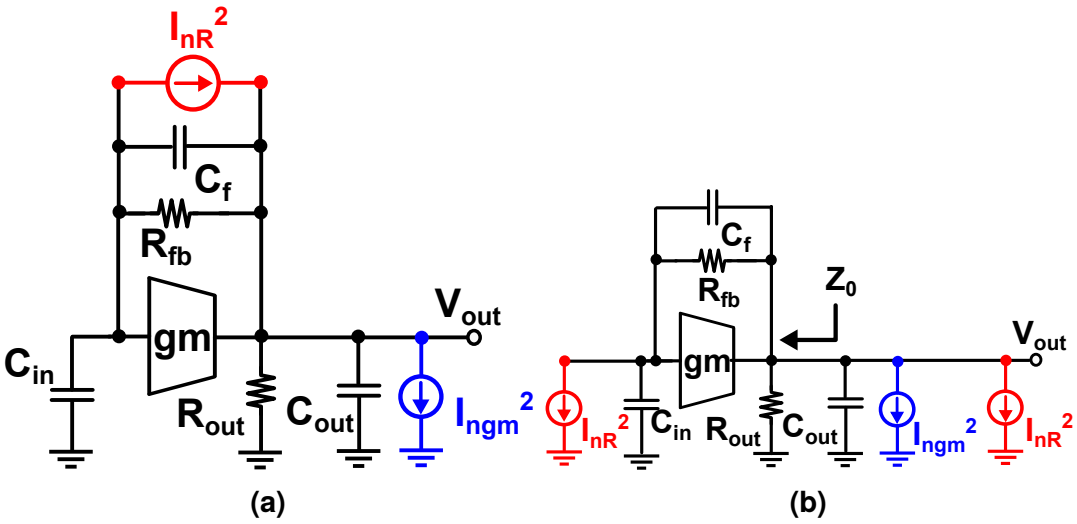

Figure 3.7: (a) Noise sources in TIA (b) Equivalent split model.

where

$$D_{o1} = C_{in}(R_{out} + R_{fb}) + C_{out}R_{out} + C_f R_{fb}(1 + gmR_{out}) \tag{3.13}$$

$$D_{o2} = R_{fb}R_{out}(C_{in}C_{out} + C_{in}C_f + C_f C_{out}) \tag{3.14}$$

Thus, for any gm and R_{fb} we can calculate V_{nout} and find the signal to noise ratio under worst case ISI condition (SNR_{wc}) as follows:

$$SNR_{wc} = \frac{V_{Signal}^2}{V_{nout}^2} \tag{3.15}$$

Fig. 3.8 shows SNR_{wc} for different TIA bandwidths. The figure clearly demonstrates the optimum SNR_{wc} corresponds to roughly 0.4 times the data-rate (25 GHz for 64 Gb/s NRZ). Note that SNR_{wc} given in Fig. 3.8 is normalized to 1 μA input. One can easily calculate the optical sensitivity by using the equation:

$$Sens = dBm\left(\frac{2N}{\sqrt{SNR|_{1\mu A}}Res}\right) \tag{3.16}$$

where N is the number of standard deviations (σ) required to reach certain bit-error-rate (BER) (for 10^{-12} BER $N = 7$), and Res is the responsivity of the photo diode in terms of μA/W.

Equalization

The SNR of an optical RX can be further improved by lowering the bandwidth (which increases DC gain) below 0.4 times the data-rate and using equalization techniques to recover the bandwidth loss, provided that the noise added by the

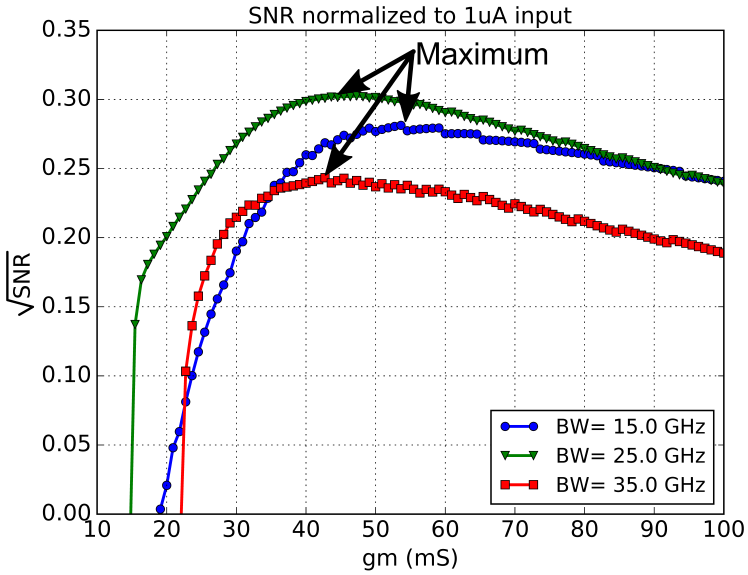

Figure 3.8: SNR_{wc} for 3 different bandwidths without equalization.

equalizer is small enough. For instance, in [25] a continuous time linear equalizer (CTLE) is cascaded to a low bandwidth TIA to achieve a sensitivity of −8 dBm OMA with a PD of 0.45 A/W at 25 Gb/s. When projected to 64 Gb/s design this approach has several shortcomings:

1. At high speed pushing the CTLE bandwidth to high frequency is power hungry and often requires the use of passive inductors which are undesired in compact multi-channel optical RX design.

2. At high frequency the CTLE has limited capability to compensate for the multiple poles in the signal path.

3. CTLE amplifies high frequency noise as well as generating noise itself degrading SNR compared to an ideal equalizer where only ISI is canceled.

In [26] and [13] a low bandwidth front-end is combined with double sampling and dynamic offset modulation to achieve −4.7 dBm sensitivity at 24 Gb/s and −6.8 dBm sensitivity at 25 Gb/s, respectively. [26] Uses a resistor to convert the photo-current to a voltage while in [13] the resistor is replaced by an SFR TIA leading to higher sensitivity. The drawback of this technique at higher speeds is the difficulty of driving the sample and hold capacitors. Moreover, transient effects such-as kickback noise are expected to become more and more important as the timing between the two consecutive samples shrinks.

Decision feedback equalization (DFE) is another well-known technique, which has the capability to remove post-cursor ISI with small or no noise penalty. Infinite-impulse-response (IIR) DFE approximates a long tail of the pulse re-

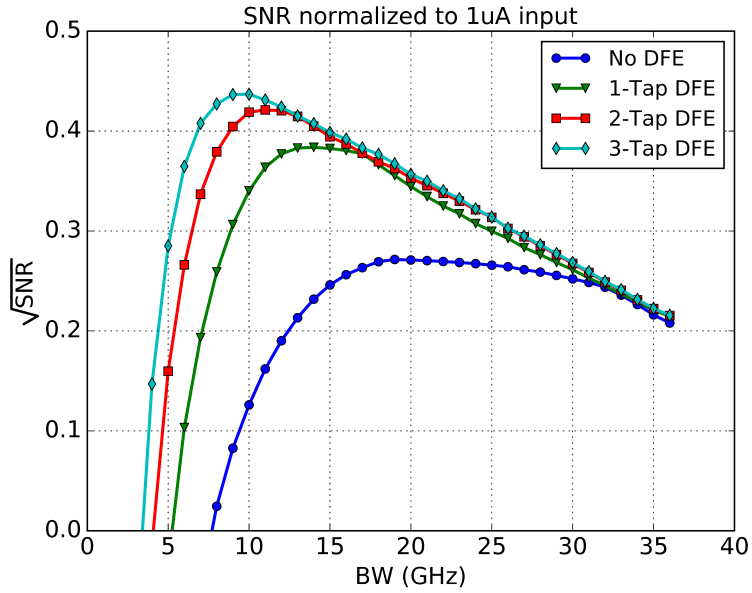

Figure 3.9: Maximum SNR for n-tap DFE.

sponse using a passive RC circuit as feedback filter, and subtracts the approximated tail from the input signal, removing all the post cursors with a single tap. However, the pulse response characteristics of the passive RC network should match those of the TIA for accurate ISI cancellation. This restricts the design of the TIA to first-order circuits. A more important limitation comes from the requirement that the total feedback delay must be less than 1 unit-interval(UI). [27] and [22] implemented this technique to achieve -5 dBm sensitivity at 9 Gb/s and -5.8 dBm sensitivity at 20 Gb/s, respectively.

A more conventional way of implementing the DFE is adding an FIR filter in the feedback path. Assuming the first m post cursors are equalized by the DFE, the signal after equalization can be expressed as follows:

$$V_{Signal_{Eq}} = V_{Tap(0)} - |V_{Tap(-1)}| - \sum_{n=m+1}^{\infty} |V_{Tap(n)}| \tag{3.17}$$

Fig. 3.9 shows SNR plots versus TIA bandwidth for different number of DFE taps based on the model provided above. The figure shows the maximum achievable SNR and the required TIA bandwidth. Without equalization, the optimal bandwidth is $20-25$ GHz. A 1-tap DFE lowers the TIA bandwidth for maximum SNR down to 15 GHz while improving SNR by approximately a factor of $\sqrt{2}$ which corresponds to a 1.5 dB sensitivity improvement. Additional DFE taps provide marginal SNR gain, while increasing the power cost of the underlying circuit implementation , as will be discussed in Section 3.2.2.

Similar analysis can be found in [22] and [28] leading to similar conclusions. However, the circuit model in [22] omits parasitic feedback capacitance (C_f) and

isolates the output node from feedback resistor R_{fb} via an ideal buffer neglecting the loading effect of R_{fb}. Moreover, in both publications the zero and pole locations of noise transfer function were set based on certain assumptions. Although the analysis provides good insight for certain conditions, it does not derive the complete analytical solution. In this study, all the equations are derived directly from any given set of parameters without simplifying the TIA model, which covers a larger design space.

DFE can be implemented either as direct feedback or as speculative DFE. On the one hand, direct feedback DFE enables many taps to be equalized with relatively low complexity. But the feedback loop delay still needs to be less than 1 UI, making this solution unattractive for the target data-rate of 60 Gb/s. On the other hand, in speculative DFE implementation, complexity grows exponentially with the number of taps. Nevertheless, the timing restriction can be relaxed using certain techniques as will be explained in Section 3.2.2. Since more than 1-tap DFE gives only marginal advantage in terms of SNR while increasing circuit complexity and power consumption significantly, we decided to use a 1-tap speculative DFE.

TIA

The schematic of the proposed SFR TIA is given in Fig. 3.10a. The feedback path is composed of a 1.1 kΩ resistor and NMOS transistors in parallel to adjust the equivalent resistance down to 250 Ω. Since the signal swing is small the transistors stay in linear region behaving as linear resistances. This approach reduces the parasitic capacitance as compared to a solution which consists of an array of passive resistors with switches. This is because typically passive resistors have larger parasitics than transistors. Moreover, the switches need to be large enough to minimize the on resistance, further increasing the area and capacitive load. A series inductance is added to extend the bandwidth of the TIA. TIA transimpedance as a function of the series peaking inductance value is plotted in Fig. 3.11a together with the pulse response in Fig. 3.11b. A 400 pH inductance provides a maximally flat response which corresponds to minimum group delay.

In this design, the input node of the TIA is used as the negative output ($Vout_n$) to serve as a differential signal to the output node of the TIA ($Vout_p$), rather than placing a replica TIA to generate a reference voltage [29], as shown in Fig. 3.10b. As a result, the transimpedance gain becomes R_{fb} instead of $R_{fb}(A_{eq}/(A_{eq}+1))$ where A_{eq} is equal to $gm(R_{fb} \parallel R_{out})$. This improvement is shown in Fig. 3.12. Both single ended outputs and the differential voltage ($V_{outp} - V_{outn}$) are given in the figure. Note that ($V_{outp} - V_{outn}$) is shifted to the right in order to match the

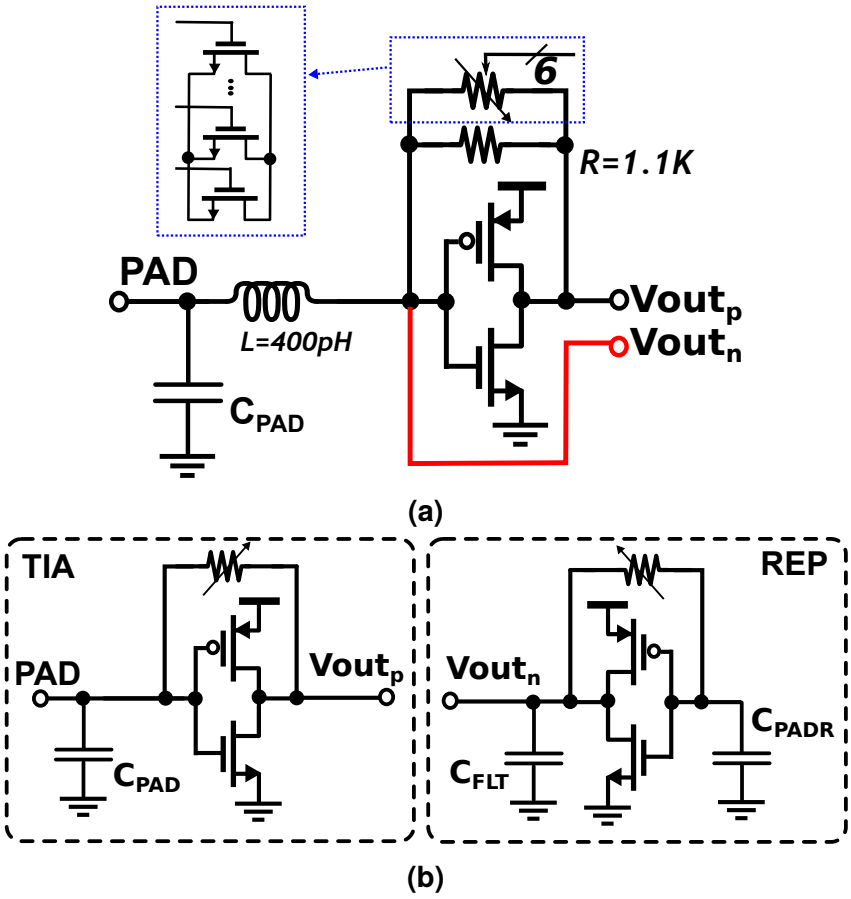

Figure 3.10: TIA schematic (a) Proposed (b) Replica.

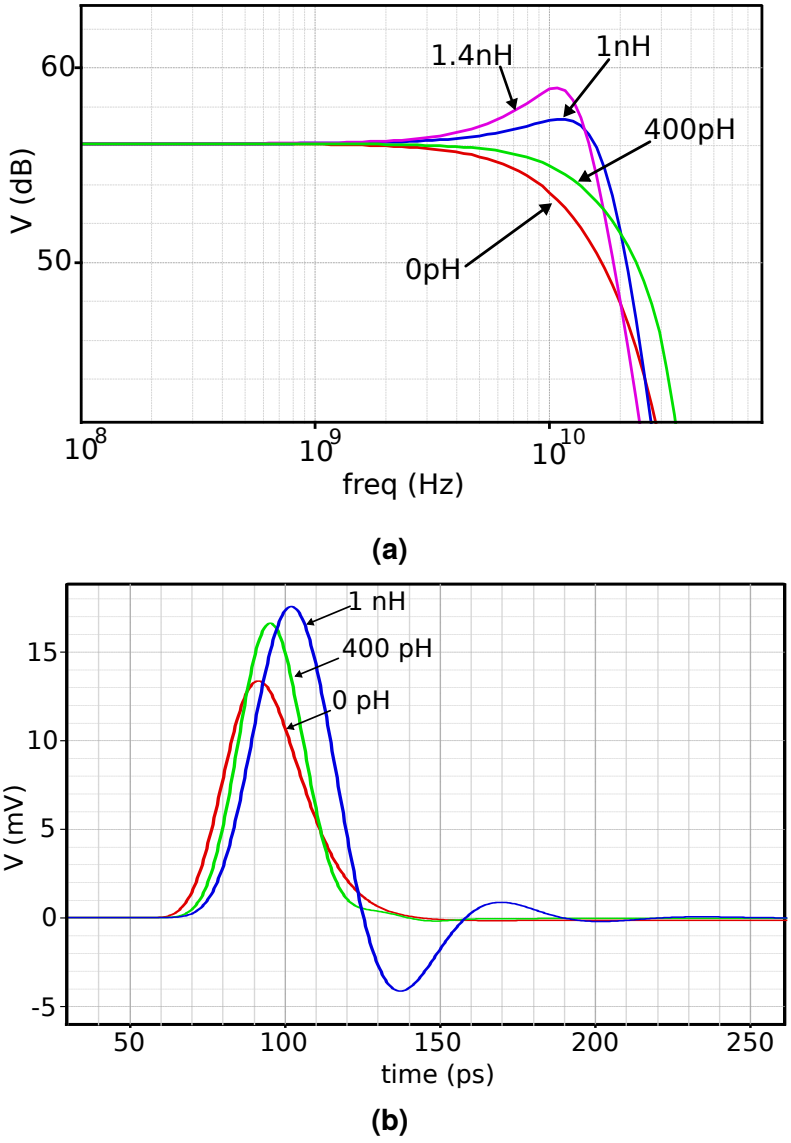

Figure 3.11: Bandwidth extension with series inductance (a) Frequency response (b) Pulse Response.

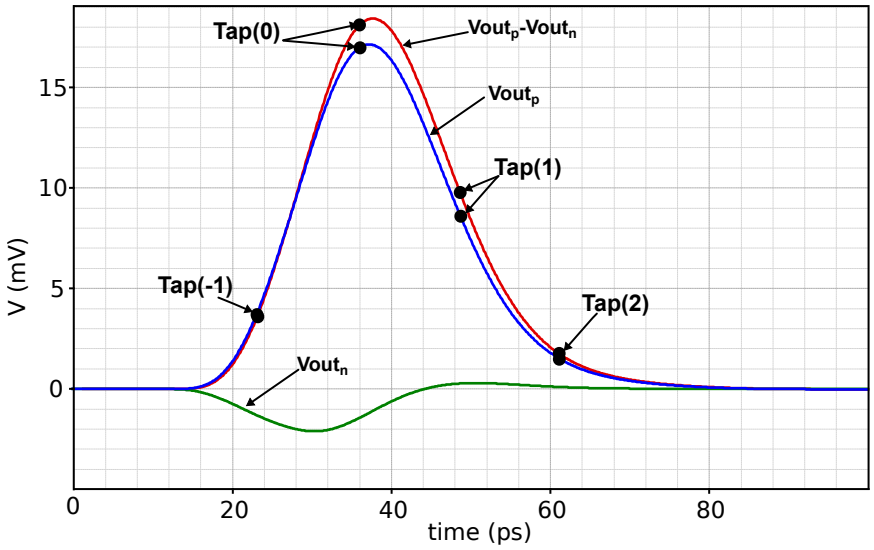

Figure 3.12: Self referenced TIA transient pulse response.

sampling points for better comparison of the two cases. The main cursor ($V_{Tap(0)}$) of $V_{outp} - V_{outn}$ is larger than the main cursor of the single ended output (V_{outp}) whereas $V_{Tap(-1)}$ and $V_{Tap(2)}$ are the same. Since $V_{Tap(1)}$ will be equalized by the DFE, the increase in this ISI term does not degrade signal integrity.

Another advantage of using the self referenced TIA is that it generates less noise compared to the replica design. In Fig. 3.13 three noise spectral densities (NSD) are given. The red solid line is the NSD of the TIA with a replica with no filtering capacitor (C_{FLT}) at its output. Adding a 600 fF of C_{FLT} shapes the NSD as indicated by the blue dotted curve. The green dashed curve is the NSD of the proposed self referenced TIA. There are two main reasons for the reduction in noise. First one is that there is no replica to generate noise. Note that the replica generates as much noise as the TIA itself increasing the integrated noise by a factor of $\sqrt{2}$. High frequency noise of the replica TIA can be filtered out by using a large capacitance at the output node. However, this would prevent the replica TIA from tracking the main TIA behavior for high frequency supply disturbances compromising power supply rejection ratio (PSRR). The second reason for noise reduction is that in self referenced TIA the low frequency noise components of the transistors are converted into common mode noise. This explains why no flicker noise is observed in the NSD of the self referenced TIA as illustrated in Fig. 3.13.

To investigate the PSRR of the self-reference TIA, it is critical to separate the input and output capacitance connected to either VDD or GND as seen in Fig. 3.14. It is easy to deduce that the currents i_i and i_o become zero if the

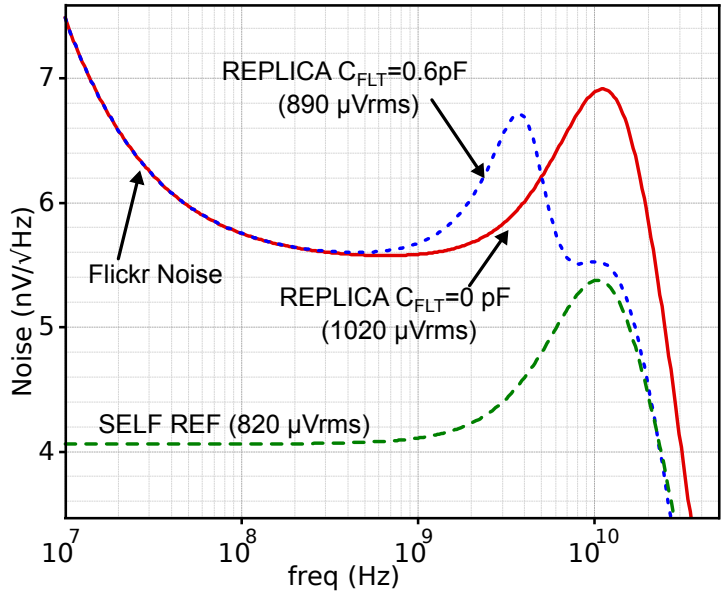

Figure 3.13: Self referenced TIA noise spectral density.

following condition is met:

$$\frac{C_{i1}}{C_{i2}} = \frac{gm_p}{gm_n} = \frac{C_{o1}}{C_{o2}} \tag{3.18}$$

In that case the current through the feedback resistor becomes zero resulting in perfect cancellation of power supply ripple. In our implementation the TIA drives a VGA whose input stage consists of a CMOS inverter with equal sized PMOS and NMOS transistors. Thus, the TIA output capacitance is divided equally between VDD and GND ($C_{o1} = C_{o2}$). Also the PMOS and NMOS transistors that compose the CMOS inverter of the TIA are sized equally which matches the two transconductances in this technology ($gm_p = gm_n$). Dividing the input capacitance equally between VDD and GND is more challenging. It consists of three parasitic capacitances. The first one is the C_{gs} of the transistors, which is already split equally between GND and VDD due to sizing of the transistors. The second parasitic capacitance at the input node is the pad capacitance. In general this capacitance couples the pad to substrate (connected to GND) creating an imbalance between C_{i1} and C_{i2}. One solution to circumvent this problem is to add a power grid below the pad in the lowest metal layer to couple the PAD equally to VDD and GND. In the used 13 level metal stack, this modification corresponds to a pad capacitance increase of approximately 5%, which has negligible impact on sensitivity. The last portion of the input capacitance comes from the PD. This capacitance is coupled to the supply voltage of the PD outside the chip and cannot be balanced as required for perfect PSRR. However, it is decoupled from the TIA input by both the bondwire and peaking inductances at high frequencies. On the

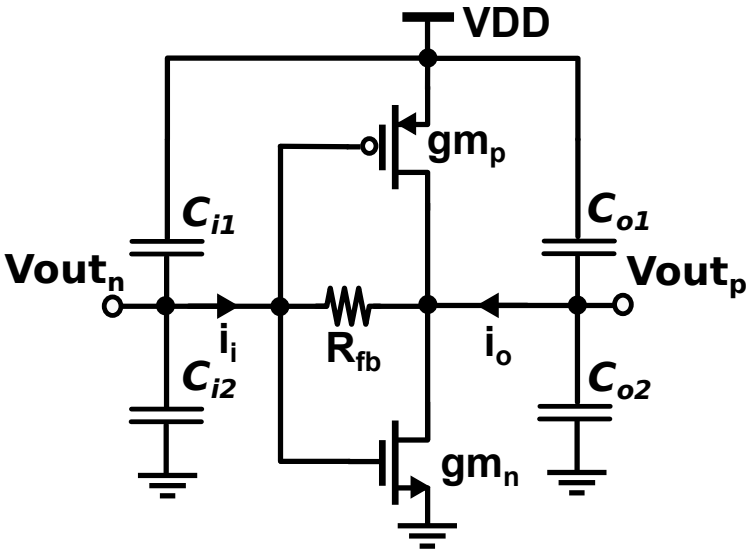

Figure 3.14: Self referenced TIA schematic for PSRR analysis.

other hand, in a replica TIA design, the PD capacitance and the bondwire inductance also creates an imbalance unless a dummy PD is placed in the packaging which may not be desirable in practical applications since it would increase both the cost of packaging and the pitch width of multi-channel design.

The PSRR simulation results of self referenced TIA and replica TIA are compared in Fig. 3.15. The worst PSRR performances of all process corners for both cases were also provided in the plot. As expected, the worst case was slow-NMOS-fast-PMOS corner (fast-NMOS-slow-PMOS performs only slightly better) since it is the corner that degenerates the gm_p/gm_n ratio the most.

To summarize, the proposed self-referenced TIA provides larger swing, lower noise and similar PSRR performance as compared to a replica TIA while consuming half the power and layout area. Moreover, the TIA has zero offset by design.

VGA

The high losses in the optical path may result in very small current signals on the PD. As an example a -12 dBm OMA signal on a 0.5 A/W responsivity PD corresponds to a 32 μA_{pp} photo-current. This signal is converted into a voltage signal with a DC gain of around 700 Ω resulting in a 22.4 mV at the output of the TIA. ISI further reduces the signal down to $10-15$ mV. Moreover, as explained in Section 3.2.1, the capacitive load at the output node of the TIA must be low which means the slicers, creating approximately 100 fF load, can't be driven by TIA directly.

In order to amplify the signal and drive the slicers, a VGA was designed and

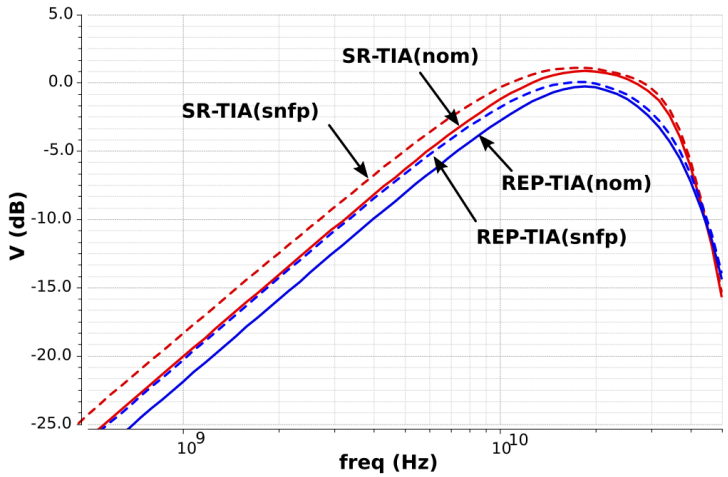

Figure 3.15: PSRR comparison of self-referenced TIA and replica TIA.

placed after the TIA. The schematic of the VGA is given in Fig. 3.16. It consists of two trans-admittance trans-impedance (TAS-TIS) stages. This structure is also known as Cherry-Hooper amplifier in literature [30]. The first TAS-TIS stage is a CMOS based design to match the common mode output of the TIA, which is around VDD/2. It must be noted that the voltage gain on the TAS is very small (around 1) due to the low input impedance of the TIS stage. This reduces the Miller effect on the C_{gd} of the input transistors minimizing the equivalent input capacitance. The DC gain of the 1^{st} stage is gm_1R_{f1} and can be controlled by changing R_{f1}.

The output common mode of the 1^{st} stage is adjusted to match the input common mode requirements of the 2^{nd} stage by injecting current into the input of the TIS stage creating a voltage drop on the feedback resistors R_{f1}. The output signal of the 1^{st} VGA stage is still pseudo differential. That is, the TIA outputs $Vout_n$ and $Vout_p$ are amplified separately (by the same gain) resulting in larger swing in Y_p. As a result, the formal definition of the output common mode $(Y_p + Y_n)/2$ is not a constant signal. That's why the output common mode is sensed from the low swing output node via a low-pass filter as illustrated in Fig. 3.16.

The 2^{nd} TAS-TIS is CML based and converts the pseudo differential signal at its input to a fully differential signal at its output. The input is connected to two differential pairs. The inner pair is sized at a quarter of the outer pair. And the current generated by the inner pair is multiplied by 4 on the NMOS mirrors to double the transconductance provided by the outer pair. Compared to a standard CML stage, the effective transconductance increases by a factor of 2, whereas the power consumption and input capacitance increases only by a factor of 1.25.

The resistors R_p extend the bandwidth of the VGA. In the TIS both NMOS

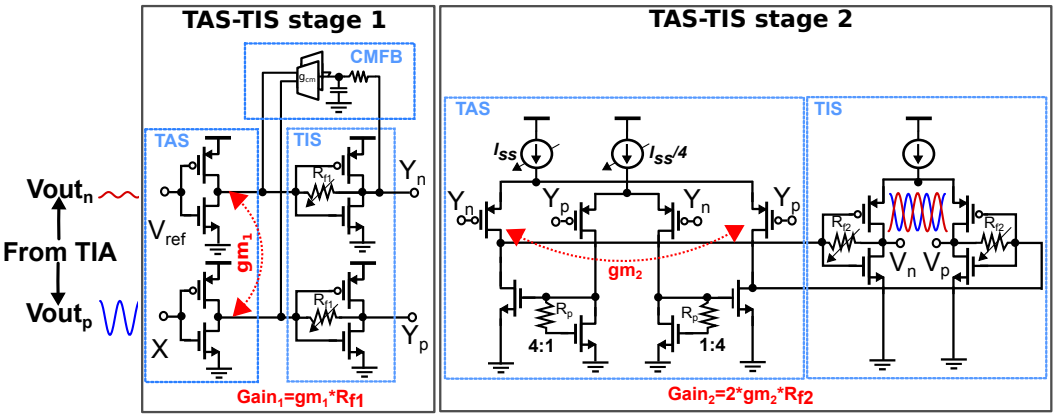

Figure 3.16: VGA schematic.

and PMOS differential pairs contribute to gain by reusing the tail current, which minimizes the power consumption. The gain of the 2^{nd} TAS-TIS can be controlled both by changing the feedback resistors R_{f2} and by changing the tail current I_{SS}.

In nominal settings the VGA provides a gain of around 20 dB with a bandwidth of 20 GHz while driving a load of 100 fF. The total power consumption of the VGA is 21 mW and its input capacitance is 25 fF. Its input referred noise is around 350 μV_{rms}. Since the TIA output integrated noise is 820 μV_{rms}, the VGA noise reduces the SNR by 10%.

3.2.2 DFE

As was explained in Section 2.2.2, in conventional speculative DFE implementations the following timing requirement should be met [31], [32]:

$$t_{c2q} + t_{mux} + t_{setup} < 1UI \tag{3.19}$$

where t_{c2q} is the clock-to-Q delay, t_{mux} is the mux delay time, and t_{setup} is the setup time of the latch. In addition to that, in a quarter rate design the t_{c2q} of the comparators sampling the speculative decisions should be smaller than 2-UI which comes as an additional timing constraint. By moving the DFE equalization into the digital domain both of those problems can be avoided at the cost of increased circuit complexity and power consumption. This is achieved via the look-ahead DFE technique ([33] [34]).

The block diagram of a 1-tap look-ahead DFE implementation for quarter rate sampling is given in Fig. 3.17. In the block diagram the subscripts denote speculations and the indexes denote the sample time. The logic works as follows: The 8 speculative decisions $D_{0,1}[0:4]$ are resolved for the two assumptions that the previous bit is equal to 0 and 1 ($S_{-1} = 0$ and $S_{-1} = 1$), and 8 look-ahead

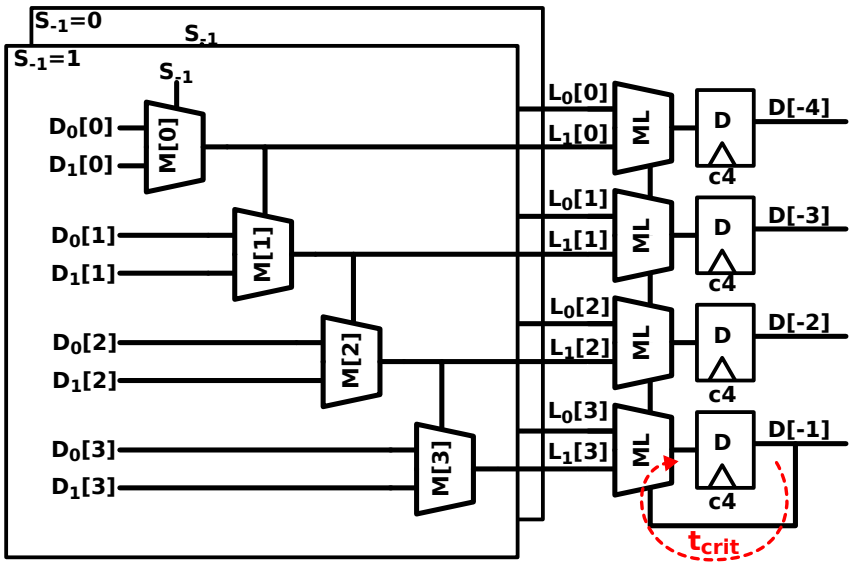

Figure 3.17: 1-tap look ahead DFE at quarter rate

decisions are generated ($L_{0,1}[0:4]$). In the final step the look-ahead decisions are resolved via the multiplexers ML with the decision $D[-1]$ (there is no subscript as this decision is not speculative anymore) used as the select signal.

Since the first part of the circuit that generates look ahead decisions ($L_{0,1}[0:4]$) is feed-forward, it can be pipelined as much as necessary relaxing constraints in this part. The only timing limitation comes from the feedback path through the multiplexers ML and flip-flops D:

$$t_{c2q,D} + t_{mux,ML} + t_{setup,D} < 4UI \tag{3.20}$$

where $t_{mux,ML}$ is the select to output delay of the multiplexer ML.

The extension of the look-ahead DFE technique into 2-taps is given in Fig. 3.18. In this case the number of initial assumptions increase to 4 (combinations of $S_{-1}S_{-2}$), just like the number of look-ahead decisions per bit ($L_{00,01,10,11}[n]$). Eventually increasing the number of blocks in the DFE. The total number of 2-to-1 multiplexers (assuming a 2^n-to-1 mux has $2^n - 1$ number of 2-to-1 muxes) in an n-tap look-ahead DFE for quarter rate is given by:

$$N_{mux21}(n) = 4(2^n - 1)(2^n + 1) \tag{3.21}$$

This results in 60 and 224(!) muxes for 2-tap and 3-tap look-ahead DFE logic, respectively. Also, since the feed-forward delay through the muxes $M[0:3]$ increases significantly, more pipeline stages needs to be placed on this path, leading to even more complicated circuitry and more power consumption. Furthermore, although the critical path for timing does not change, ML becomes a 4-to-1 mux and an

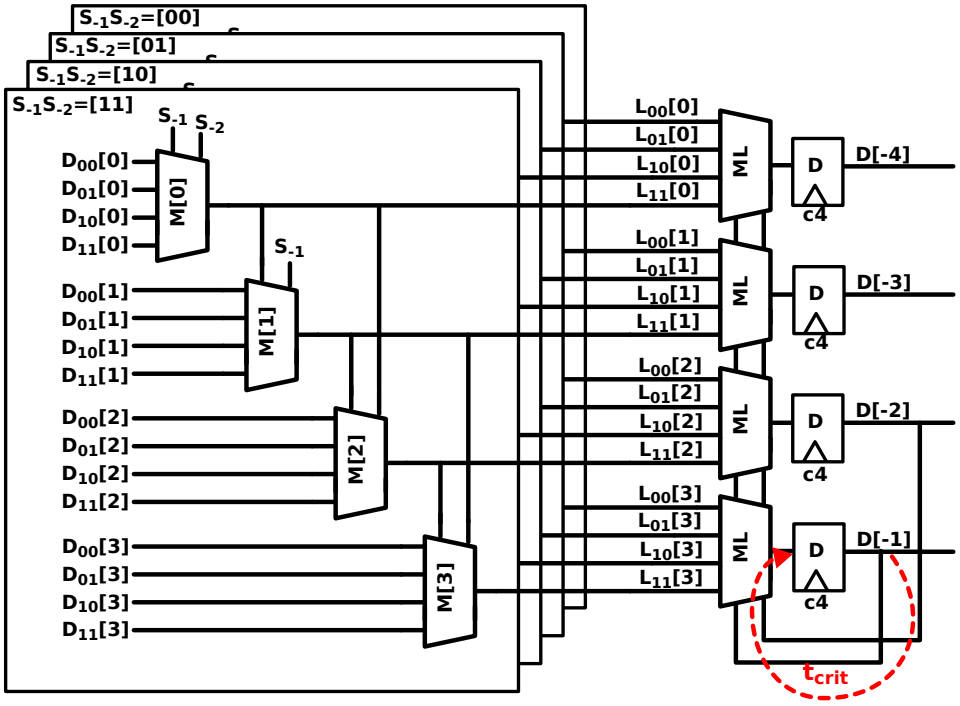

Figure 3.18: 2-tap look ahead DFE at quarter rate

8-to-1 in 2-tap and 3-tap DFEs resulting in longer $t_{mux,ML}$ delay times.

To conclude, look-ahead DFE architecture relaxes the critical timing significantly at the cost of increased circuit complexity and power consumption. This cost increases exponentially with each additional DFE tap. In Section 3.2.1 (Fig. 3.9), the SNR vs number of DFE taps analysis shows that most of the SNR gain is achieved via the 1^{st} DFE tap and the additional DFE taps contribute only marginally. As a result, a 1-tap look-ahead DFE architecture was chosen as an optimal trade-off between power consumption and circuit complexity and SNR improvement.

In Fig. 3.19 the block diagram of the implemented 1-tap DFE is given. The differential input ($V_{p,n}$) is sampled by the comparators which are driven by quarter rate clocks. After that, the signals are aligned to a single quarter rate clock and the look-ahead DFE resolves the speculation. Then, look-ahead signals $L_H(n)$ and $L_L(n)$ are calculated from the speculative decisions $D_H(n)$ and $D_L(n)$. Finally, $D(3)$ resolves the speculation. The dependence of each bit to the previous bit is broken in the new speculative array which results in a relaxed timing constraint of:

All the digital logic up to $L_H(n)$ and $L_L(n)$ is feed-forward and can be pipelined to meet the timing. In this application a two-stage pipeline was required to close timing. The clock is also feed-forwarded to enable deeper logic between the two flip-flops. In RC extracted simulations the look-ahead DFF logic was functional

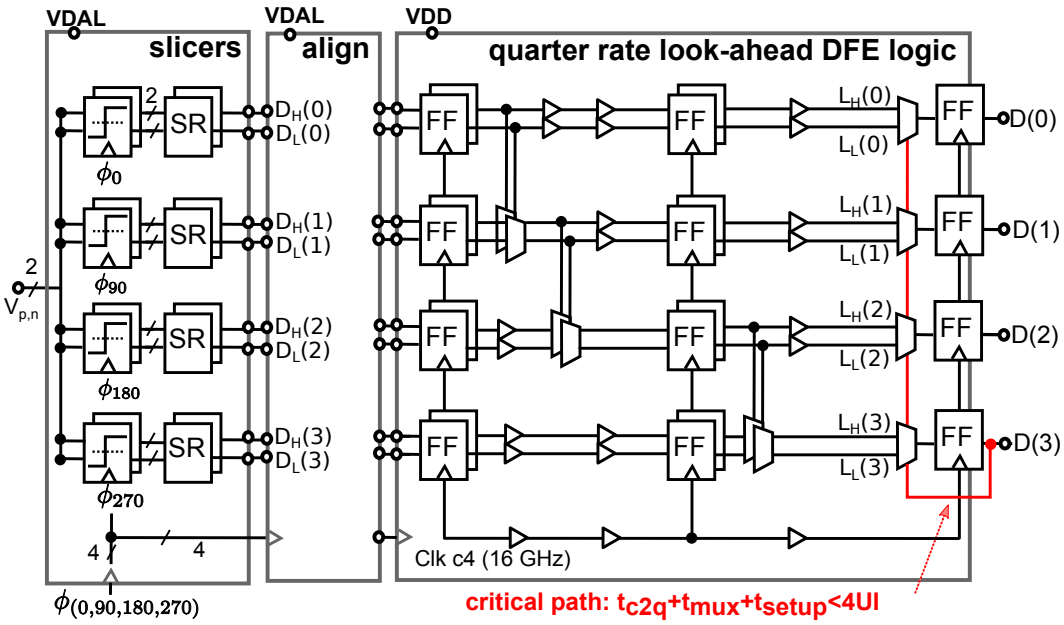

Figure 3.19: Slicers and look-ahead DFE block diagram.

up to 85 Gb/s at 800mV supply. Therefore DFE is not limiting the data-rate of the optical RX.

The schematic of the comparator is given in Fig. 3.20. The first stage consists of two differential pairs whose sources are connected directly to VDD. The clock transistors are connected as cascode as in a Lewis-Grey comparator [35]. Compared to a conventional dynamic comparator where the clock transistors are located on the tail of the input pair, this topology enables to increase the common mode for the same CK-Q delay. The nominal common mode of this latch is 350 mV, which matches well with the output common mode of the AFE. The second stage is a self timed latch with cross coupled inverters. The threshold of each comparator is controlled via a 9-bit voltage DAC (VDAC) whose control bits are stored in on chip registers. The VDAC covers a range of ±VDD/2 with a resolution of VDD/512. In order to minimize the layout area, the resistor string network of the VDAC is realized using the parasitic resistances of low level metals [32]. The integral non-linearity (INL) of a test VDAC was measured to be below ±1LSB (±1.75mV).

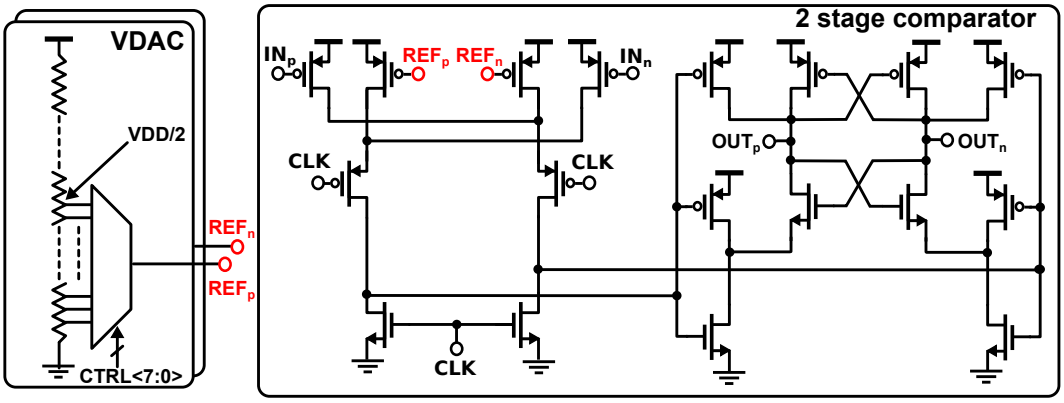

Figure 3.20: Comparator schematic.

3.3 Clock and Data Recovery [1]

This section explains the reasons to use the selected clock and data recovery (CDR) architecture as well as describing the circuit techniques used to implement the CDR.

In general IO links can be classified under two categories: source-synchronous and asynchronous. Source-synchronous links require the transmission of the reference clock as well as the data within the bus, increasing the number of lanes to be used for the same total bandwidth. The additional lane increases both power consumption and system cost. Moreover, at high-speed links the sampling phase needs to be adjusted by a phase tracking loop on the receiver. Furthermore, depending on the channel characteristics the transmitted clock may be significantly degraded in terms of signal integrity.

In contrast, asynchronous links do not require the transmission of the clock making them cost and power efficient. However, a sampling clock needs to be generated by a clock and data recovery circuit on the receiving side.

The asynchronous architectures can be categorized under two main branches according to the clock generation principle: with and without reference clock [36]. Referenceless CDR architectures incorporate a voltage controlled oscillator (VCO) and analog loop filter to generate the sampling clock as illustrated in Fig. 3.21a. The CDR architectures with reference clock replace the VCO with a digital-to-phase converter (alternatively referred as phase rotator) driven by a phase accumulator and analog loop filter with a digital one. This digitization has several advantages [37]:

[1] This section is based on: I. Ozkaya, A. Cevrero, P. A. Francese, C. Menolfi, T. Morf, M. Brändli, D. M. Kuchta, L. Kull, C. W. Baks, J. E. Proesel, M. Kossel, D. Luu, B. G. Lee, F. E. Doany, M. Meghelli, Y. Leblebici, and T. Toifl. "A 60 Gb/s 1.9 pJ/bit NRZ Optical Receiver With Low-Latency Digital CDR in 14 nm CMOS FinFET". IEEE Journal of Solid-State Circuits PP.99 (2018), pp. 1-11.

- Referenceless clock CDR may suffer significantly from thermal and supply ripple induced noise that is injected into the VCO control voltage. Digital solution is affected much less by the noise injection.
- In VCO based CDR applications, the sensitivity to noise may enforce a much lower CDR loop bandwidth compared to PR based CDR, resulting in longer settling times, which is an important parameter in burst mode applications such as this work
- Analog loop filter and VCO may be affected by the process, voltage and temperature (PVT) variation whereas digital alternative is insensitive to PVT variation.
- Digital circuits in the CDR with reference clock are easier to port across different technologies.
- Digital solution allows design parameters to be controlled easily.

The disadvantage is that an external reference clock needs to be generated in the system. On the other hand, in large systems applications such as processors, a reference clock is required for the rest of the system anyway. Thus, the reference clock is likely to exist already, to be reused by the receiver bus.

The phase detection schemes can also be divided into two main groups: linear (e.g. [38]) and non-linear (e.g [39]) phase detection schemes. Schematics of implementation of both schemes are given in Fig. 3.22.

In the presence of data transitions, the Hogge PD in Fig. 3.22a generates pulses whose width are proportional to the phase error between the data and clock, and compares them to fixed-width reference pulses. The average difference between the proportional and reference pulses indicates the magnitude and polarity of the phase error [40]. In high data-rate applications, the XOR gates which generate the pulses, may limit the phase detection performance as the rise and fall times become comparable to 1 UI. Moreover, in the existence of large ISI components the pulse width depends on the data pattern and it is impractical to pattern filter the analog pulses. As a result, it is not feasible to use a linear phase detection technique considering the data-rate target of this work and the bandwidth limitation introduced to improve sensitivity that generates a large post-cursor ISI (as explained in Section 3.2.1).

In the non-linear phase detection scheme, the data and edge samples are compared to define the polarity of the phase error. As the detection is done on the sampled data, non-linear phase detection is more robust in the sense that it does not rely on the analog performance of the blocks. Also digital post-processing, such as pattern filtering, is an option if required by the input signal characteristics.

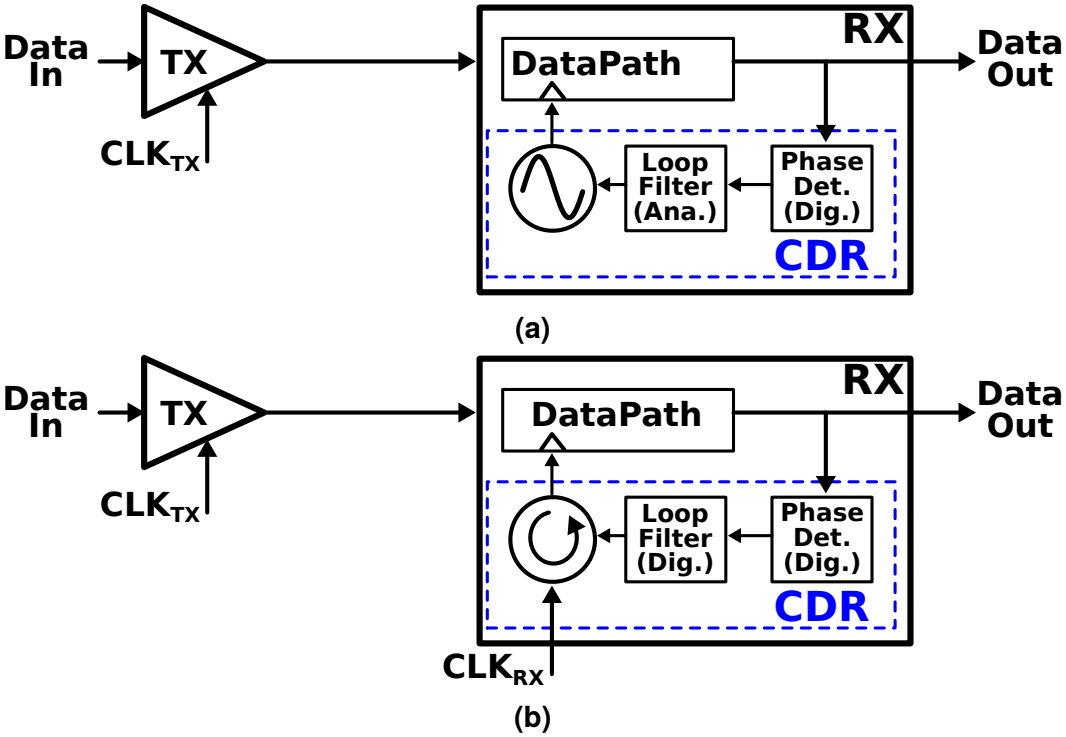

Figure 3.21: CDR clock generation architectures (a) Without reference (b) With reference

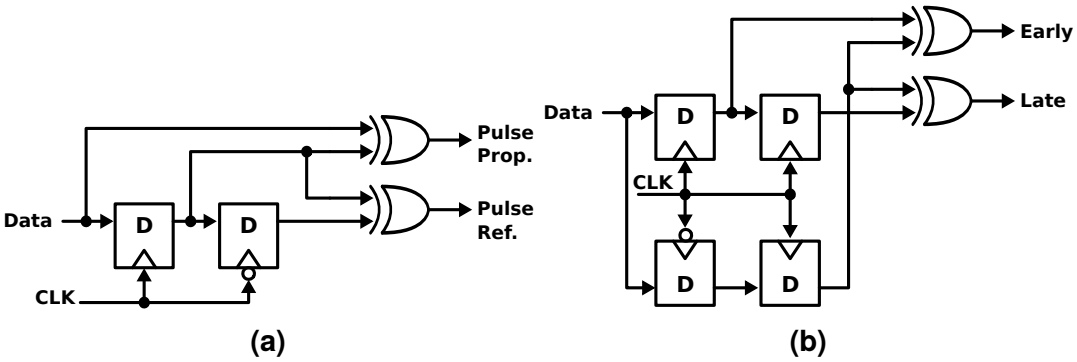

Figure 3.22: Phase detection schemes (a) linear (b) non-linear

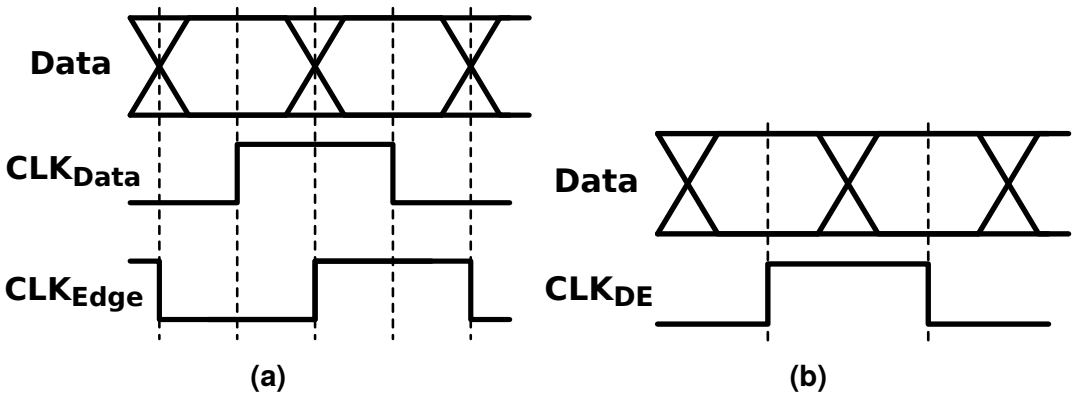

Figure 3.23: Non-linear phase detection techniques (a) edge-sampled (b) baud-rate

Non-linear phase detection scheme can be divided into 2 main sub-classes: edge-sampled and baud-rate. Edge-sampling scheme requires clock phases for sampling the edges of the data, as illustrated in Fig. 3.23a. Baud-rate eliminates the need for those extra phases Fig. 3.23b making the clock generation easier. However, most of the baud-rate implementations rely on certain ISI conditions to be met by the analog data. For example, one of the conventional baud-rate phase detection technique introduced by and named after Mueller-Muller [41] requires a symmetric pulse as the algorithm drives the sampling point to the center of geometry of the pulse. Especially in optical links where the signal amplitude depends on many parameters such as coupling losses, mechanical stress on the fiber, and even ambient temperature; it is very difficult to maintain the same signal characteristics over a long period of time and across a wide data-rate range as targeted in this work. Therefore, edge-sampled non-linear phase detection scheme has been selected for CDR implementation in this study.

In this work a phase interpolation (PI) CDR is designed and implemented with a half rate reference clock signal, which can be provided by a central PLL to drive the complete I/O bus in the server.

In the remainder of this section the critical parameters affecting the jitter tolerance performance will be determined based on the provided CDR model and the circuit details of the blocks used to realize the CDR will be explained in detail.

3.3.1 CDR Modeling

In this section, an insight into the loop dynamics of the CDR will be provided to motivate the design choices explained in the following sections. The linearized model to be used for the analysis is given in Fig. 3.24.

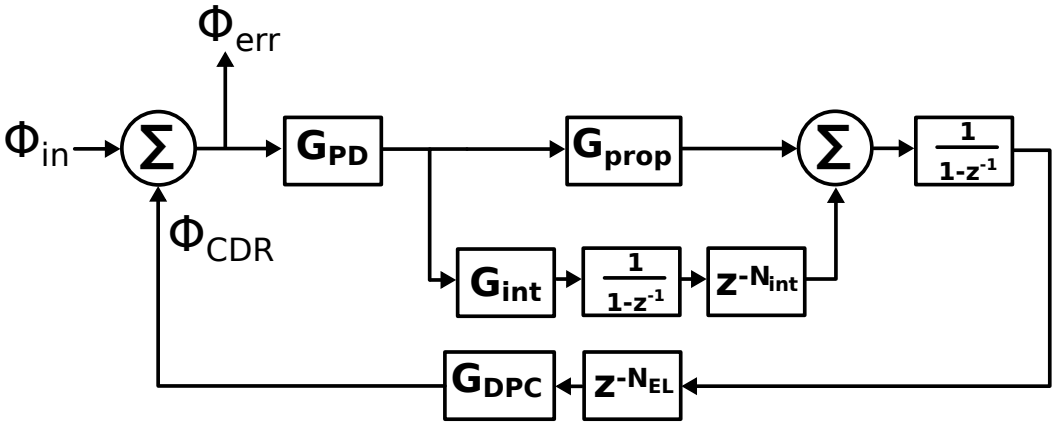

Figure 3.24: Linearized CDR model

The open loop transfer function of this model is given by the Eq. (3.22):

$$OLTF = G_{PD}(G_{prop} + G_{int}\frac{z^{-N_{int}}}{1-z^{-1}})\frac{z^{-N_{EL}}}{1-z^{-1}}G_{DPC} \tag{3.22}$$

where G_{prop} is the proportional path gain, G_{int} is the integral path gain, N_{int} is the extra delay on the integral path in terms of unit interval (UI), N_{EL} is the complete loop delay in terms of UI, G_{DPC} is the digital to phase converter gain (which is a phase rotator (PR) in this implementation), and G_{PD} is the phase detector gain which includes a bang-bang phase detector (PD) followed by a tree-style 4-to-1 majority voting in our implementation. Also ϕ_{in} is the input phase, ϕ_{CDR} is the sampling phase, and ϕ_{err} is the phase error. ϕ_{err} defines the performance of the CDR loop as the jitter tolerance (JTOL) is directly proportional to the reciprocal of the ϕ_{err}.

Despite the fact that a bang-bang phase detection and majority voting both have nonlinear responses, in the existence of noise (and jitter) a linear model can be extracted for those operations as shown in [37]. Although the linearized CDR model is based on the one provided in [37], the focus in this study will be on the impact of the loop latency on the CDR dynamics. Also in [42] a similar analysis is provided with results supporting the analysis provided in this work.

The analysis will begin with a simplified CDR model which does not include the integral path ($G_{int} = 0$). In that case the open loop transfer function becomes:

$$OLTF = G_{PD}G_{prop}\frac{z^{-N_{EL}}}{1-z^{-1}}G_{DPC} \tag{3.23}$$

The bode plot that corresponds to this model with various loop delay values (N_{EL}) is given in Fig. 3.25. The loop latency values are chosen close to the simulated latency of around 60-70 UI. There are 3 main parameters that determine the open

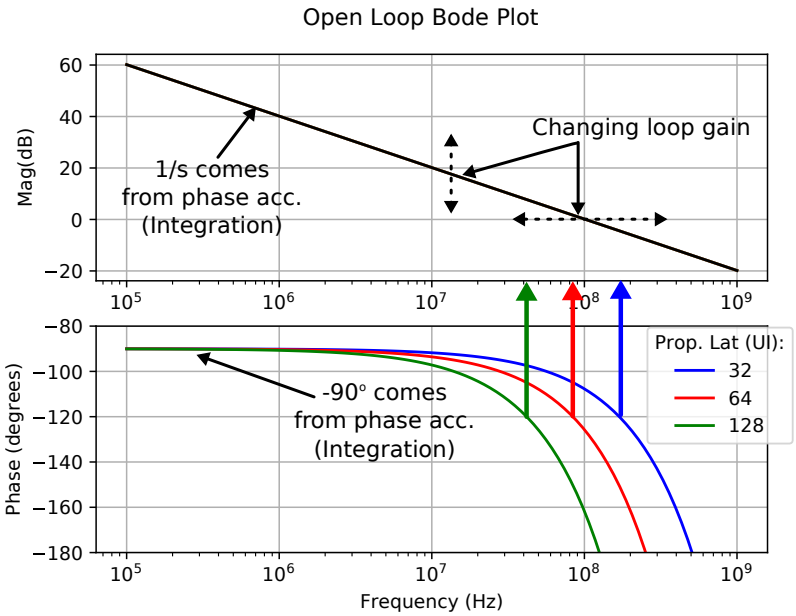

Figure 3.25: CDR model open loop transfer function for different N_{EL} ($G_{int} = 0$)

loop transfer function: 1) integration which shapes the amplitude response to $1/s$ and adds a $-90°$ phase shift, 2) the loop gain which moves the amplitude vertically, and 3) loop delay (N_{EL}) which corresponds to a negative phase change proportional to the frequency resulting in the bending of the phase response downward. The bigger the N_{EL}, the earlier the bent occurs. As a result as N_{EL} increases the phase margin goes down resulting in peaking in the closed loop transfer function. The impact of this peaking on the JTOL is illustrated in Fig. 3.26.

In order to have an optimum settling behavior that does not create peaking in JTOL, $60°$ phase margin must be satisfied by changing the loop gain. The change in JTOL with respect to loop gain (at a fixed loop latency of 128 UI) is given in Fig. 3.27. As the loop gain is reduced the phase margin increases and the peaking disappears whereas the JTOL corner frequency also drops. The loop gain is controlled by the loop filter described in Section 3.3.2

Now that the relation between loop latency and JTOL function is explained, the integral path can be introduced to study its effects. The bode plot for the open loop transfer function of the CDR for various integral path gain (G_{int}) with other parameters constant is given in Fig. 3.28. Also the integral path delay is assumed to be 0. The introduction of an integral path increases the slope of the amplitude in the lower frequencies to 40 dB/decade and shifts the phase by another $-90°$ at 0 input frequency. As G_{int} is increased, the gain at lower frequencies increases whereas the phase margin decreases. The JTOL functions corresponding to the same G_{int} values are given in Fig. 3.29. The introduction of an integral gain

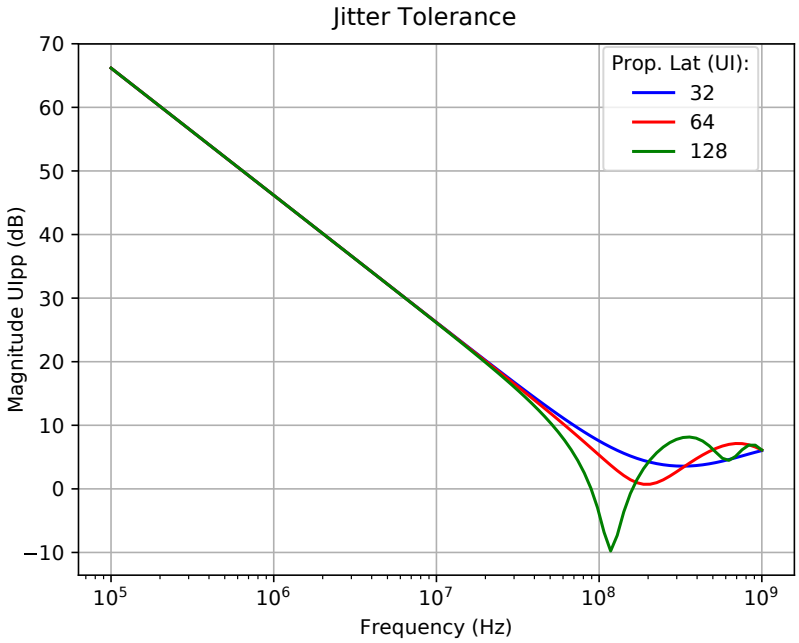

Figure 3.26: JTOL for different N_{EL} ($G_{int} = 0$)

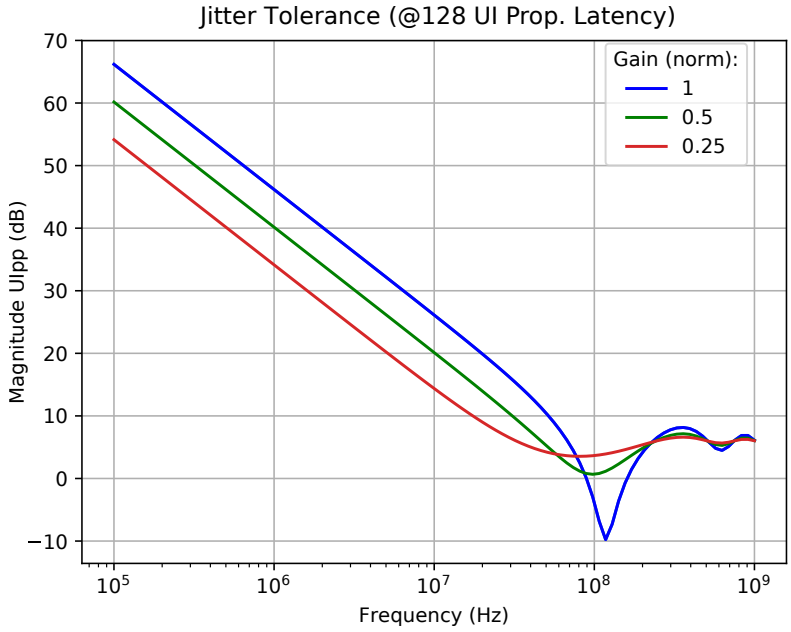

Figure 3.27: JTOL for different loop gain ($N_{EL} = 128UI$ and $G_{int} = 0$)

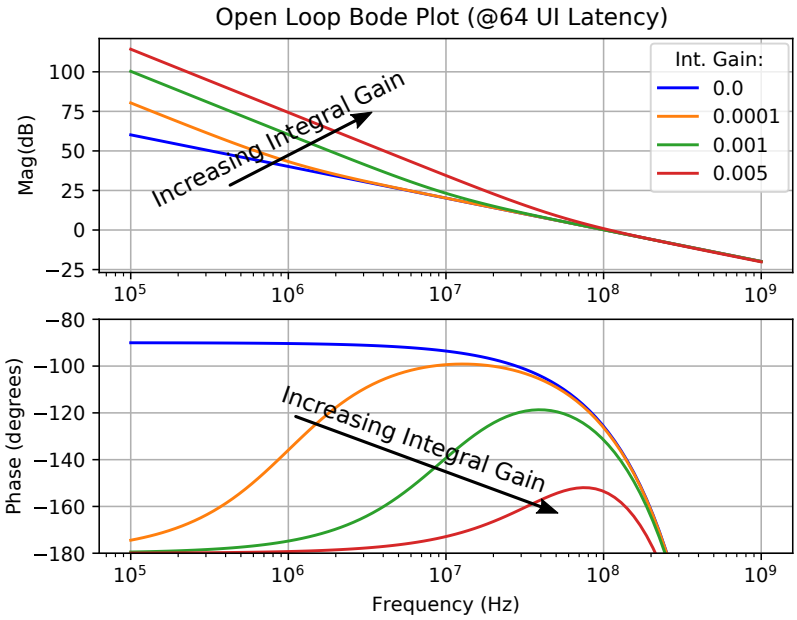

Figure 3.28: Open loop transfer function for different G_{int} ($N_{EL} = 64UI$)

improves the JTOL response in the lower frequencies significantly. On the other hand a large G_{int} starts to create peaking (as expected from open loop transfer function) compromising the JTOL performance.

In the final part of the analysis N_{int} will be increased while keeping the other parameters constant at their nominal values. The effect of this additional integral path delay on JTOL is given in Fig. 3.30. The graph indicates only a small change in the jitter tolerance plot with increasing N_{int}. Even a delay of 256 UI does not compromise the JTOL function significantly.

The conclusion from this study is that the proportional path delay directly determines the JTOL performance of a CDR loop whereas the integral path delay has almost no impact as long as the G_{int} coefficient is kept small enough not to compromise the phase margin. Thus, when designing a CDR circuit much of the effort must be spent on minimizing the proportional gain while digital synthesis can be used for the design of integral path as it is quicker than custom design and allows control of parameters easier.

This implementation, whose details are explained below, includes a full custom CDR circuit that runs at a quarter rate clock of 15 GHz at the maximum data rate of 60 Gb/s to minimize the proportional path delay. The integral path was omitted from this version. However, the design of the integral path is relatively easy as discussed above. The simulated proportional path latency of the CDR (including the analog delay of the clock path) is between 60-70 UI which according the analysis provided above results in a corner frequency of around 80

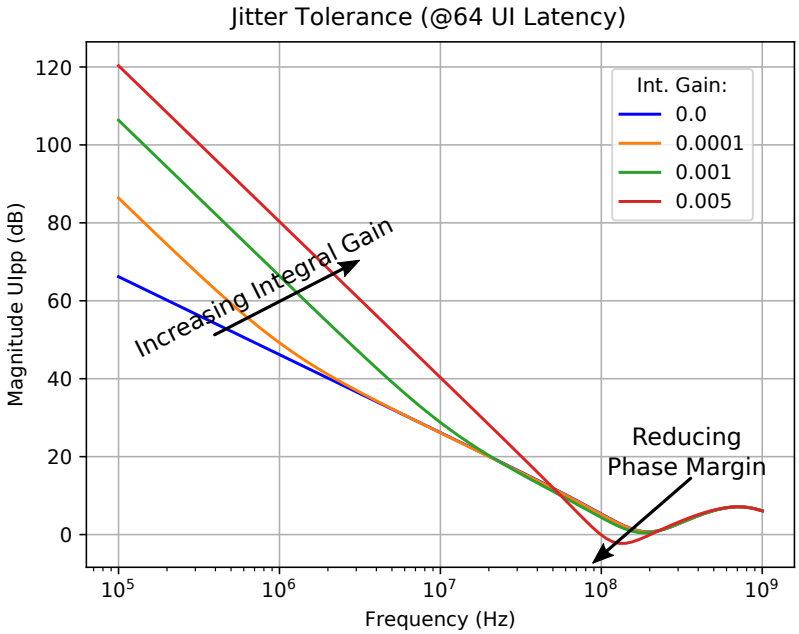

Figure 3.29: JTOL for different integral gain ($N_{EL} = 64UI$)

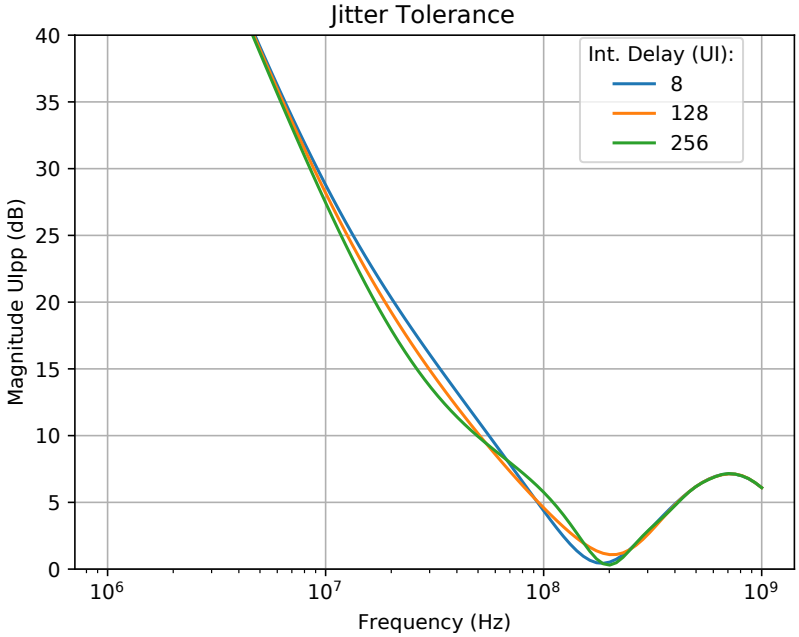

Figure 3.30: JTOL function for different N_{int} ($N_{EL} = 64UI$)

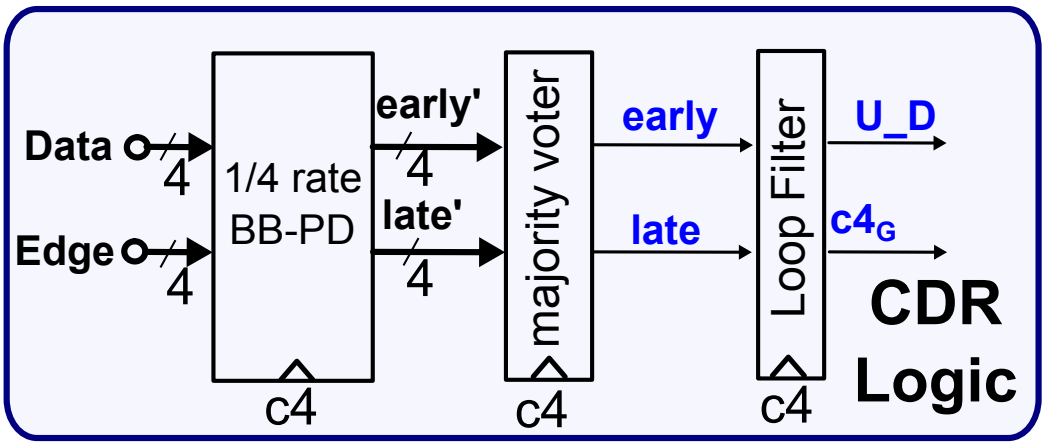

Figure 3.31: Block diagram of CDR logic

MHz at 60 Gb/s data rate that is supported by the measurement results as will be shown in Section 5.2.

3.3.2 CDR Logic

In order to minimize the latency of the CDR loop, the CDR logic is implemented with a quarter rate clock. All of the blocks in CDR path were custom designed to satisfy timing requirements of this high update rate. The CDR logic consists of 3 stages: bang-bang phase detector, majority voter, and loop filter. Its block diagram is given in Fig. 3.31.

Bang-bang Phase Detector

Phase detectors are key elements in the CDRs. As the name suggests they are used to detect the phase offset of the sampling clock with respect to the incoming data. The bang-bang phase detector architecture was proposed by J. D. H. Alexander in his fundamental article [39]. Although the bang-bang phase detector has a non-linear response, at high speed its simplicity and accuracy makes it advantageous over the linear phase detectors ([37]).

An idealized eye diagram for bang-bang phase detection is given in Fig. 3.32a with sampling phases of $d[n]$ for data and $e[n]$ for edges, producing digital outputs of $D[n]$ and $E[n]$ (Fig. 3.32b), respectively.

The bang-bang phase detector logic uses 2 consecutive data bits $D[n]$ and $D[n+1]$ to detect the existence of a transition and the edge signal $E[n]$ to decide whether the phase is early or late. It has 2 output signals: $Eearly[n]$ and $Late[n]$. When there is no transition both outputs are at logic 0. The corresponding truth table for the phase detection is given in Table 3.1. The “X”s in the table represent

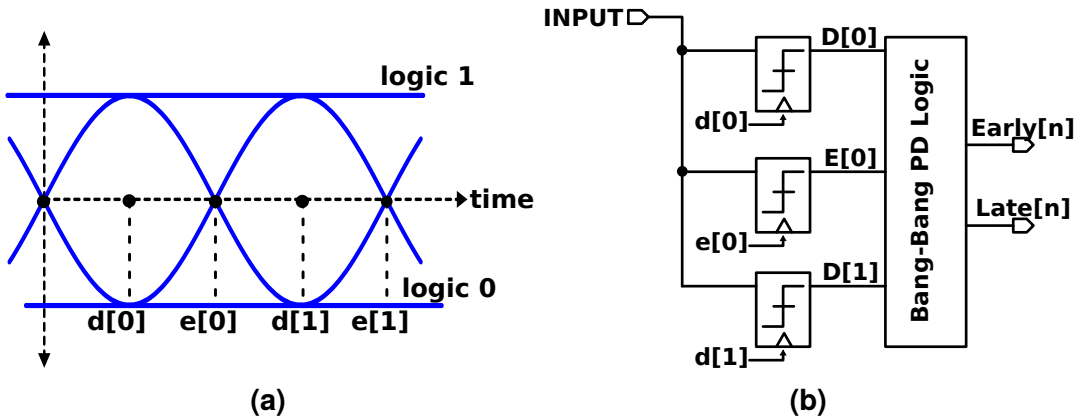

Figure 3.32: (a) Input eye diagram and sampling instances for bang-bang phase detection (b) Bang-bang phase detection circuit.

Table 3.1: Bang-bang phase detection truth table

$D[n]$	$E[n]$	$D[n+1]$	$Early[n]$	$Late[n]$
0	0	0	0	0
0	1	0	X	X
1	0	1	X	X
1	1	1	0	0
0	0	1	1	0
0	1	1	0	1
1	0	0	0	1
1	1	0	1	0

the "don't care" condition as it is impossible to get $E[n]=0$ while $D[n]=1, D[n+1]=1$ and $E[n]=1$ while $D[n]=0, D[n+1]=0$. The logic implementation of the truth table is:

$$\begin{aligned} Early[n] &= D[n+1] \oplus E[n] \\ Late[n] &= D[n] \oplus E[n] \end{aligned} \tag{3.24}$$

Majority Voter

Majority voter receives 4 early and 4 late signals from the 4 phase detectors and decides whether there are more early or late signals, and returns 1 bit early and 1 bit late signal. The straight forward method to implement the required logic would be to sum the 4 early and 4 late signals from the phase detectors and compare the two sums to make a final decision. However, this method would require complicated circuitry with a relatively deep logic that has to be pipelined,

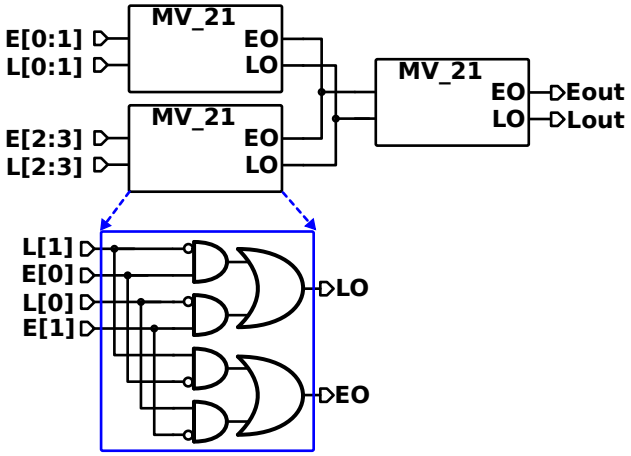

Figure 3.33: Block diagram and logic implementation of majority voter

resulting in increased CDR loop latency. In order to minimize latency, a divide an conquer approach was used in this design that simplifies the logic significantly, just like the one proposed in [42]. In this approach the 4 early and late signals coming from the phase detector are divided into two to be processed by 2 2-to-1 majority voters (MV_21). And the outputs of those majority voters are processed by another 2-to-1 majority voter to generate the final decision. The block diagram and logic implementation of this function is given in Fig. 3.33.

It must be noted that the majority voter 'loses' some valid information for certain input conditions. Such an example can be when the 4 inputs are 'EEL0': in that case the first MV_21 with 'EE' input will generate an early output whereas the second one with 'L0' at its input will generate a late output. Thus, the final MV_21 will have an 'EL' signal at its input resulting in undecided output of '0'. Obviously, this means a loss of information as the number of the early inputs were more than late inputs. This loss of data corresponds to a slight decrease in the phase detection gain around the optimum sampling point. As the sampling point moves away from the edge, the probability of receiving that sort of information from the RX drops and the phase detection gain increases again.

Loop Filter

Loop filter is used to set the proportional path gain (G_{prop}). In this study, the loop filter functionality is achieved by accumulating early and late information provided by the majority voter on a counter. When there is an "early" signal at its input, the counter incremented by 1 and when there is a "late" signal, the counter is decremented by 1. When the counter reaches a certain threshold in any direction (positive or negative), it updates the PR control by 1-step in the

respective direction and resets itself.

The G_{prop} is inversely proportional to the threshold of the counter. As the threshold is increased, the value of G_{prop} drops and as the threshold is decreased, the value of G_{prop} increases. In this implementation, the threshold of the counter was set to a minimum value that satisfies the required CDR loop phase margin at maximum sampling speed based on the transient simulation results.

The counter consists of a bidirectional shift register (BSR) and a finite state machine (FSM) checks for the trigger condition. The block diagram of the loop filter is given in Fig. 3.34 together with its state diagram. The simple implementation of the bidirectional shift register allows it to be clocked at the quarter rate clock (clk4) minimizing CDR latency.

The circuit that realizes the loop filter (Fig. 3.34) works as follows: Initially the BSR is reset to '0000000'. Then, depending on whether an early or late signal is received, it starts to fill '1's from one end, which corresponds to a clockwise or counter clockwise step in the state diagram. At the end, when the BSR is full of '1's, the FSM resets the BSR to initial position and triggers an up-or down rotation in PR_C block. It must be noted that the state can change in any of the two directions at any point in state diagram. For example, after reset if the loop filter receives an early and a late signal consecutively it will first go to state '10000000' then back to the '00000000' state.

The maximum output rate of the loop filter in this implementation is 1-step per 10 clk4 cycles limiting the frequency offset tracking range of the RX to ±780 ppm with a PR of 32-step per UI resolution. In practice this becomes > ±500 ppm error free tracking range as will be shown in Section 5.2. For frequency offsets between ±780 ppm and ±500 ppm, the RX clock is still locked to the input clock but due to the lack of an integral path the phase error becomes large and the received data is not error free any more. The frequency tracking limitation can be improved by the integration of an integral path to the CDR loop.

3.3.3 Phase Rotator

Phase rotator is most commonly implemented as a four quadrant mixer that interpolates between 4 quadrature clock phases: IP, IN, QP, and QN (Fig. 3.35a). The ideal constellation of the phase rotator would be a circle with equal phase steps and no amplitude variation as shown in Fig. 3.35b. In [43], the authors attempt to realize such a circular constellation by adjusting the I and Q weights with 5-bit controls, each. Despite the rather complicated circuitry, the measurement results are far from ideal.

There are other approaches where the ideal circular constellation is approximated with polygon constellations. The polygons can be implemented by adding

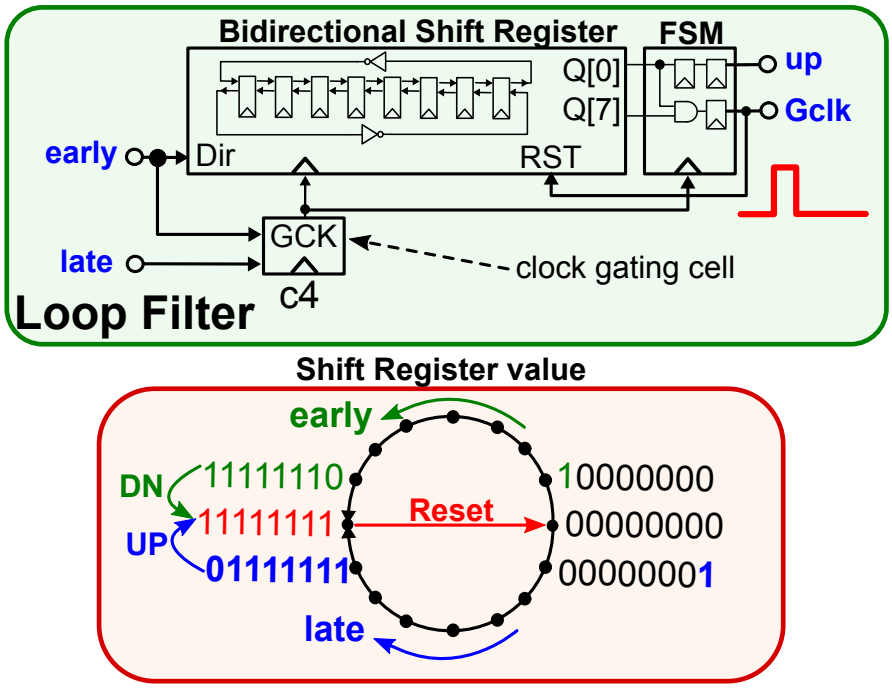

Figure 3.34: Block diagram and logic implementation of loop filter

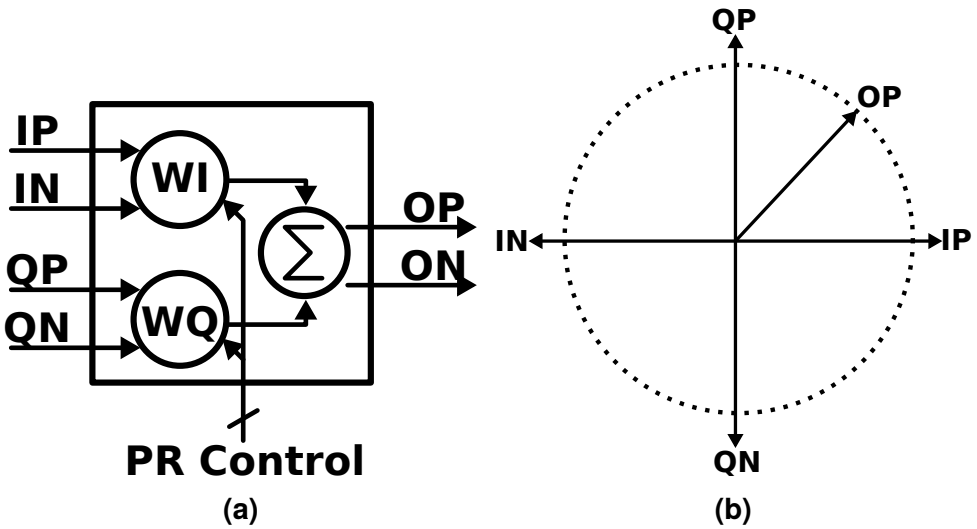

Figure 3.35: (a) Phase rotator (b) Ideal constellation.

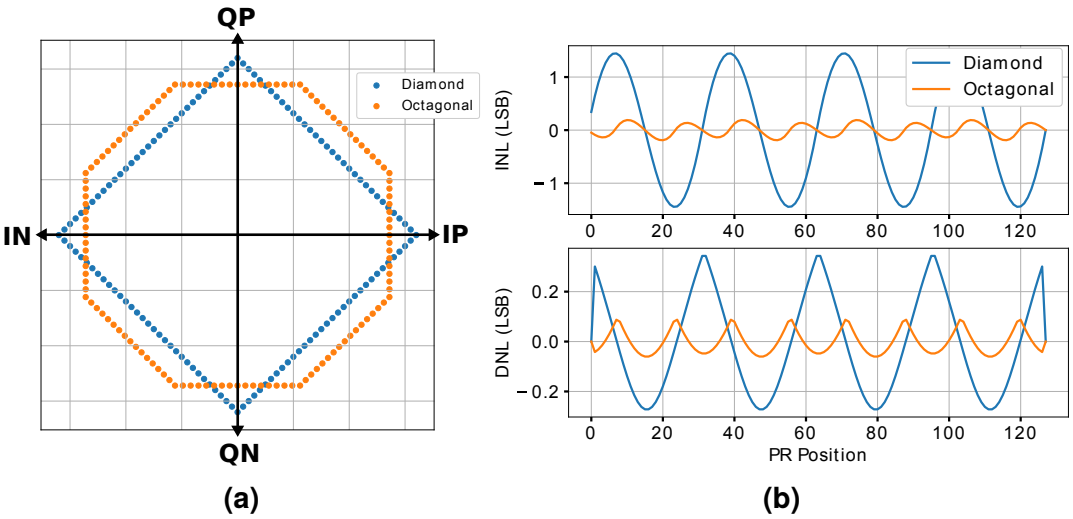

Figure 3.36: (a) Diamond and octagonal constellations (b) INL and DNL for diamond and octagonal constellations.

or subtracting unit weights of input quadrature phases which allows simpler and scalable circuit design. In [44] a diamond shaped constellation implementation is targeted. However, in order to avoid the high integral non-linearity (INL) phase error intrinsic to diamond constellation (Fig. 3.36), the step sizes along the constellation were not kept uniform resulting in a non-uniform layout. A better approximation with an octagonal constellation is proposed in [45], and later an improved version was demonstrated in [46]. This architecture reduces the INL phase error well below the quantization level for uniform weighted control. The INL and DNL comparison for 32 steps in a quadrant ($1LSB = 2.8125°$) for the two constellations is given in Fig. 3.36b. Another advantage of octagonal constellation over diamond constellation is that it has much less amplitude variation, which may cause additional phase error to be generated in the block that is driven by the PR.

The proposed phase rotator architecture is based on the one presented in [46] whose schematic, control logic, and constellation diagram are given in Fig. 3.37. The PR in [46] consists of unit CML stages with piecewise thermometric control bits to generate interpolation between I and Q inputs. The vertical and horizontal segments of the constellation diagram are controlled by IQ1 DAC and diagonal segments are controlled by IQ2 DAC. The binary phase accumulator block processes the early and late information coming from the phase detection scheme. Then, a binary-to-grey converter is used to avoid glitches in the phase output by assuring only 1-bit transition for every update. However, the glitch that comes from the charge injection of the polarity switches can not be avoided, and is a known problem. The proposed modifications on the PR allow a glitch-free

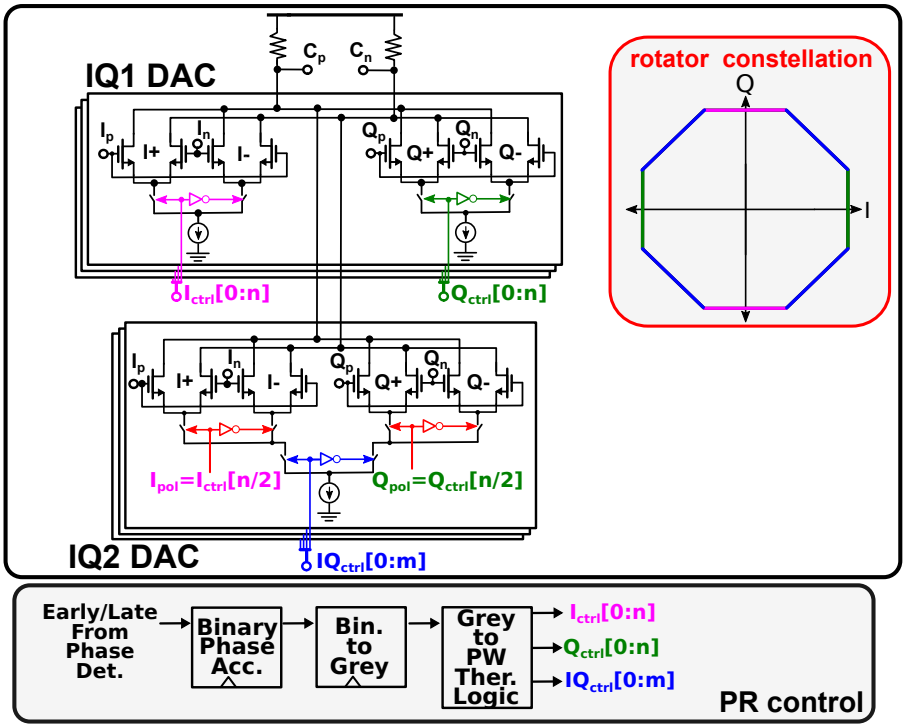

Figure 3.37: PR schematic, control logic and constellation diagram of the phase rotator presented in [46]

operation as well as simplifying the PR control logic as described below.

The schematic of the proposed PR, its control logic, and the constellation diagram are given in Fig. 3.38. The main difference is in the control of the diagonal segments of the constellation. In [46], the diagonal transitions are carried out by switching one unit current from I to Q directly by the control bits $IQ_{ctrl}[0:m]$, or vice-versa. In the proposed solution, the IQ2 DAC is replaced by two DACs (IQ2 and IQ3 in Fig. 3.37) that do not have polarity switches. The diagonal transitions are carried out by switching one unit current on both I and Q vectors, effectively creating a diagonal step. The number of IQ DACs and the required number of control bits have increased by a factor of 1.5 with the proposed change. However, the improvements in PR performance and control logic outweigh the overhead of increase in number of stages.

Some of the advantages of the proposed PR can be listed as follows:

1. Glitch-free operation: As the polarity switches are removed, the phase glitch observed during quadrature transition of the PR in [46] does not exist in the proposed architecture. Quadrature transitions are identical to any other transition and do not require any special treatment.

2. Simplified control logic: The control logic, previously consisting of a binary-phase accumulator, a binary-to-grey converter, and a grey-to-piecewise ther-

mometric decoder as given in Fig. 3.37, is simplified into a single bidirectional shift register with a single inversion as extra logic circuitry. The bidirectional shift register can be constructed from a single tile consisting of one flip flop and one digital MUX. Moreover, a single bit change is guaranteed by design contributing to glitch-free operation.

3. Faster update rate of the control logic: This shift register directly generates the required piecewise thermometric control bits without any complicated logic operation such as binary-to-grey conversion. This allows the control logic to be clock at very high speeds, such as 15 GHz c4 clock in this work. This is an important advantage especially in burst mode CDR which will be explained in Chapter 4, since the higher update rate allows faster phase search.

4. Reduced latency: As the control logic is not pipelined and can be run at very high speed, its contribution to the CDR loop latency is minimized.

5. Improved scalability: Changing of the length of the bidirectional shift register changes the resolution directly. In [46], increasing the resolution by 1 bit would require a complete redesign of the binary-to-grey and grey-to-piecewise thermometer decoder.

6. Improved portability: The control logic is very easy to port across many processes due to its tile-based simple structure.

7. Reduced number of cascode transistors: The removal of polarity switches reduces the number cascoded transistor, allowing larger headroom for PR to operate.

8. Reduced parasitic capacitance at common source: The parasitic capacitance introduced by the polarity switches does not exist in this architecture, improving the high frequency performance of the PR.

3.3.4 IQ Generator

Quadrature oscillators are widely used to generate 90° phase signals. In [47] and [48] the presented structures are named 'tetrahedral' oscillators and the functional dynamics are explained in a rather complicated way. In [49] the same circuit is drawn as a 2-stage pseudo differential ring oscillator (2xPDRO). The latter version is easier to analyze, hence in this study the latter notation will be used.

The evolution of the 'tetrahedral oscillator' into a 2xPDRO is given in Fig. 3.39 together with its conceptual block diagram. The cross coupled loading inverters

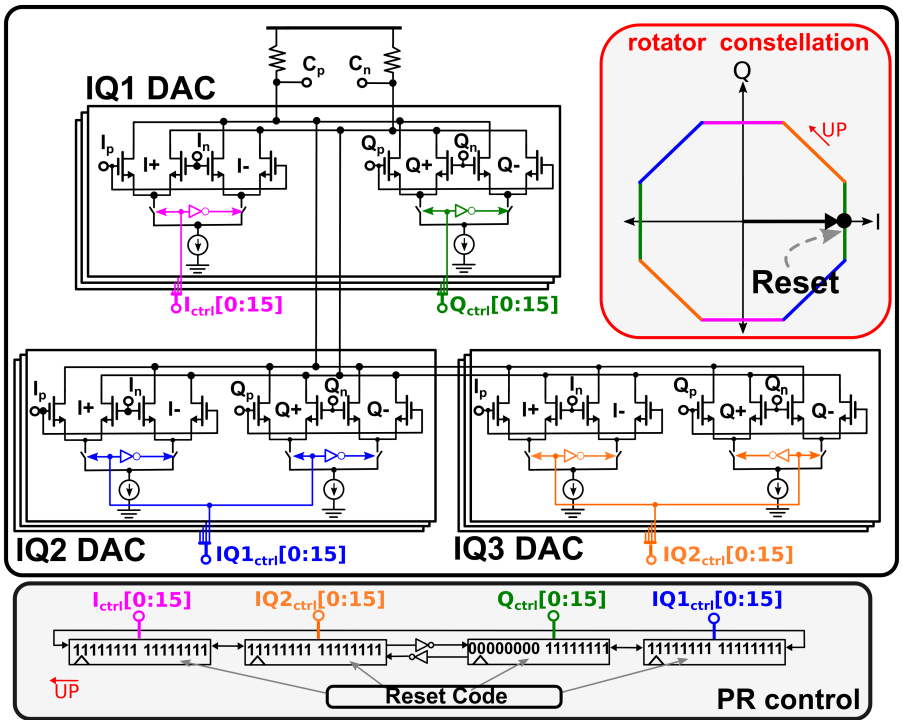

Figure 3.38: PR schematic, control logic and constellation diagram of the proposed PR

(sized K) in 2xPDRO has two main functions: 1) They couple the two main inverter (sized 1) outputs converting 2 single ended inverters into a 1 stage pseudo differential amplifier. It should be noted that when $K = 0$ the oscillator becomes a positive feedback loop consisting of 4 inverters of size 1 and does not oscillate. In that case $I = IB$ and $Q = QB$. Eventually, K must be increased to the point that the positive feedback loop is broken. Transistor level simulations show that a $K > 0.5$ should be satisfied for oscillation in the 14 nm technology with nominal supply voltage. 2) They create a hysteresis in the transfer function of the 1 stage pseudo differential amplifier. This hysteresis prevents a direct small signal analysis of the system. On the other hand, one can model this hysteresis as a delay element whose value is defined by the time it takes for the differential input to reach the hysteresis threshold level starting at 0 analog level. The hysteresis thresholds and how it is mapped as the delay of the 1 stage pseudo differential amplifier is illustrated in Fig. 3.40. It must be noted that this additional delay due to hysteresis allows a 2 stage amplifier with negative feedback to oscillate. A standard 2-stage amplifier with negative feedback is unlikely to oscillate since it has only 2 main poles resulting in a phase change of only 180° at infinite frequency. However, for the oscillation to occur there must be more than 180° phase shift at a finite frequency with a higher than unity gain. That's why at least 3 gain stages are used in ring oscillators. The introduction of additional delay in

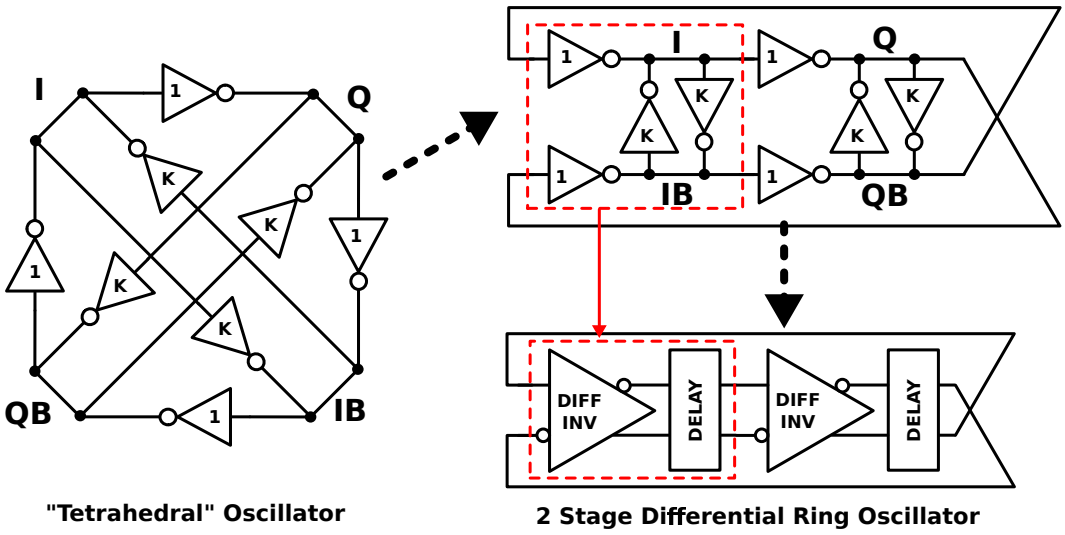

Figure 3.39: Quadrature Oscillator

the loop would lead to a larger than 180° phase shift even in a 2-stage amplifier and would result in oscillation if other criterion are also met.

In [47] a quadrature oscillator based injection locked oscillator (ILO) is presented. The 'tetrahedral' figure is used in the publication with one additional inverter to increase the number of control parameters and a model is derived to analyze how cascaded stages improve the phase error performance over a wide input frequency range.

A simplified ILO with multiple stages in 2xPDRO notation is given in Fig. 3.41. The size of the injection inverter (J) determines the tradeoff between phase correction factor per stage and frequency tracking bandwidth. As J increases the frequency tracking bandwidth increases and error correction factor decreases.

At quarter rate, edge based CDR technique requires 45° phases rather than 90° generated by the presented 2xPDRO. The conventional solution to this problem is to use two separate clock paths for data and edge sampling requiring 2 phase rotators with different control signals (to satisfy 45° phase offset between edge and data clocks) and 2 quadrature generators. However, this solution has several drawbacks. First of all, the integral non-linearity (INL) of the 2 PRs will be uncorrelated leading to a factor of $\sqrt{2}$ increase in effective INL. Moreover the random jitter generated on the 2 separate clock paths will also be uncorrelated increasing the effective random jitter created on the clock paths by $\sqrt{2}$ as well. Instead, this study proposes generating 45° phases directly from a single 4 stage ILO.

A 4xPDRO with injection inverters (J) whose schematic and functional block diagram is given in Fig. 3.42 is a candidate for generating 45° phases directly. Nevertheless, the delay introduced by the cross coupled inverters of this structure

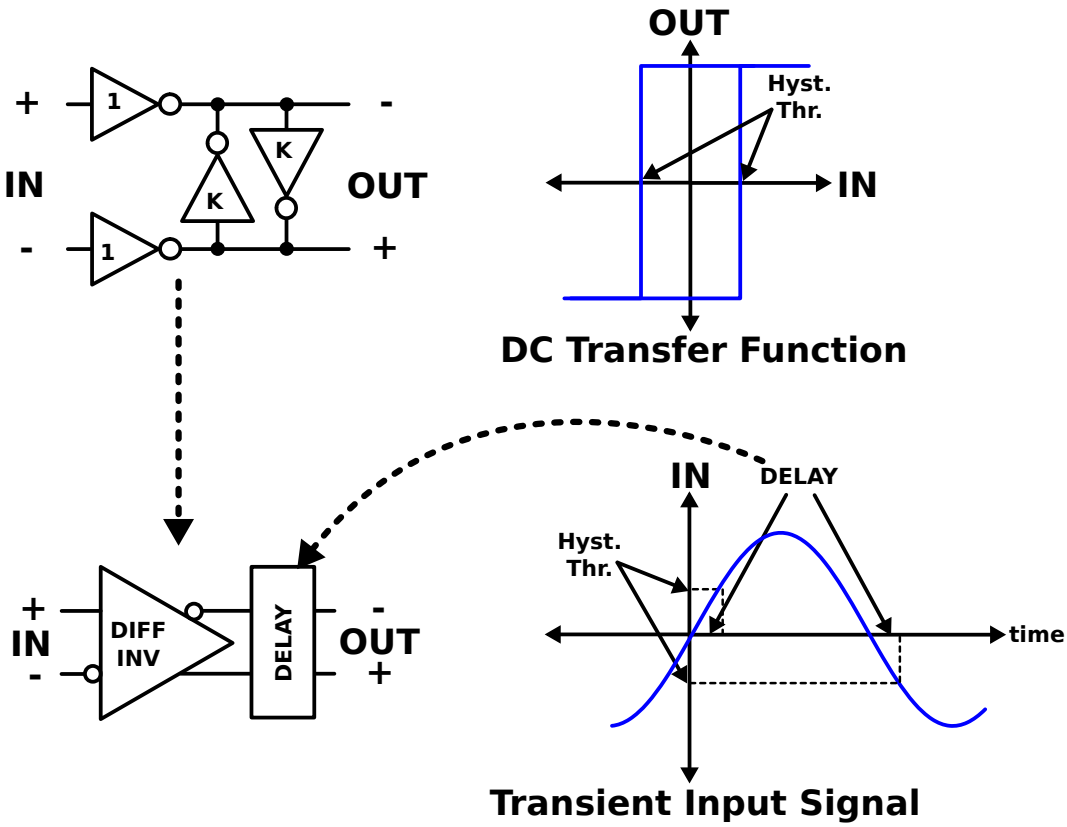

Figure 3.40: Modeling of 1 stage pseudo differential amplifier

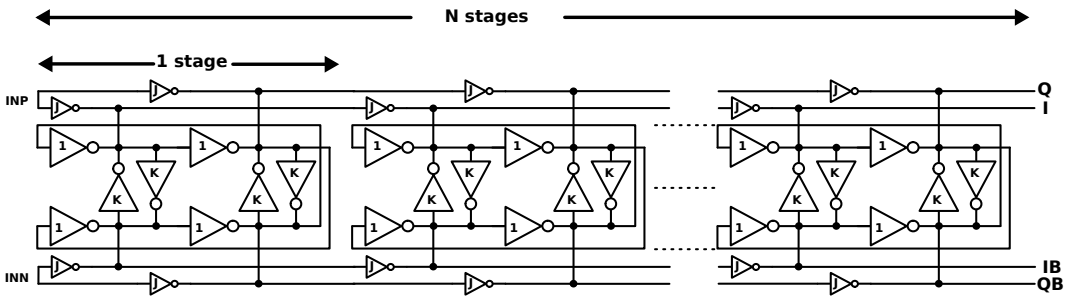

Figure 3.41: Multistage ILO

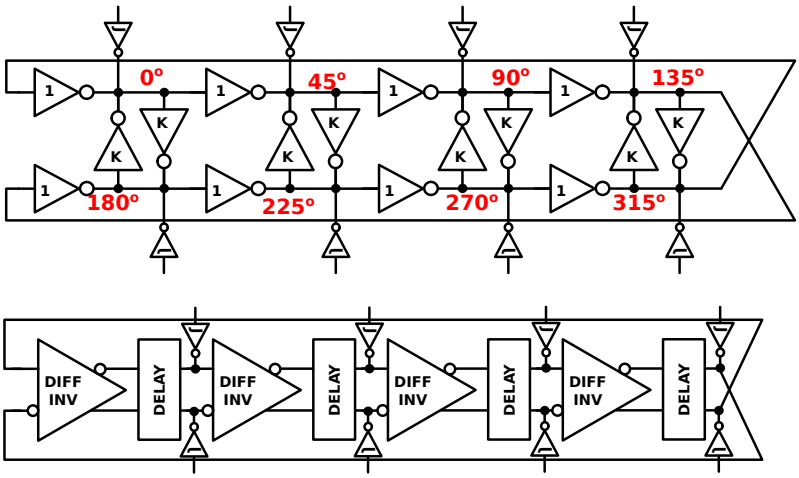

Figure 3.42: The schematic and functional block diagram of a 4xP-DRO

limits the maximum oscillation frequency significantly. It is also not possible to remove the hysteresis (by setting K=0) in order to reduce the delay it produces since the feedback gain of the loop becomes positive as mentioned above. On the other hand, a CML stage is inherently differential and there is no need to introduce hysteresis that results in additional delay in the loop. As a result, a 4-stage CML based ring oscillator (CML-RO) can be run at a much higher frequency compared to the CMOS 4xPDRO alternative. Therefore a 4 stage CML-RO with frequency injection was used in this work. An additional benefit of using a CML based ILO is its higher power supply ripple rejection ratio (PSRR).

The implemented IQ generator architecture used to generate the edge and data phases is given in Fig. 3.43. At its input there are 4 CML stages that generate a coarse estimate of 45° phases to drive the first ILO stage. Following that, there are 3 cascaded ILOs that gradually suppress the phase error introduced by the coarse estimation to produce fine 45° phases for data and edge sampling. Each ILO has a 4 stage CML based RO with injecting CML pairs (INJ CML). The ratio of RING CML to INJ CML determines the trade-off between the error suppression factor and frequency locking range: as the INJ CML get stronger, the frequency locking range increases but error suppression factor decreases. In this implementation, the injection pair is sized half the main RING CML, which satisfies large enough locking range and good error suppression at the end of 2 cascaded stages. The natural oscillation frequency of the RO can be adjusted by changing the tail current and load resistance via 2 digital control bits. Finally CML buffers are placed to drive the following stage.

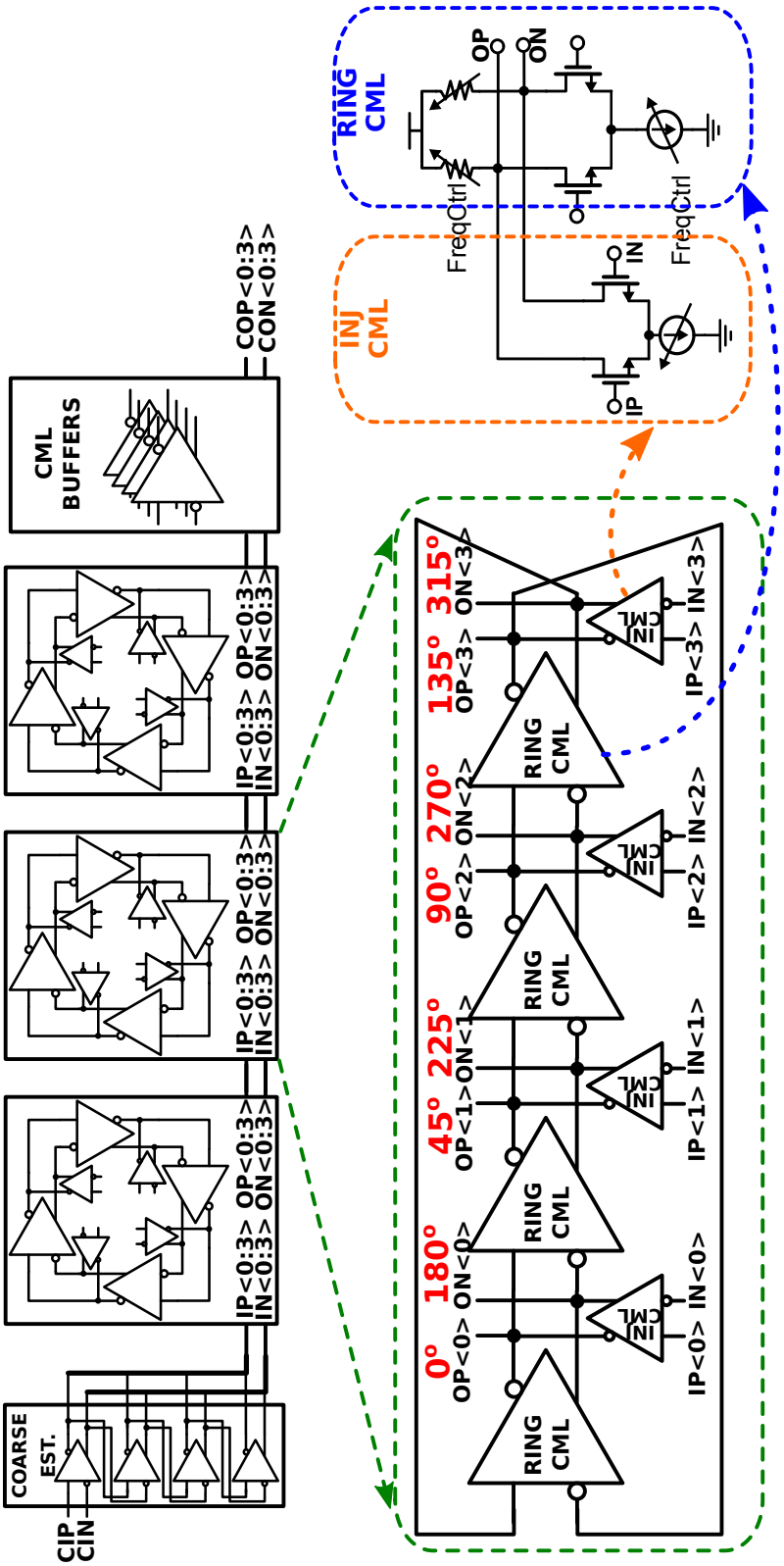

Figure 3.43: Schematic of the IQ generator

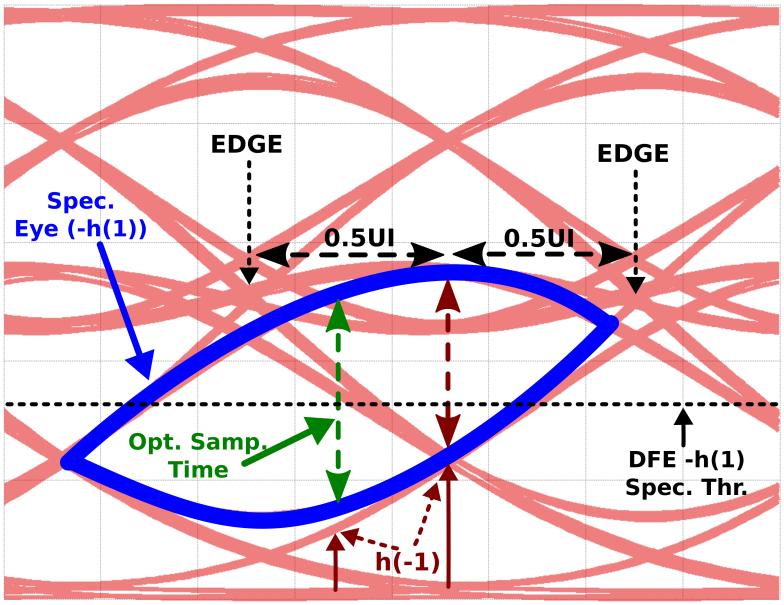

Figure 3.44: Optimal sampling point with the existence of h(-1) on DFE eye

3.3.5 IQ Correction

DFE technique enables the cancellation of post cursors introduced by the limited bandwidth of the data path. However, most data paths have at least a second order transfer function leading to non-zero pre-cursor(s) which cannot be canceled by the DFE. Even so, the effect of the pre-cursor can be reduced by shifting the data phase back in time, just as in this work. The improvement obtained by shifting the data phase is illustrated in Fig. 3.44 on the measured eye diagram of the implemented RX at 60 Gb/s. As the data sampling point moves towards left of point between the two edges, the size of the pre-cursor ($h(-1)$) drops faster than the main cursor ($h(0)$) size, resulting in larger signal.

This shift in time is implemented with the IQ correction block. The block diagram of the IQ correction block is given in Fig. 3.45 together with the data to edge space timing diagram. It comprises 4 identical differential phase interpolators. Each one has 5 digital control bits to adjust its output within a range of ±0.25 UI with a resolution of 0.5 phase rotator step (260 fs at 60 Gb/s). Thus a 1-UI data to edge spacing is supported. During measurements the optimal time-shift is found experimentally with a sweep.

Since the control signals of the phase interpolators are independent, any residual phase error (for example due to mismatch in IQ generator transistors) can also be canceled by the IQ correction block, if necessary.

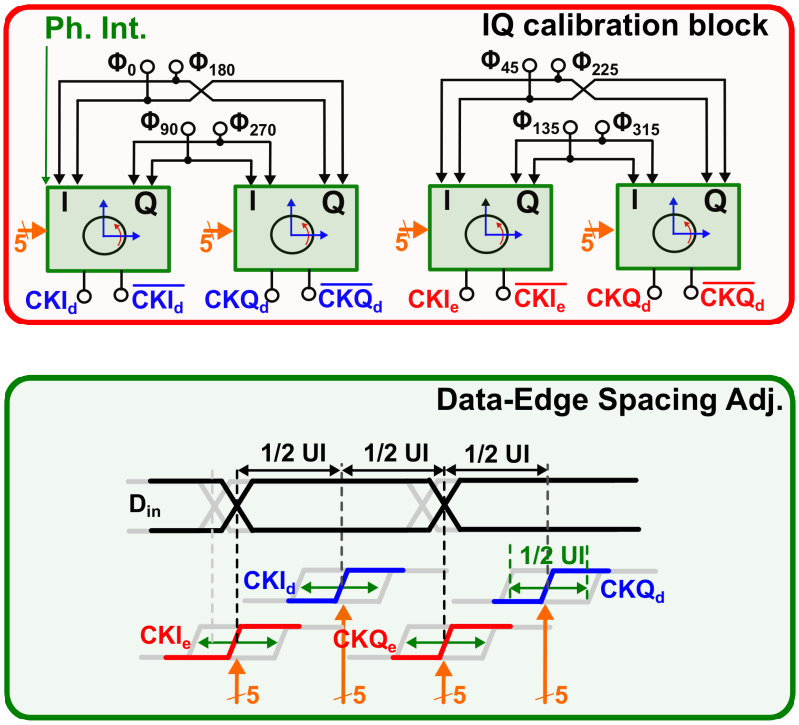

Figure 3.45: IQ correction block

Chapter 4

Adaptive Receiver Design

As was explained in Chapter 1, the link utilization in big data centers is on the orders of a few percent only. In the rest of the time the I/O links are idle consuming power and eating up the thermal budget of the system. On the other hand, an adaptive I/O link can be powered down during idle operation and powered up only when there is bandwidth demand, allowing a larger thermal budget for the rest of the system.

This chapter explains the design details of the proposed adaptive optical receiver. The adaptivity functionality is achieved by modifying the optical RX which was previously presented in Chapter 3.

This chapter is organized as follows: In Section 4.1, various system level design choices to achieve adaptivity will be analyzed and compared. Then, in Section 4.2, the data protocol that is proposed to optimize the adaptivity performance is explained. In Section 4.3, the top level block diagram of the adaptive optical RX is given and the rapid power-on operation is illustrated. And finally, in Section 4.4, the burst mode CDR that is implemented to minimize the phase lock is analyzed.

4.1 System Level Design Options

In the rapid power on operation of the RX, there are two main tasks to be completed before the valid data is received: the biases of the analog blocks need to be settled to their nominal operating points and CDR should be locked to the incoming data. These two tasks determine the power on time. Among the two tasks, the CDR locking is expected to be the bottleneck for power-on time, as it requires the CDR loop to settle to its final operating point. Thus, minimizing the CDR locking time is quite important for rapid power-on performance. Below, various solutions to minimize the power-on time are analyzed and compared in terms of performance, stability and design complexity.

4.1.1 One Lane Always On

The first design approach relies on the fact that the I/O links are mostly used in multichannel applications (such as PCI Express x4 and PCI Express x16).

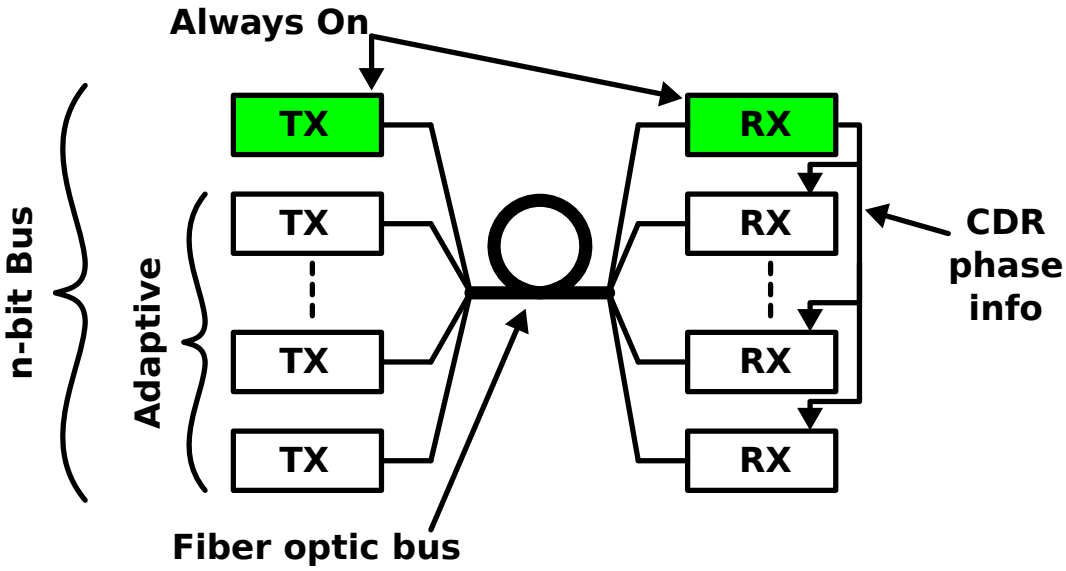

Figure 4.1: Multichannel link block diagram for one lane always on approach

Assuming that latencies of all the lanes in the fiber optic bus are constant with respect to each other, one of the links can be kept powered-on all the time to track the changes in the phase of the incoming data. As a result the CDR of the RX of this lane would be locked to the incoming data all the time. The other lanes in the bus are adaptive and powered down during idle time. When the bus is activated the CDRs of the adaptive receivers would be initialized at the correct point thanks to the phase information provided by the always on lane. This way, the CDR lock time is minimized. This solution is illustrated in the block diagram provided in Fig. 4.1.

However, this approach has several drawbacks. First of all, since one lane is kept on all the time, the minimum power consumption is limited to $1/n$ (when utilization ratio is 0) of the active power, where n is the bus width. Thus, for buses with few lanes the idle power is still significant. The other drawback stems from the fundamental assumption that the latencies of all channels in the bus will be constant with respect to each other. Nonetheless, this is not necessarily true for all applications. Especially longer optical cables will have higher latency variations due to mechanical handling or temperature variation of the environment. As a result, although the power-on time can be minimized with this approach, its use is limited to the applications where optical connection is relatively short and environment is well controlled.

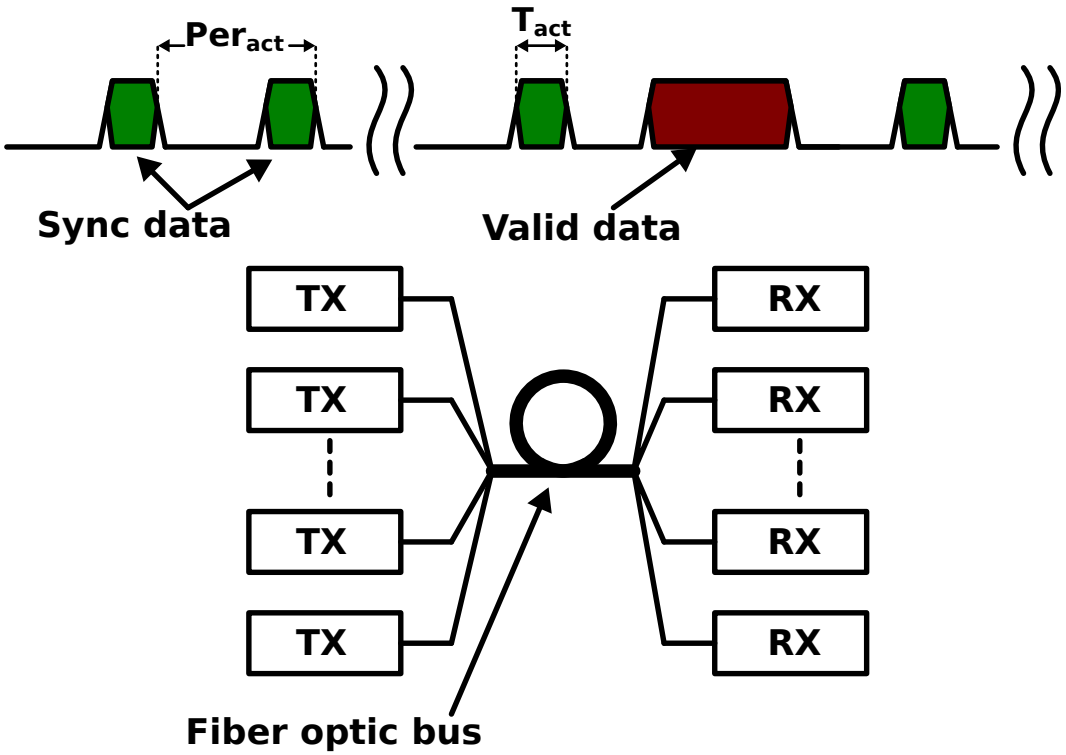

Figure 4.2: Multichannel link block diagram for periodic power on approach

4.1.2 Periodic Power-on

Another solution to rapid power on problem can be to periodically power-up the optical link such that it never loses the phase lock. The block diagram and signal characteristics for this approach is given in Fig. 4.2. Unlike the previously explained approach, all of the links are treated the same way: they are all powered-up periodically so that each RX is synced to its own incoming data. This solution removes the requirement of latencies for different channels in a bus being the same. Moreover the power saving can be significantly improved even for the buses with few bits by keeping the active time percentage low during the idle operation.

In this approach the power consumption in idle mode is given by:

$$P_{idle} = P_{ON} \frac{t_{act}}{Per_{act}} \tag{4.1}$$

where P_{ON} is the power consumption of the RX when it is active, Per_{act} is the activation period, and t_{act} is the active time during synchronization. Thus, in order to minimize power overhead t_{act} should be minimized and Per_{act} must be maximized. t_{act} is limited by the time it takes the analog voltages to be settled, CDR to update the sampling phase, and finally the digital logic to recognize the end of synchronization data and power off the RX. Under optimum conditions,

all those operations would realistically take around 10 ns.

In source synchronous systems the main sources of phase drift between RX and TX are environmental changes such as temperature and fiber-optic cable position. All those changes result in very slow phase drifts allowing Per_{act} to be set to relatively high values (10s of μseconds) compared to $t_{act} = 10ns$. Hence, the power overhead of this approach can be as low as 0.1%.

Nonetheless, the analysis provided above is valid only for source synchronous applications. In order to find the maximum allowed Per_{act} for non-source synchronous case, the accumulated timing error in n UIs is given by the following equation:

$$t_{ErrAcc} = \frac{n}{fclk_{RX}} - \frac{n}{fclk_{TX}} \tag{4.2}$$

where $fclk_{RX}$ and $fclk_{TX}$ are the full-rate clock frequencies of receiver and transmitter, respectively. Eq. (4.2) can be reorganized as

$$t_{ErrAcc} = \frac{n}{fclk_{RX}} \frac{(fclk_{TX} - fclk_{RX})}{fclk_{TX}} = \frac{n}{fclk_{RX}} f_{OS} \tag{4.3}$$

where f_{OS} is the frequency offset. In order to receive error-free data on the receiver, the accumulated time error must be smaller than half of the timing margin of the receiver:

$$t_{ErrAcc} < \frac{t_{margin}}{2} \Longrightarrow nUIf_{OS} < \frac{t_{margin}}{2} \Longrightarrow n < \frac{t_{margin}}{2UIf_{OS}} \tag{4.4}$$

where UI is equal to $1/fclk_{RX}$ and t_{margin} is the timing margin measured on the RX (with a bathtub measurment) for a certain error-rate.

The condition given in Eq. (4.4) sets an upper bound to the Per_{act} with respect to the eye opening in the RX and frequency offset. As an example, placing the optimistic values of 0.2 UI timing margin and 100 ppm frequency offset in the Eq. (4.4), the maximum value of n is found as 1000. At 56 Gbps data-rate, this value allows a maximum off time of 17.8 ns only! So, considering a synchronization time (t_{act}) of 10 ns, the activation period of this technique is limited to 27.8 ns and power consumption is approximately $0.36P_{ON}$, even with optimistic assumptions.

4.1.3 Fully Automated Power-On Per Lane

The third approach to rapid power-on problem is to design fully automated and independent RXs that can sense an incoming data, power themselves on and quickly find the phase of the incoming data via a burst-mode CDR. Unlike previous approaches, it does not suffer from the power overhead of keeping one lane always on or of periodic power-on. Thus, the idle power is independent of the

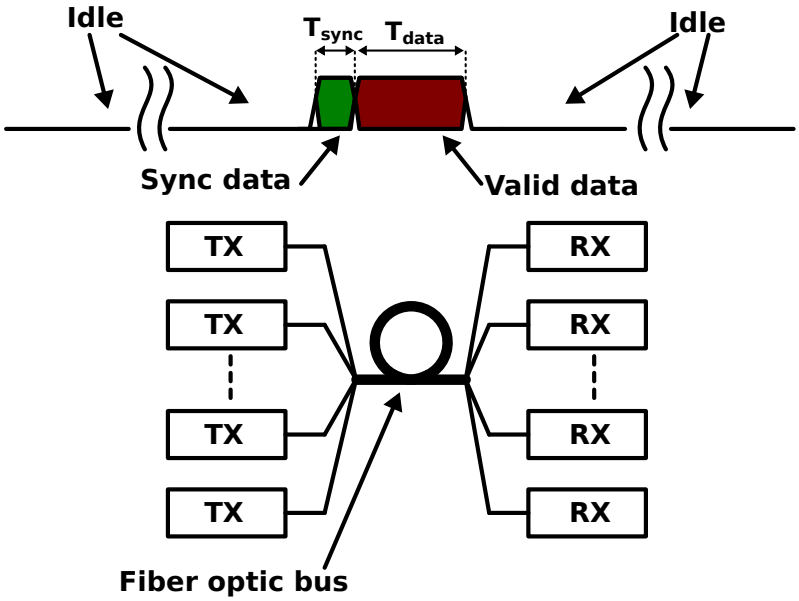

Figure 4.3: Multichannel link block diagram for fully automated power-on per lane approach

on-state power consumption. The RX can be designed such that, the only power consumption in the off-state (P_{OFF}) is due to the circuit used to sense the beginning of incoming data during and leakage currents.

The power consumption for this solution is given by the equation:

$$min(P_{ON},(P = UR\frac{T_{sync}+T_{data}}{T_{data}}P_{ON}+(1-UR\frac{T_{sync}+T_{data}}{T_{data}})P_{OFF})) \tag{4.5}$$

where UR is the utilization ratio, T_{sync} is the time required to power on RX and find the optimum sampling point, and T_{data} is the duration of the valid data. The variables UR and T_{data} are determined by the system that uses the adaptive I/O link. Having those variables fixed, in order to minimize the total power consumption P_{ON}, P_{OFF}, and T_{sync} need to be minimized.

Another advantage of this solution is that there is no need for communication between the RXs in the bus, which reduces the complexity and makes the solution scalable to any bus size.

4.1.4 Comparison and Design Choice

Three different approaches to the adaptive RX design were presented and analyzed in the previous subsections.

The first one, one lane always on, offers a limited power saving especially in applications where the bus consists of only few bits. Moreover its operation relies

on the assumption that the delay variation between the lanes of the same bus is small, which is not necessarily true for applications that need long fiber-optic cables.

The second approach, periodic power on, does not suffer from delay variation problem as each RX independently locks its own phase to the incoming data. It also has a potential for huge power saving, but only for source synchronous applications. In other applications, either the power saving is very limited or the periodic power on technique is not applicable at all for higher frequency offsets.

The third one, fully automated power on per lane, has the highest potential for power saving as all the lanes can be powered down for the whole duration of the idle time. On the other hand, in order to get the maximum performance the synchronization period before the valid data needs to be minimized which may be achieved through a burst mode CDR leading to higher design complexity.

In this work fully automated power on per lane approach is chosen to be implemented since it has the highest potential for power saving and widest range of applications. In the following sections, the techniques to minimize the power-on time will be analyzed in detail.

4.2 Defining the Data Protocol

In order to minimize the power consumption of the optical IO link, and CDR lock time of the RX a data protocol is proposed. The protocol also marks the beginning and end of the valid data and allows a precise BER check and power-on time measurement as will be explained in Chapter 5. The implemented data protocol is given in Fig. 4.4. In this section, the reasons affecting the data protocol definition will be explained in detail.

The optical signal generation occurs in the VCSEL. Its bias current is modulated to generate high and low power optical levels which correspond to the digital levels of 1 and 0. Especially, for high data-rate applications, the VCSEL biasing and modulation comprises a significant portion of the total power consumption of the optical transmitter. Thus, the VCSEL needs to be set at the minimum biasing condition during idle time, which leads to the conclusion that the data

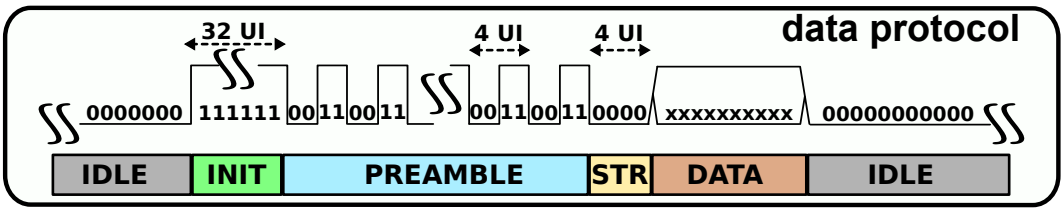

Figure 4.4: Data Protocol

transmitted during idle time should be 0. A custom design TX for adaptive optical links can further reduce the VCSEL bias to generate a level much below the 0 level of active period. This would still be interpreted as a 0 in the RX side, still satisfying the data protocol.

When the VCSEL bias is reduced for long periods of time (during no data transmission) the VCSEL starts to lase only on one optical mode. Under this initial condition, increasing its bias to start transmitting data will not result in immediate increase in optical power which was previously reported in [50] and [51]. A detailed explanation for this behavior is given in [52] as follows: At low bias, the VCSEL lases only on one optical mode; when its bias is increased, this mode will saturate and burn a spatial hole into the gain, which will prevent further optical power increase. At the same time, there is not yet sufficient gain outside this hole for a higher order mode to start lasing. Therefore, a short time delay is observed before the other modes kick in and continue the optical power increase. During that time, a rapid rise in the junction temperature occurs locally, shifting the threshold of all modes, most likely to higher threshold currents, which will result in a further delay in all modes turning on.

Also measurement results are provided to further investigate the turn on behavior of the VCSEL in [52]: Assuming a balanced signal right after the idle period, the turn on delay is measured for the 2 proposed bias conditions in idle mode: 3 μs of logic 0 level and 3 μs of reduced bias level (lower than logic 0). The measurement results are given in Fig. 4.5. The turn on time for the VCSEL is approximately 30 ns and 40 ns, respectively. Considering typical Ethernet packages of 10 ns to 256 ns lengths, 40 ns of VCSEL settling time corresponds to power overheads of 400% to 23%.

The insertion of a wake-up pulse right before the balanced signal reduces the turn-on time of the VCSEL significantly. The improvement with a 4 ns wake-up pulse is given in Fig. 4.6. The optical signal is already stable right after the initial wake-up pulse: almost 10X faster turn-on. Applying a stronger wake-up pulse (higher level than logic 1) reduces this delay even further. It must be noted that, since the duration of this strong wake-up pulse is short it does not create any reliability issues [53].

Since this work is focused on the adaptive RX design, the VCSEL bias levels on the TX side were kept constant at levels of logic 0 and logic 1 during measurements. The reduced bias and strong wake-up pulse techniques were not used. The wake up pulse consisted of a certain number of logic 1 data. The length of this pulse depends on parameters such as: the VCSEL type and bias, the data-rate, and nominal modulation amplitude.

The implementation of reduced bias and strong wake-up pulse features into a custom design adaptive TX is not expected to increase the circuit complexity

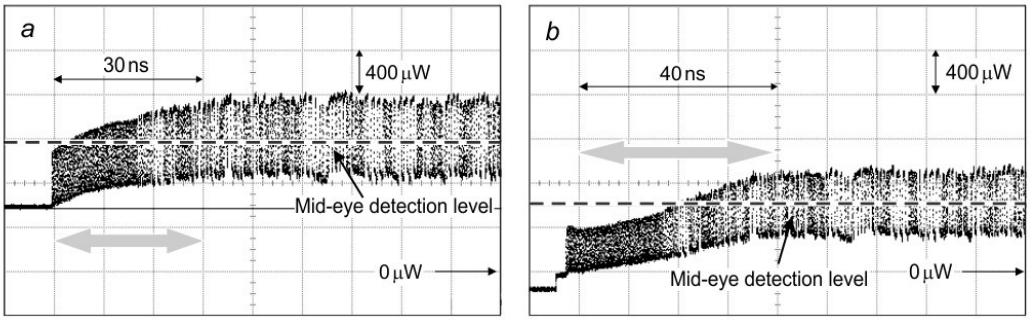

Figure 4.5: Measured turn-on delay of an MM-VCSEL at T=35°C [52] for idle bias conditions: a) logic 0 b) reduced bias

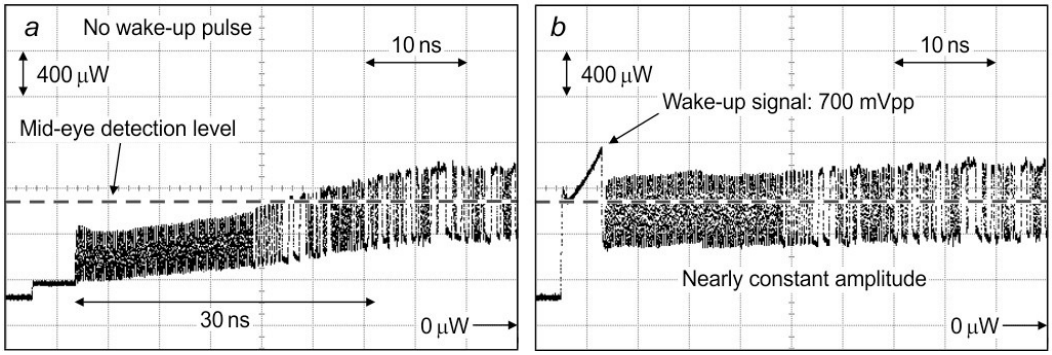

Figure 4.6: Measured turn-on delay of an MM-VCSEL at T=35°C [52] for with reduces bias in idle mode: a) balanced signal at data transmission b)wake-up pulse of logic 1 for $4ns$

significantly.

After the wake-up pulse, VCSEL settles to its nominal operating point and the optical signal levels at the RX side are well defined. Now, the CDR in the RX needs to lock to the phase of the incoming data. The operation of CDR relies on the information obtained from the transitions on the incoming data. The more transitions there are, the faster the CDR locks. Hence, in order to minimize the CDR lock time the transition density of the input signal needs to be maximized. The straightforward choice to satisfy this need is to send a preamble signal of ″..0101..″, which gives the highest possible transition density in NRZ signaling. However, as described in Section 3.2.1, in order to increase the sensitivity, the analog front bandwidth was limited. As a result, a preamble of ″..0101..″ generates a very small signal (Fig. 4.7) for the edge comparators of the bang-bang CDR to detect reliably. An optimized point for the trade-off between the transition density and signal amplitude at the AFE output is found to be a preamble of ″..0011..″. Although the the transition density is half of the previous preamble signal, it has a much higher swing and sharper transitions that increase the SNR

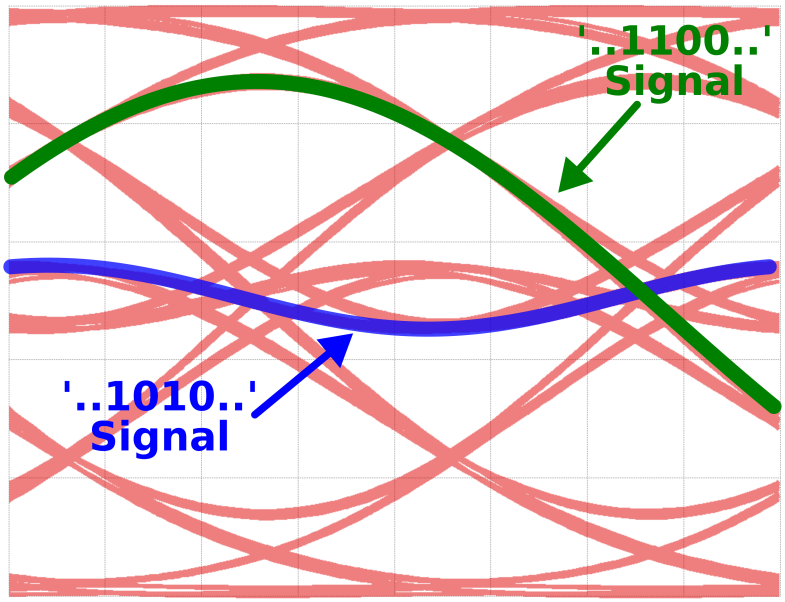

Figure 4.7: The 2 preamble signal options (″..0101..″ and ″..0011..″) are marked on the simulated eye diagram of the AFE

in phase detection significantly as illustrated in Fig. 4.7. The exact preamble length is not determined at this stage of the design as it depends on the CDR performance and actual transmitted signal quality (amplitude, bandwidth etc...).

At the end of the preamble a certain signal must be placed to mark the beginning of the valid data. Since it does not affect the power-on performance of the RX it can be a very short predefined sequence. In this application it was chosen to be ″0000″. The data after this marker is interpreted as the valid data to be transmitted. After the valid data transmission is finished, the TX goes into idle mode, sending a stream of ″0″s to the RX. When the receiver detects 64 consecutive ″0″s, it interprets this as the beginning of another idle period, and powers itself down to wait for the next cycle of the data transmission.

4.3 Adaptive RX Architechture

The adaptive optical RX architecture is a modified version of the RX presented in Section 3.1. The block diagram of the RX is illustrated in Fig. 4.8. The blocks and signals that realize the power on and off functionality are given in red color.

The power cycle of the adaptive RX occurs as follows: When the RX is active, the RX digital block checks for a certain number of consecutive “0″s (64 “0″s in this application) to detect the beginning of an idle period. Once the beginning of an idle period is detected, the RX digital generates the PWROFF pulse to reset the PON SENSE block. As a result, EN signal goes down powering down the

clock path and VGA. When the clock path is powered down, CMOS blocks do not consume any dynamic power, hence they do not need to be powered down explicitly. During the idle period, apart from some biasing auxiliary circuits, only TIA and PON SENSE blocks are active. At the end of the IDLE period the initialization pulse INIT is sensed by the PON SENSE block, which sets EN signal to 1 powering on the clock path and VGA through PON bias boost block. Also the reset of the PON FSM block is released.

PON FSM is a finite state machine that controls the power on algorithm. When it is initialized, it sets the BMDONE signal to '0' enabling the burst mode CDR (BM-CDR) path. Then it runs the BM-CDR algorithm, which will be explained in detail in Section 4.4, and at the end sets BMDONE signal to '1' to activate the bang-bang CDR path for normal operation.

The BMDONE signal is also used to dynamically increase the AFE dc offset cancellation loop bandwidth during power on. This is possible since the input signal is a preamble with a balanced high frequency single tone signal and it does not contain any low frequency components. Eventually, the offset current is canceled within 2 ns, and at the end of the BM-CDR period the bandwidth of this loop is decreased not to create baseline wander in normal operation.

After the BMDONE signal is set high, the RX digital block starts to check the incoming data for the STR sequence ('0000') in order to determine the beginning of the valid data. Once STR is detected, it starts to check the incoming data for the next idle period.

4.3.1 PON Sense

PON Sense is the block that enables the circuit at the beginning of a data burst. During the idle period it is kept active together with TIA.

The schematic of the PON Sense is given in Fig. 4.9. It consists of two main blocks: a differential to single ended amplifier and a set reset (SR) latch. The threshold of the differential amplifier is controlled by the source degeneration resistors.

The set (S) input of the SR latch is connected to the amplifier (WUP) and the reset (R) input is connected to the PWROFF signal generated by the RX digital block. The output of the SR latch (EN) is used as the enable signal for the RX. The signal flow throughout a data burst is also given at the bottom of the Fig. 4.9. Once the WUP output goes to '1', EN output also goes to '1'. After that the WUP signal may go high or low logic values depending on the incoming data. Those changes do not effect the EN output. It must be noted that PWROFF signal is created after 64 '0's of the idle period are detected by the RX digital block, which means that WUP will be at '0' by the time PWROFF rises, ensuring the SR latch

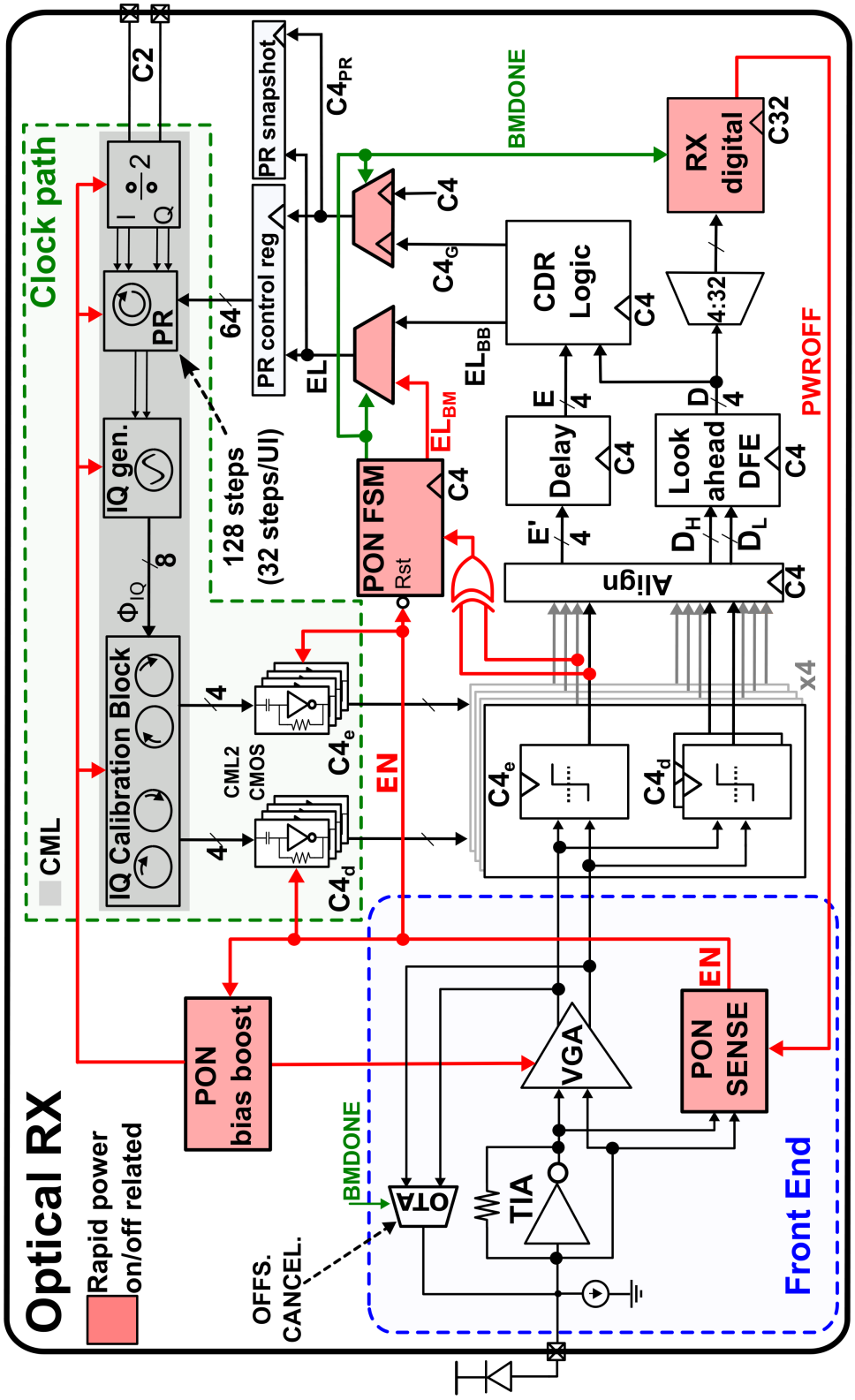

Figure 4.8: Adaptive RX top level block diagram

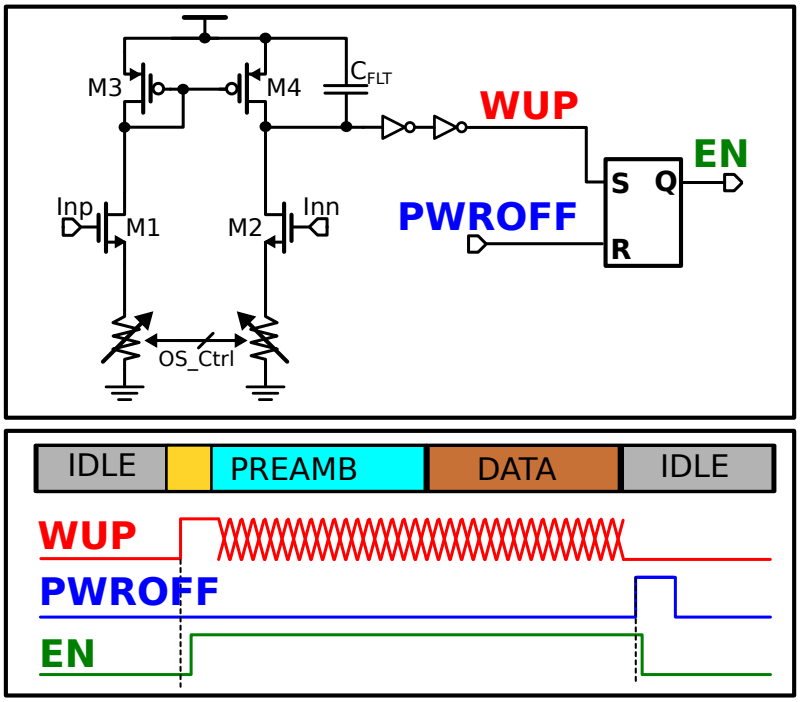

Figure 4.9: PON Sense schematic

does not enter the forbidden state of $S = 1$ and $R = 1$.

4.3.2 Offset Cancellation

There are 2 units that contribute to the cancellation of the common mode current of the photo-diode: the current DAC for the coarse estimation and the analog feedback loop for residual offset cancellation. The current DAC control register is set to a value to cancel the main portion of the DC current from the photo diode and it is not changed during power on and off periods. On the other hand, the analog offset cancellation loop has to be turned off so that it does not try to cancel the incoming signal of constant '0's during the idle mode. Hence, it has to turn on and rapidly cancel the residual current offset at the beginning of the power on event.

During data transmission this loop has to have very low bandwidth in order not to create baseline wander for long streams of '0's or '1's. Yet, due to this constraint, the residual DC current cancellation at the beginning of a data burst could take unacceptably long time. The proposed solution to this problem is to dynamically increase the bandwidth of the analog feedback loop during the preamble period, and decrease it back for normal operation. It must be noted that the high bandwidth offset cancellation loop will not create baseline wander since the preamble does not have low frequency components.

The dynamic bandwidth control is done inside the OTA block, which is connected on the feedback path of the analog front end (Fig. 4.8). The schematic of the OTA is given in Fig. 4.10. Its input stage consists of 2 unity feedback transconductance amplifiers (OTA_{BUF}) to buffer the 2 differential inputs from the highly capacitive loads required for dominant pole generation on V_{XP} and V_{XN} nodes. The buffered signals V_{XP} and V_{XN} are then amplified through the transistors $M1-4$. Two capacitors (C) are connected between the output and input nodes of the amplifier via the source followers consisting of transistors $M5-6$ and their tail currents. By Miller theorem, the capacitive load seen at the nodes V_{XP} and V_{XN} are equivalent to:

$$C_{eq} = (A+1) * C \tag{4.6}$$

where A, is the gain of the amplifier consisting of transistors $M1-4$. Finally the amplified voltage signal is converted to an output current ($IOUT$) via a transconductance amplifier OTA_{OUT}.

The bandwidth switching can be done by changing either the effective capacitance or resistance of the node that determines the dominant pole location. If the bandwidth reduction is done by adding capacitance, then the initial voltage of this additional capacitance needs to be set equal to the node it is connected to. Otherwise it would create a jump in the voltage that has to be canceled by the low-bandwidth loop afterward. In order to avoid this problem, the bandwidth switching is done by changing the resistance which does not store any charge, which means it does not create any disturbance during switching.

The dominant pole comes from the nodes V_{XP} and V_{XN}, and for both configurations its value is given by the following formulas:

$$p_{hbw} = \frac{C_{eq}}{g_{OTA_{BUF}}} \tag{4.7}$$

$$p_{lbw} = C_{eq} * R \tag{4.8}$$

where, $g_{OTA_{BUF}}$ is the transconductance of the OTA_{BUF}.

A transistor level transient simulation that shows the step response of the offset cancellation loop for high-bandwidth and low-bandwidth configurations is given in Fig. 4.11. In the simulation a 50 μA step was applied to the AFE input at 1 ns. In the low-bandwidth case (blue curve) the error signal at the AFE output goes below 5 mV only after 1.19 μs whereas in the high-bandwidth case the same error level was reached only 2.21 ns after the step. Thus the settling time is improved by a factor of more than 500 during the rapid power on.

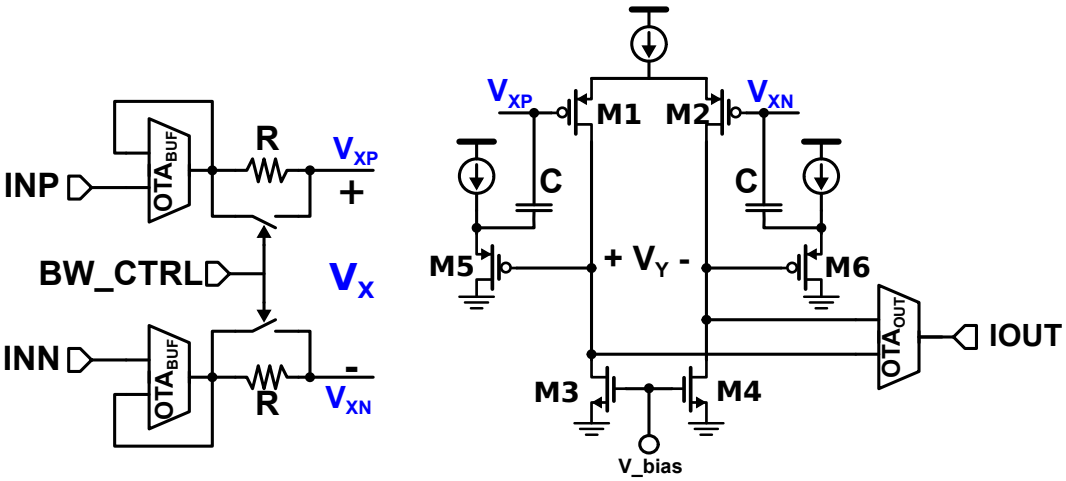

Figure 4.10: OTA for offset cancellation schematic

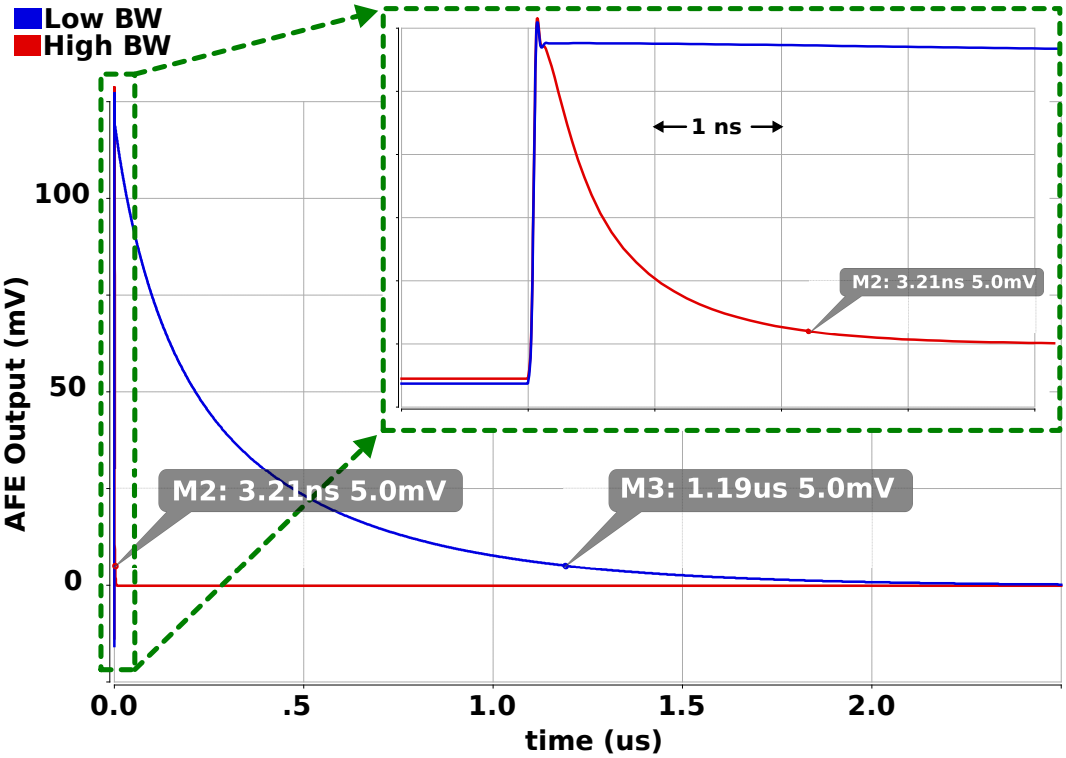

Figure 4.11: Offset cancellation loop step response

4.3.3 PON Bias Boost

During the idle mode, the CML based blocks in the RX are turned off by shorting their tail transistors' gates to ground, setting the bias voltage to 0 ($V_{bias} = 0$). In normal operation, this node is biased with a reference current (I_{ref}) that passes through a diode connected transistor. The reference current is mirrored by the tail transistors of the CML stages with the required ratios. The ratio of the total mirrored current to reference current needs to be kept relatively large ($I_{tot} > 10I_{ref}$) not to waste too much current for biasing only. However, this requirement results in relatively large output resistance of the bias stage (or diode connected transistor). Thus, during power on, the settling time of the bias voltage may compromise the total power-on time of the RX. In order to reduce settling time of this bias voltage, a charge is injected into this node at the beginning of each data burst.

The circuit realizing the charge injection and the resulting bias settling improvement is given in Fig. 4.12. The circuit operates as follows: when EN signal goes down DIS goes up shorting the BIAS node to the ground. As a result the tail currents (I_1 to I_N) for the blocks connected to the BIAS are reduced to leakage currents only. In the mean time the BST node is charged to VDD via the transistor $M7$ resulting in 0 V potential in the boost capacitor C_{BST}. When EN goes up, DIS is pulled to ground via the inverter INV1 turning off the transistor $M9$. During the initial transient BST node is also pulled to ground via the capacitor C_{BST}, turning on $M8$ that provides additional current to charge the bias node. This additional current is only turned off when the BIAS reaches the threshold of $M5$ to turn $M7$ on and charge BST to VDD.

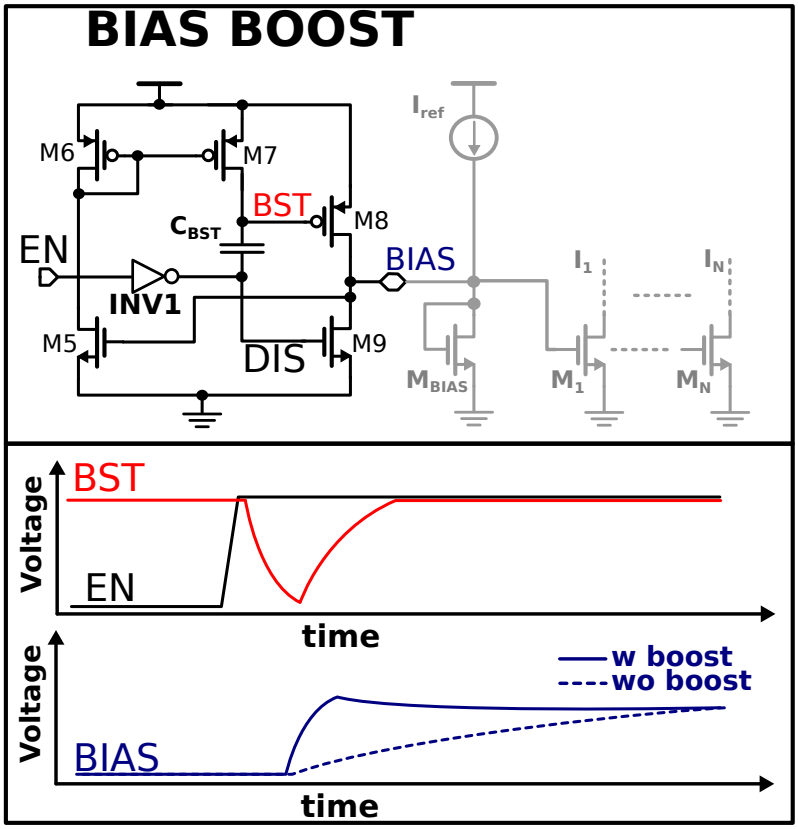

Figure 4.12: Bias boost circuit schematic (top) and bias settling with and without the boost circuit (bottom)

4.4 Burst Mode CDR

The CDR presented in Section 3.3 is designed to process the edges on a random data during normal operation. Thus, it needs the phase detector to filter out the false early or late information from edge samplers (when there is no transition), the majority voter to increase signal to noise ratio on phase detection, and the loop filter to stabilize the CDR loop. All of those additional blocks used to process the data introduce delays that eventually reduce the CDR bandwidth. Using this CDR logic to find the initial sampling phase during power-on may result in relatively long phase-lock periods, increasing the power overhead calculated by the Eq. (4.5). In this section a burst mode CDR (BM-CDR) will be introduced and explained that is used to reduce the initial phase lock time significantly.

The introduction of a deterministic preamble in the data protocol, as described in the Section 4.2, renders phase detection block obsolete as the transition phases are well defined. Hence, the outputs of the edge comparators can be directly used as early or late information without the need to know the outputs from the data samples. Since the data information is no longer needed during preamble, the phase detection block can be removed from BM-CDR path as well as the delay blocks introduced on the edge path to match the look-ahead DFE latency.

In order to further reduce the latency, the majority voter is also removed from BM-CDR path at the cost of increased phase detection noise.

The analysis provided in Section 3.3.1 shows that as the CDR latency is reduced, the loop gain can be increased for larger CDR bandwidth and faster phase lock. The loop filter block, which was described in Section 3.3.2, was introduced to lower the CDR gain via averaging to satisfy the stability in the bang-bang CDR operation. Removing this block from the BM-CDR path increases the loop-gain while lowering the latency even further. The final effect is a compromised loop stability which will be analyzed in the Section 4.4.2.

After all the aforementioned reductions, the block diagram of the burst mode CDR path is shown in Fig. 4.13. The U_D_{BB} and U_D_{BM} are the up/down signals coming from the bang-bang and burst mode CDR paths, respectively. $c4_{G(BB)}$ and $c4_{G(BM)}$ are the gated clocks generated by the bang-bang and burst mode CDR paths, respectively. The rest of the clock path (starting from the PR control logic) is not affected by the burst mode CDR.

When the RX is powered up the finite state machine (PON_{FSM} in Fig. 4.8) chooses the burst mode CDR outputs through the MUXs using the $ModeSel$ signal. At the end of the burst mode CDR this signal is inverted to hand over the control to bang-bang CDR.

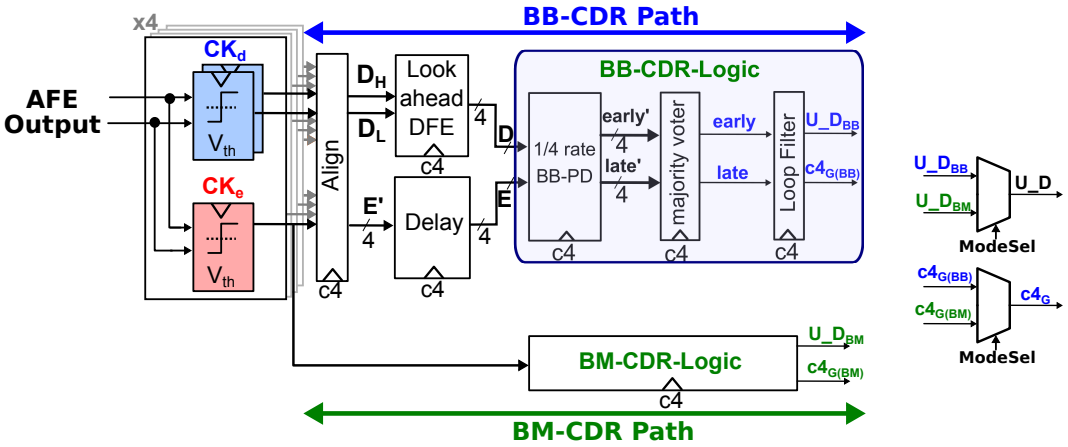

Figure 4.13: Bang-bang and burst mode CDR Paths

4.4.1 Burst Mode CDR Logic

In the existence of a preamble of ″0011″, one of the edge comparator outputs can be used directly as the up/down signal of the BM-CDR and the edge sampling clock can be used directly as the gated clock as given in Fig. 4.14. In this case when the AFE output is above 0, the phase rotator is rotated upwards (towards increasing phase) and when AFE output is below 0 the phase rotator is rotated downwards (towards decreasing phase). The PR rotation direction during burst mode is illustrated in Fig. 4.14 on the preamble signal. The blue arrows indicate an up rotation of the PR, and the red arrows indicate a down rotation. A stable point for CDR operation is where the BM-CDR logic wants to converge onto, and metastable points are the points where the BM-CDR logic will run away from in either direction.

Starting from one of the metastable points the BM-CDR would need to travel 180^o to reach one of the stable points, which would need at least 64 c4 cycles (assuming every decision taken by the edge comparator is correct!). This takes approximately 4 ns at 60 Gb/s.

A simple modification in the BM-CDR logic as illustrated in Fig. 4.15 can reduce the PR travel distance by half, saving 2 ns of CDR lock time. In this improved version, the XOR of the two consecutive edge comparator outputs ($Edge<n>$ and $Edge<n+1>$) is used as up/down signal of the PR control. In that case, all the 0 crossings of the AFE output are stable points and all the peaks are metastable points, with respect to $Edge<n+1>$ position. The maximum travel distance is reduced to 90^o, without any additional latency in the BM-CDR loop.

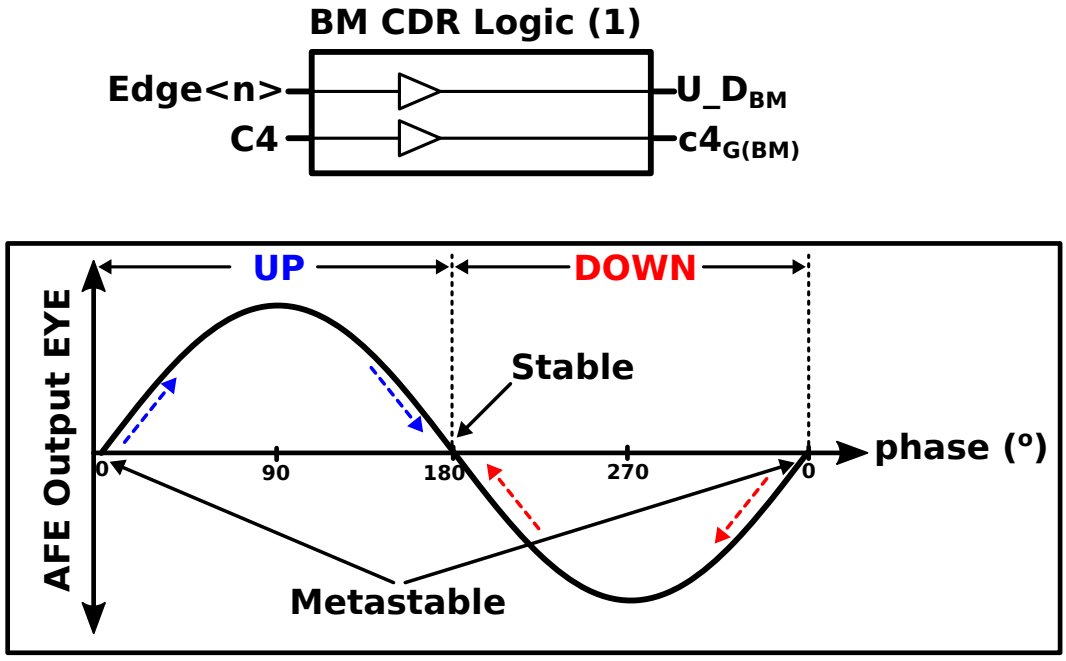

Figure 4.14: Burst mode CDR logic and the PR rotation direction on the preamble signal

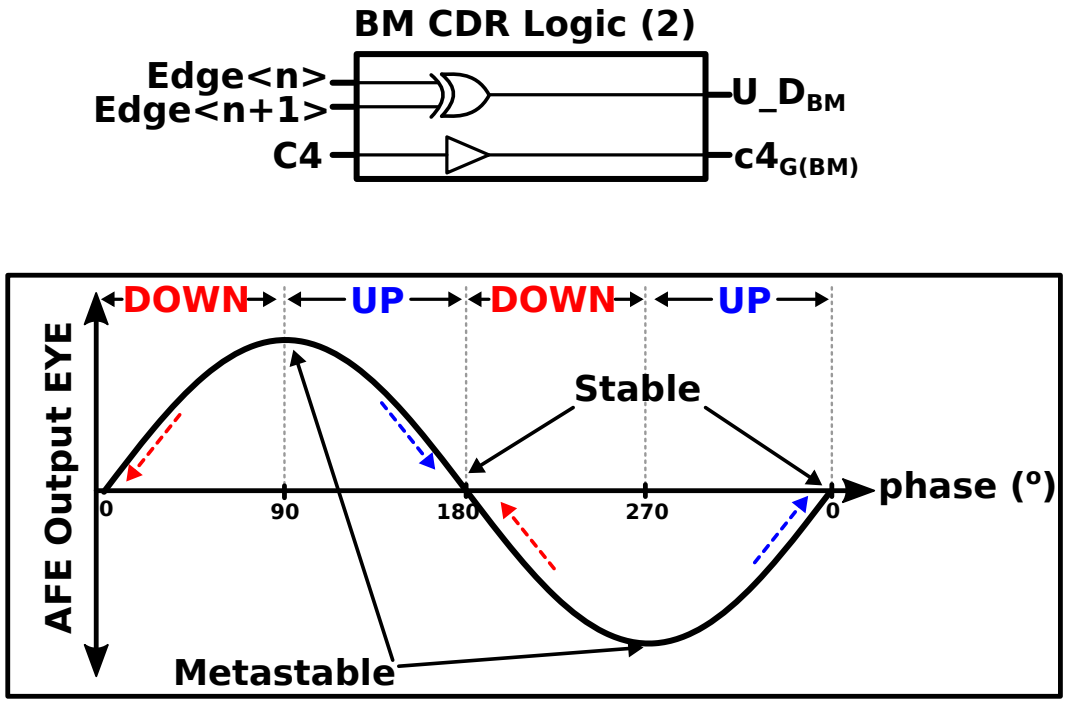

Figure 4.15: Improved burst mode CDR logic and the PR rotation direction on the preamble signal

4.4.2 Metastability Analysis of the BM-CDR Loop

The probability of receiving an $'UP'$ signal with respect to the $Edge < n+1 >$ sampling phase is given in Fig. 4.16 for ideal (black curves) and realistic (green curves) cases. In the ideal case, the probability function is not continuous and the $0 \leftrightarrow 1$ transition occurs as a step as there is no noise. This means all the decisions are correct and the sampling position will move away from the metastable points immediately. On the other hand, in the realistic case, where the AFE output signal is noisy and sampling point is jittery the $0 \leftrightarrow 1$ transition is smooth, which means there is a certain probability that the up/down decision will be erroneous. These errors may increase the convergence time of the BM-CDR loop significantly.

In order to analyze the convergence characteristics of the BM-CDR, the AFE and BM-CDR loop was modeled using python. The model is based on the block diagram given in Fig. 4.17. The AFE model has three poles: one located at 14 GHz and the other 2 are both located at 21 GHz. Its DC output amplitude is normalized to 1 V_{pp} and a parametric white noise (with a nominal value of 20 mV_{rms}) was introduced.

The phase rotator was modeled with a linear transfer function of 32 steps/UI ($1UI = 90^o$) and covers 4 UI (360^o) in 128 steps ($0:127$) just as implemented. Moreover, a loop latency parameter is introduced in the BM-CDR path to account for all the delays in loop. The nominal latency was found from the transistor level simulation and is set as 32 UI. Finally a random jitter parameter (with a nominal value of 250 fs_{rms}) was defined to account for the timing error.

The eye diagram of the AFE during preamble phase is derived and the probability of sampling a $'1'$ on the edge-comparators is calculated for all possible sampling phases in the existence of the aforementioned non-idealities. Then, the probability of an $'UP'$ signal being generated by the BM-CDR logic was found using those outputs. The results are given in Fig. 4.18. The x-axis for all the plots in the figure is the phase and it is given in terms of degrees (^{o}), where $1UI = 90^o$.

The same analysis is repeated for various noise and jitter parameter values to show their effect on the probability of generating an $'UP'$ signal, and the results are given in Fig. 4.19. As expected, as the value of those parameters increase the transition becomes less steep, which means the chances of generating a wrong $'UP'$ signal increases.

Transient simulations were run on python to analyze the BM-CDR loop characteristics. In order to isolate the loop dynamics from the random variation the noise and the jitter values were set to 0, and the PR position was observed for various loop latency values. The results are given in Fig. 4.20. The PR position converges to around 32 which corresponds to one of the stable points ($90^o = 32$ PR

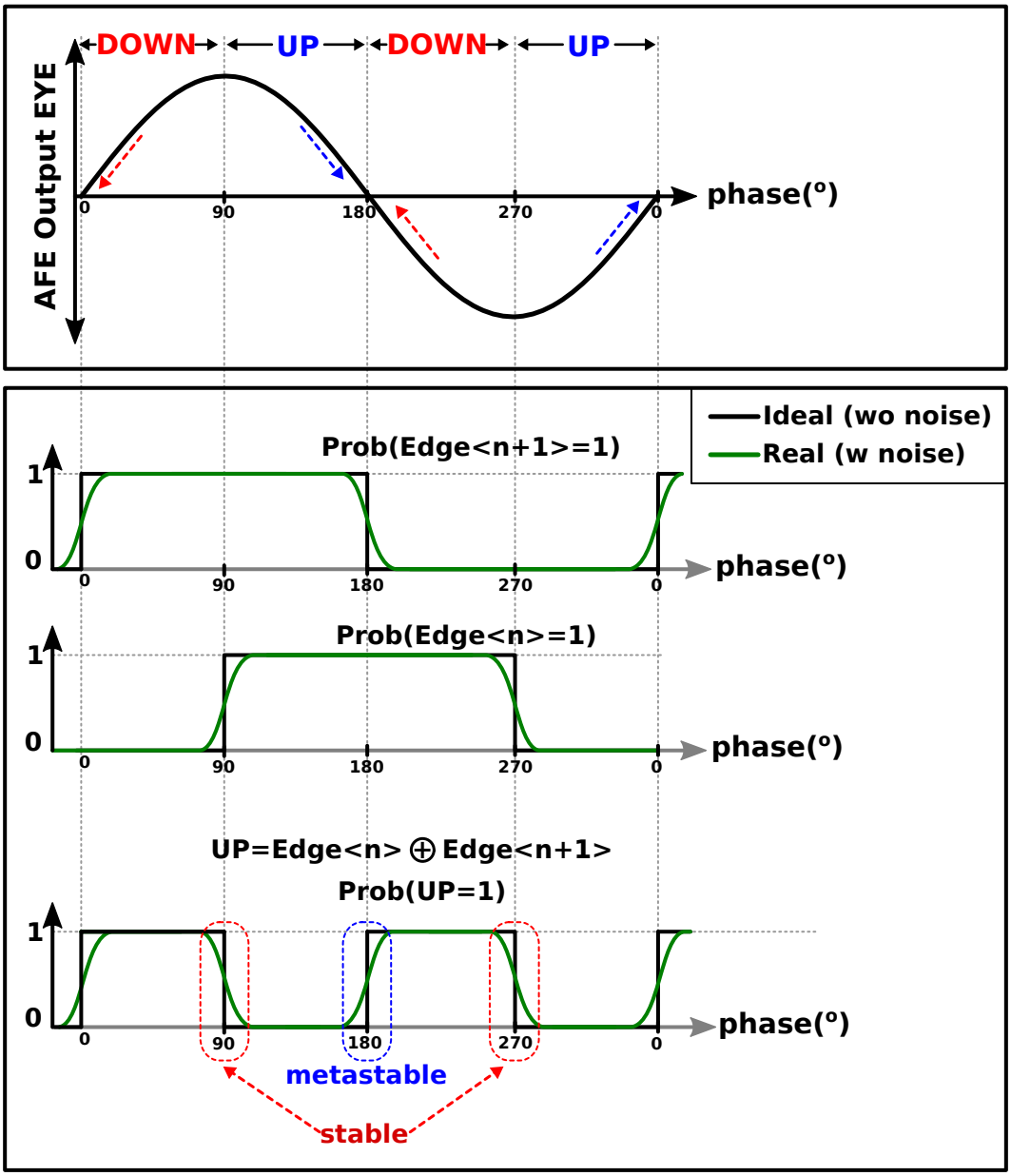

Figure 4.16: Probability of an $''UP''$ signal vs. the $Edge < n+1 >$ sampling phase

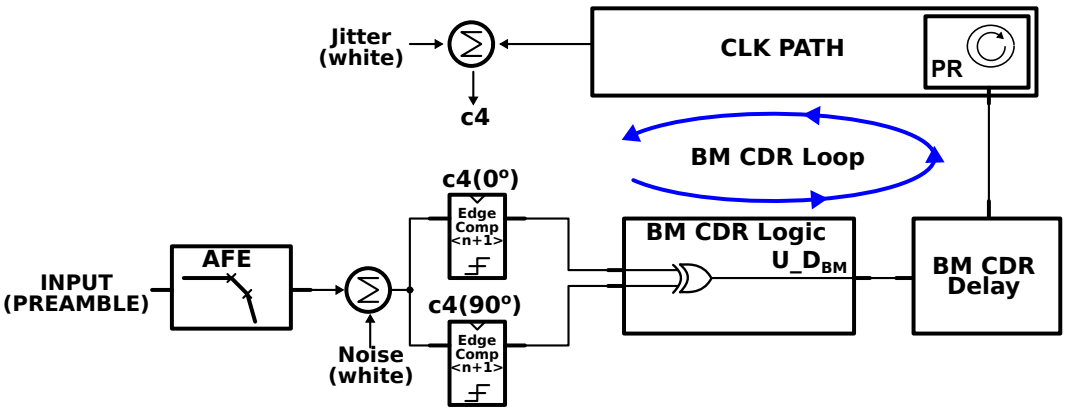

Figure 4.17: BM CDR block diagram used for modeling

steps). It must be noted that the term stable is not used in its strict definition as in the conventional loop stability analysis. Here it indicates that the loop wants to converge to this point. However, as can be seen from the figure, the PR position is not stable in the conventional sense and it oscillates around the stable point. The amplitude of the oscillation is a function of the BM CDR loop latency. Thus, the loop dynamics are well defined and once the loop latency is known (from transistor level simulation it is found to be approximately 32 UI or 8 c4 cycles at 60 Gb/s data rate) the overshoot or undershoot size is predictable and can be compensated for.

The same transient simulation was run for all possible PR initial positions with the latency parameter set to its nominal value and the results are given Fig. 4.21. The PR position diverges from the metastable points (0 and 64) and converges to the proximity of the stable points (32 and 96).

Up to this point, the random variables of noise and jitter were not included in the transient simulations, as a result the PR position is predictable and depends on the initial condition. In order to investigate the BM CDR dynamics during rapid power on, the noise and jitter parameters were set to their nominal values of 20 mV_{rms} (normalized) and 250 fs_{rms} and the transient simulation was rerun for 1000 times with random initial condition. The results are given in Fig. 4.22. For all the runs the PR position converges to approximately the same region around the stable points. Nonetheless, the convergence time for some of the runs has increased significantly. Tracing the runs that take longer time to converge to their origin in the plot, it can be concluded that they all start from metastable points of 0 and 64. In order to support that, the simulation is repeated 100 times for the initial PR positions of 5,60 and 64. The first two initial conditions are away from the metastable region whereas the third one is at the metastable initial condition. The results are given in Fig. 4.23. It is clear that all the non-metastable initial conditions start immediately to converge towards the stable point of $PR_{pos} = 32$ as

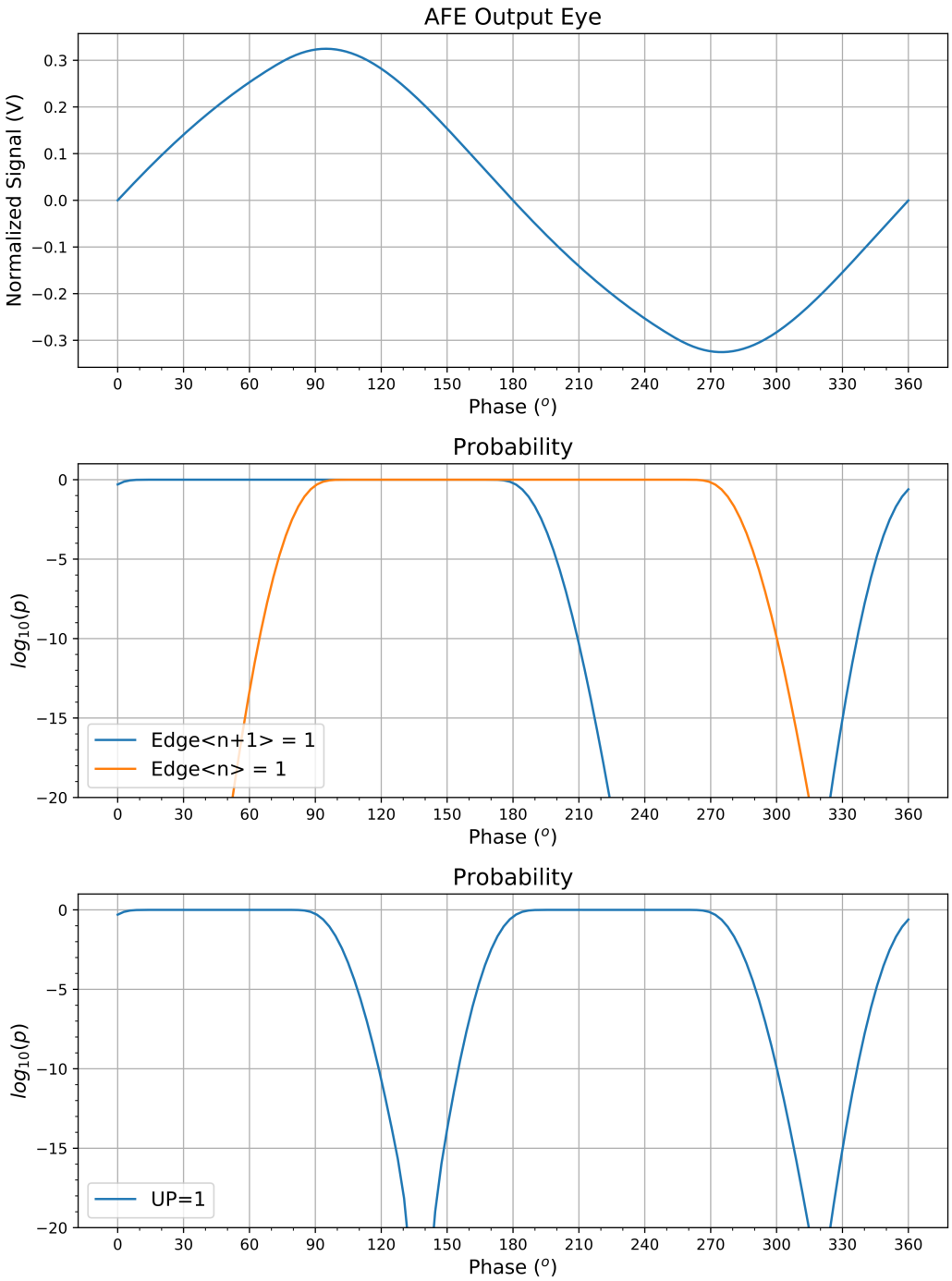

Figure 4.18: The probability of generating an $'UP'$ signal on BM-CDR logic with respect to the sampling phase on preamble signal

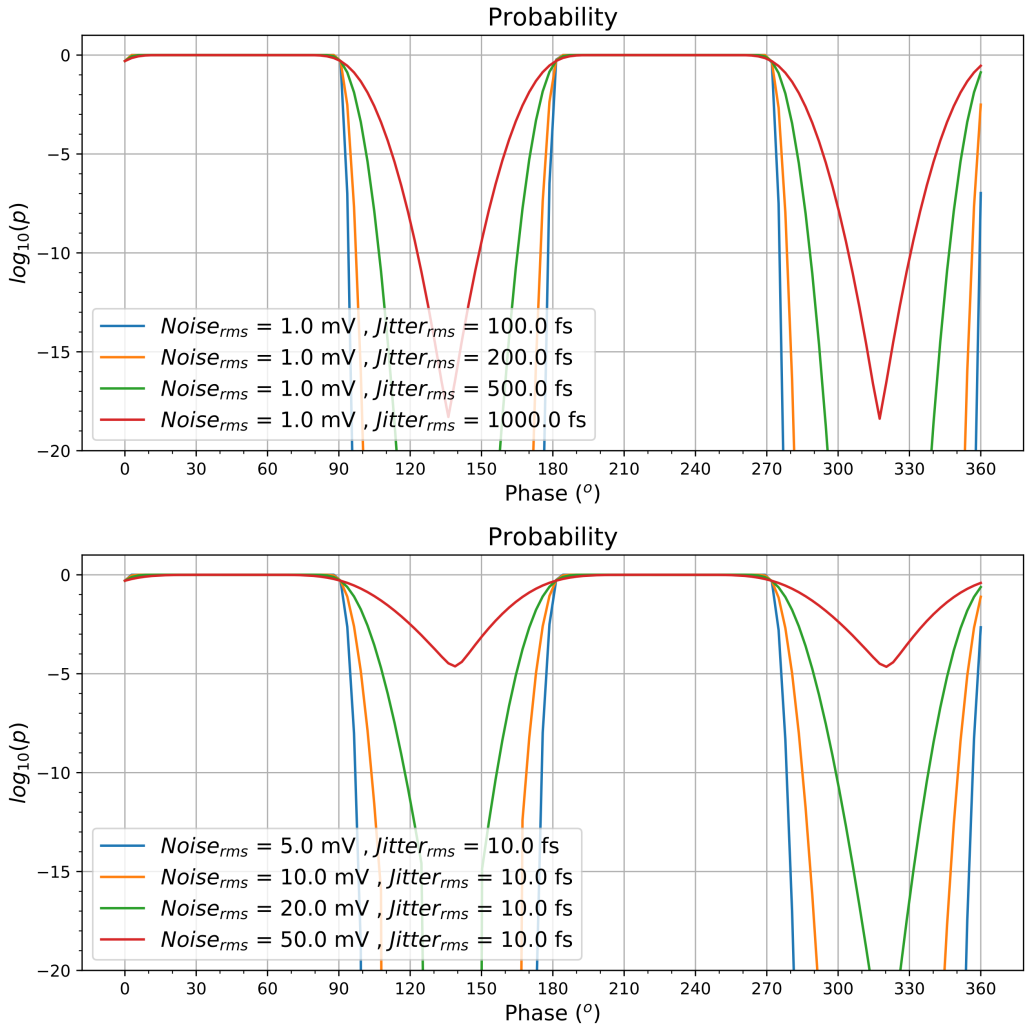

Figure 4.19: The effect of the noise and jitter parameter values on the probability of generating an $'UP'$ signal

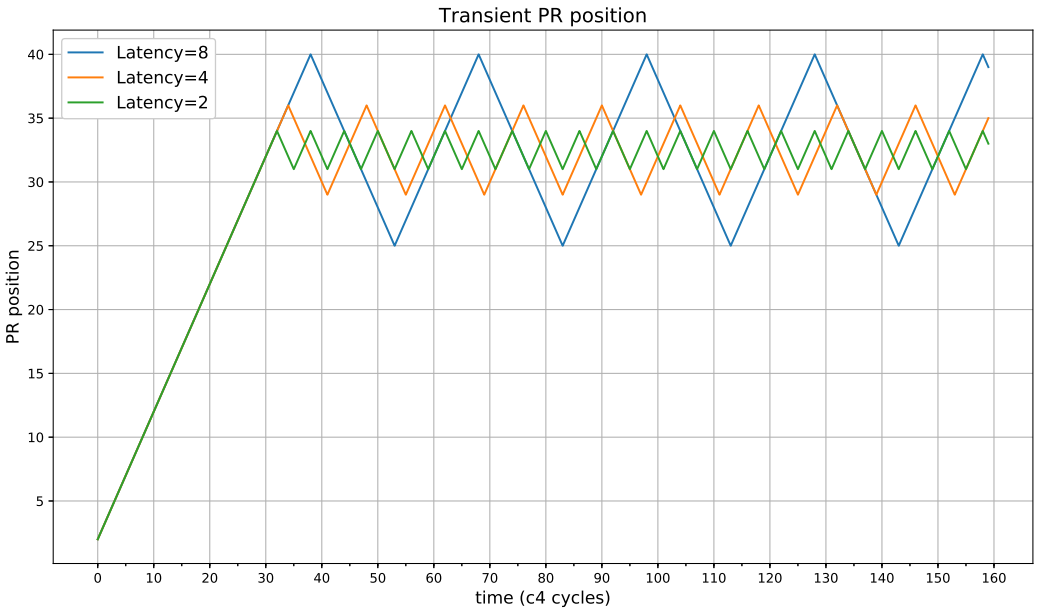

Figure 4.20: Transient PR position for BM CDR algorithm

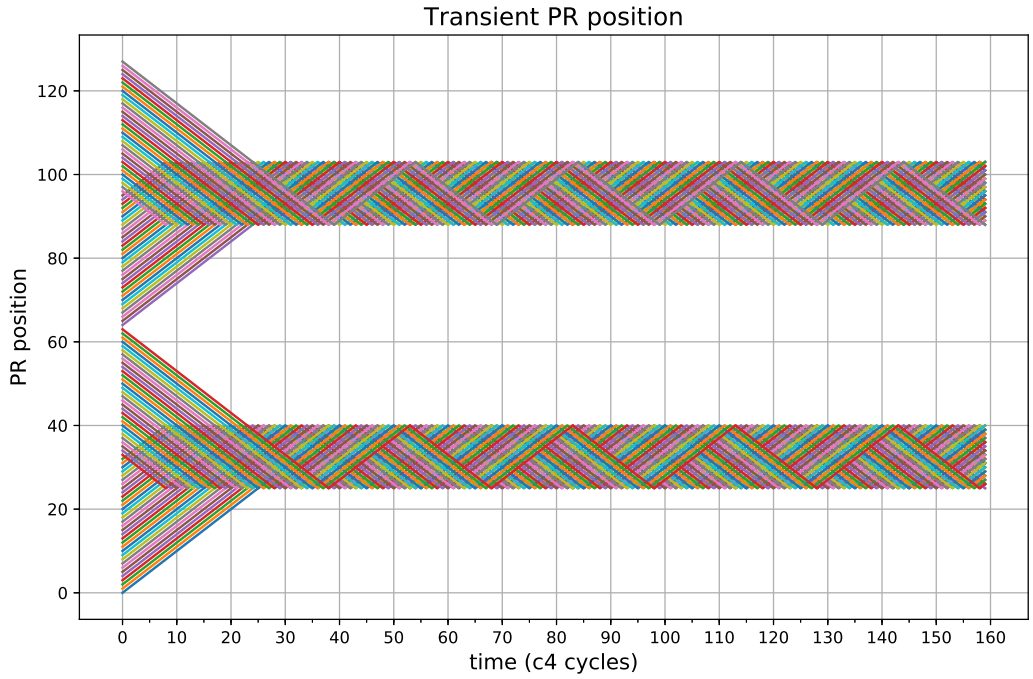

Figure 4.21: Transient PR position for BM CDR without noise and jitter

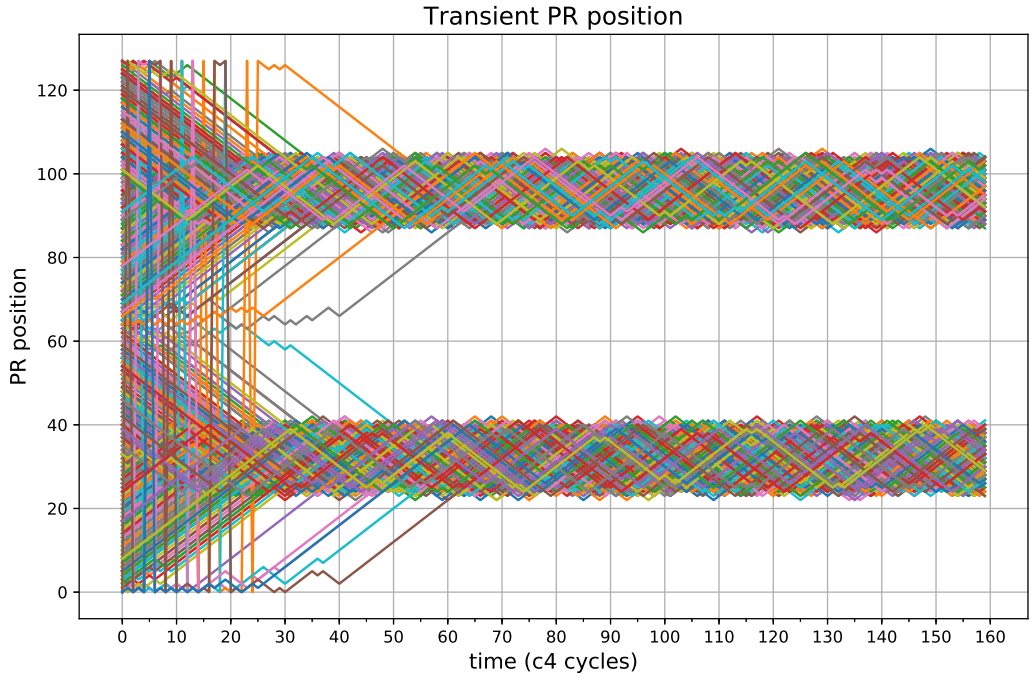

Figure 4.22: Transient PR position for BM CDR with noise and jitter

the first decisions are not affected by the noise and jitter. On the other hand, when the initial condition is metastable ($PR_{pos} = 64$ in this case) the first decisions are heavily affected by the noise and jitter, and the convergence time has already increased by 40 $c4$ cycles in 100 runs. To conclude, the metastable initial condition acts like a "trap" for the BM-CDR.

In I/O link measurements, the bit-error-rate (BER) test is conventionally run for 10^{12} bits. Assuming 1000-bit per power on cycle, the power-on test is expected to be measured for 10^9 times to get sufficient BER statistics. It is very unpractical to run and analyze the results of that many transient simulations in python environment. Thus, another BM-CDR model was developed (in Python) to directly find the probability of the PR being at each point after a certain number of steps starting from a certain initial condition. In other words, this model generates a probability density function for PR position. The model takes into account all of the parameters introduced for transient simulations: latency, noise, and jitter.

In Fig. 4.24 the probability density function of the PR after 20 updates ($'UP'$ or $'DOWN'$) is given for 3 different PR initial conditions. The figure clearly illustrates that when the PR starts from the metastable initial point (64) it has a high chance of being around that point after many steps: it is "trapped". On the other hand, starting slightly (4 points) away from the metastable point, the

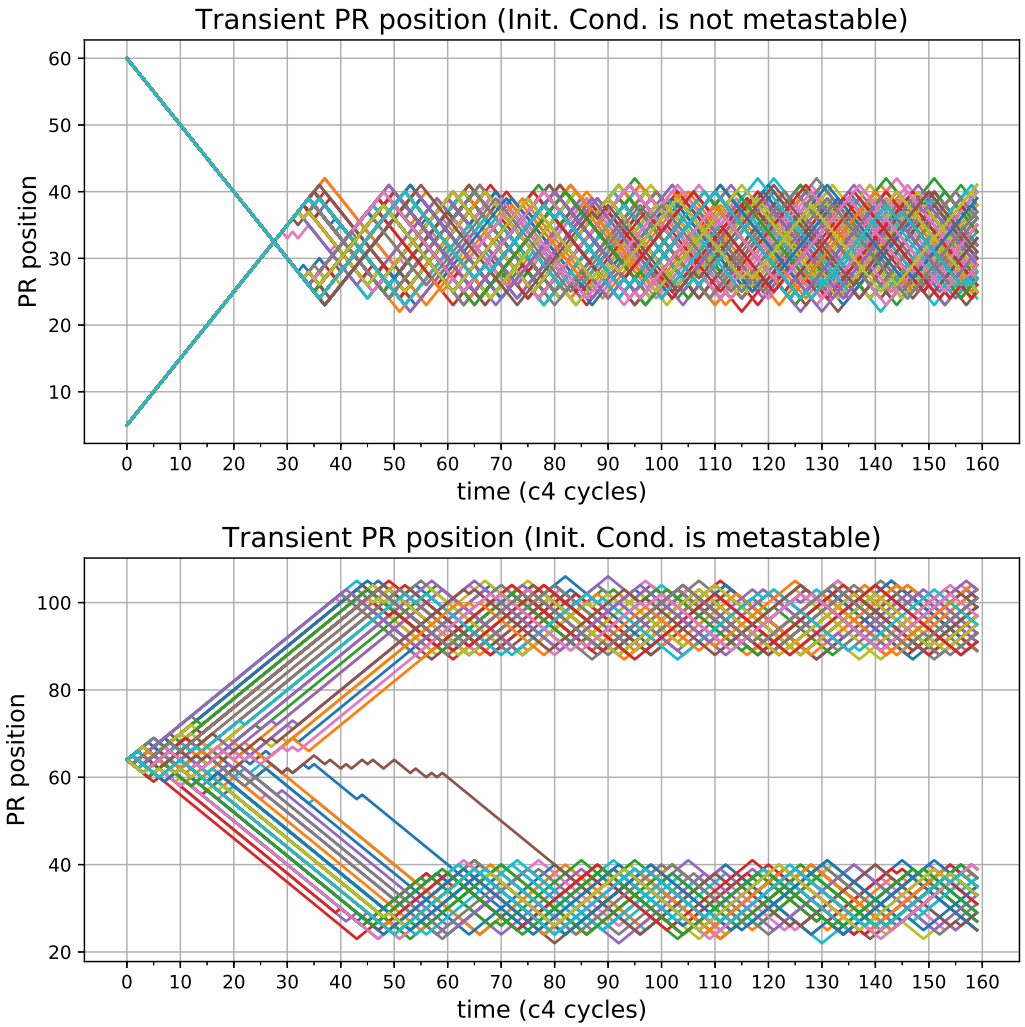

Figure 4.23: Transient PR position for BM CDR: comparison of non-metastable (top) and metastable (bottom) initial conditions

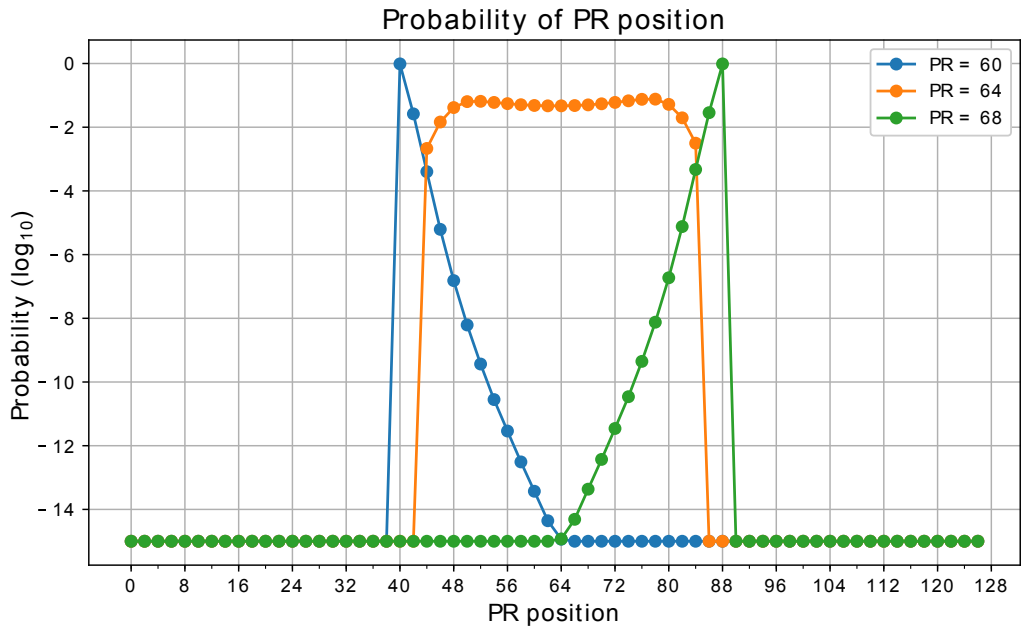

Figure 4.24: PR position probability after 20 decisions starting from initial conditions: 60, 64, 68

chances of being "trapped" already drops to below 10^{-15}.

In order to analyze the change of probability density function with respect to the number of decisions make, the simulation is run for 10, 20, and 30 steps. The results are given in Fig. 4.25. The probability of PR being around the metastable point drops with the number of steps as expected. However, the rate of this drop is very slow and it would require many more steps to go below a threshold of 10^{-9}.

Obviously, probability plots provided above are functions of noise and jitter. As the noise and jitter parameter values are reduced, the convergence will occur faster (higher chance of escaping the trap) and as they are increased the convergence will occur slower (lower chance of escaping the trap). Although it is not as obvious, the probability density function has also a strong dependency on the latency parameter. The reason is that, as the latency increases, the PR update direction becomes uncorrelated with the early or late information sampled by the edge comparators. The effect of latency on PR probability densith function for the metastable initial condition ($PR = 64$) is illustrated in Fig. 4.26. In the plot, the calculated probability is given at 20 steps after the initial condition, for 2 values of latency parameter: 2, 8 (nominal) and 14 $c4$ cycles.

To conclude, in its current state, the BM-CDR algorithm has a tendency to be trapped in one of the two metastable points depending on the initial condition. This problem increases the worst case lock time significantly, and needs to be

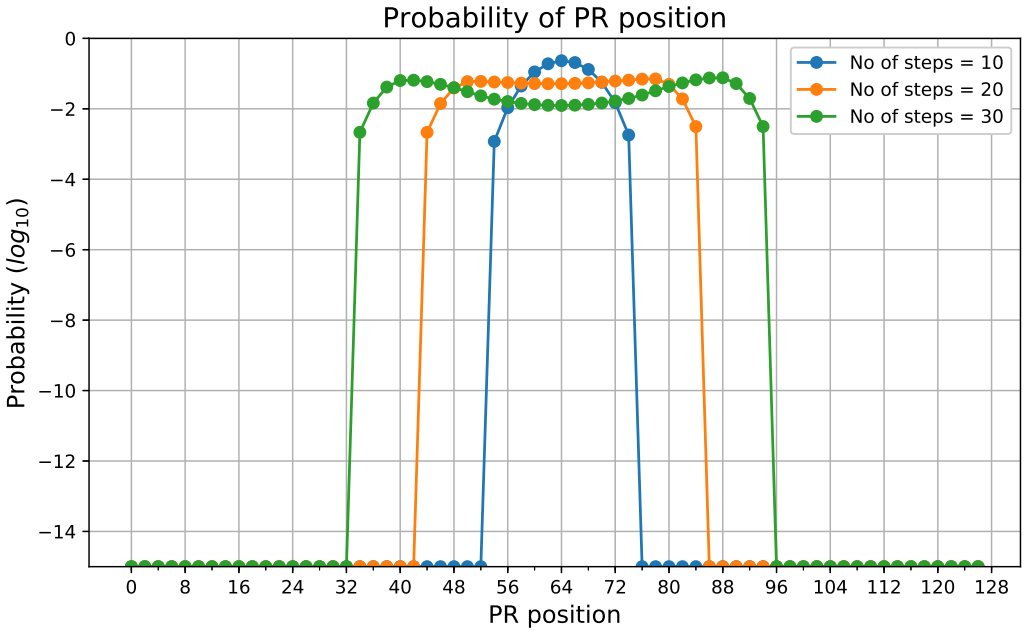

Figure 4.25: PR position probability after 10, 20, and 30 decisions starting from metastable initial condition

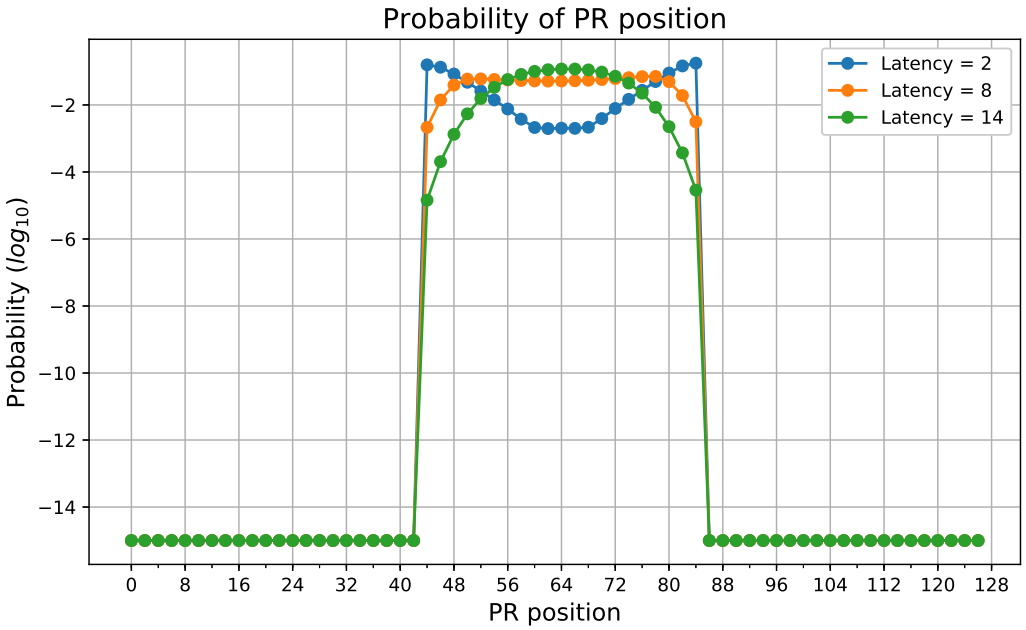

Figure 4.26: PR position probability after 20 steps starting from metastable point of 64 for latencies of 2,8, and 14 c4 cycles

tackled. In Section 4.4.3, the proposed solution to this problem will be presented and explained in detail.

4.4.3 Solution to Metastability Trap

The reason for the trap in the metastable initial condition is the unreliable edge information due to noise and jitter. For the same initial condition, in the absence of noise and jitter, the expected trajectory is a continuous ramp either up or down, as can be seen in Fig. 4.21. If this kind of behavior can be enforced artificially the metastability problem can be mitigated. Thus, the solution proposed in this work is to apply an initial ramp with a certain number of steps, whose direction is determined by the first edge information. After the ramp, the up/down updates are done as normal: 1 step/edge decision.

In order to determine the ramp length BM-CDR dynamics should be analyzed for stable and metastable initial conditions. On the one hand, for metastable initial condition, the longer the initial ramp the further the PR will move away from the trap. On the other hand, in stable initial condition the longer the initial ramp the further the PR will move away from the stable point. Actually, a careful analysis reveals that any ramp whose length is bigger than 16 increases the worst case BM-CDR lock time. Ideally, the worst case initial phase search should take 32-steps (from metastable to stable positions). If the initial ramp is longer than 16, then the stable position becomes the worst case: the PR will travel away from the stable position as long as the ramp and then travel back to the stable position. Thus, the ramp length should not be longer than 16. Moreover, there is also no motivation to make it less than 16, as the probability of ending up near the metastability trap increases and the worst case initial phase search is limited to 32 steps between the metastable and stable positions anyway. As a result, the ramp size is fixed as 16.

4.4.4 BM-CDR Algorithm

This section describes the implemented BM-CDR algorithm based on the aforementioned considerations in this chapter. The BM-CDR algorithm consists of 4 different phases ($P0-3$) and is executed by the finite state machine PON_FSM (Fig. 4.8). In order to minimize the locking time, the PON_FSM block runs at the quarter rate clock and is realized using custom digital logic.

During idle mode, PON_FSM is reset to its initial condition by the EN signal provided from the PON_SENSE block. When the reset is released ($EN = 1$), first it counts up to 16 to allow the analog bias nodes and AFE offset cancellation loop to settle to their nominal values, comprising phase 0 ($P0$). After that, it applies

the ramp of 16 steps to avoid the metastability trap as explained in Section 4.4.3, consisting phase 1 ($P1$). When the ramp is finished, it connects the BM-CDR logic output, which was described in Section 4.4.1, to PR control register and starts checking for a transition in the 'UP' signal. This period constitutes phase 2 ($P2$). The transition in the 'UP' signal indicates the crossing of the stable PR position. Once the change is detected, the final phase ($P3$) of the BM-CDR takes place where the PR is pulled 8 steps backwards to compensate for the overshoot or undershoot caused by the loop latency, which was illustrated in Fig. 4.21. At the end, the PR control is handed over to the BB-CDR loop for normal operation.

The different phases of the BM-CDR algorithm are marked on a transient simulation of a complete power-on cycle together with the AFE output and power on control signals in Fig. 4.27. The transient simulation was run using the RC-extracted model of the complete adaptive RX, except the RX digital block, given in Fig. 4.8. Moreover, the noise was also taken into account, which explains the fluctuations in the preamble signal seen at the VGA output. The PR position being stable after the PR control is handed over to BB-CDR indicates that the BM-CDR found the correct sampling position during the rapid phase search.

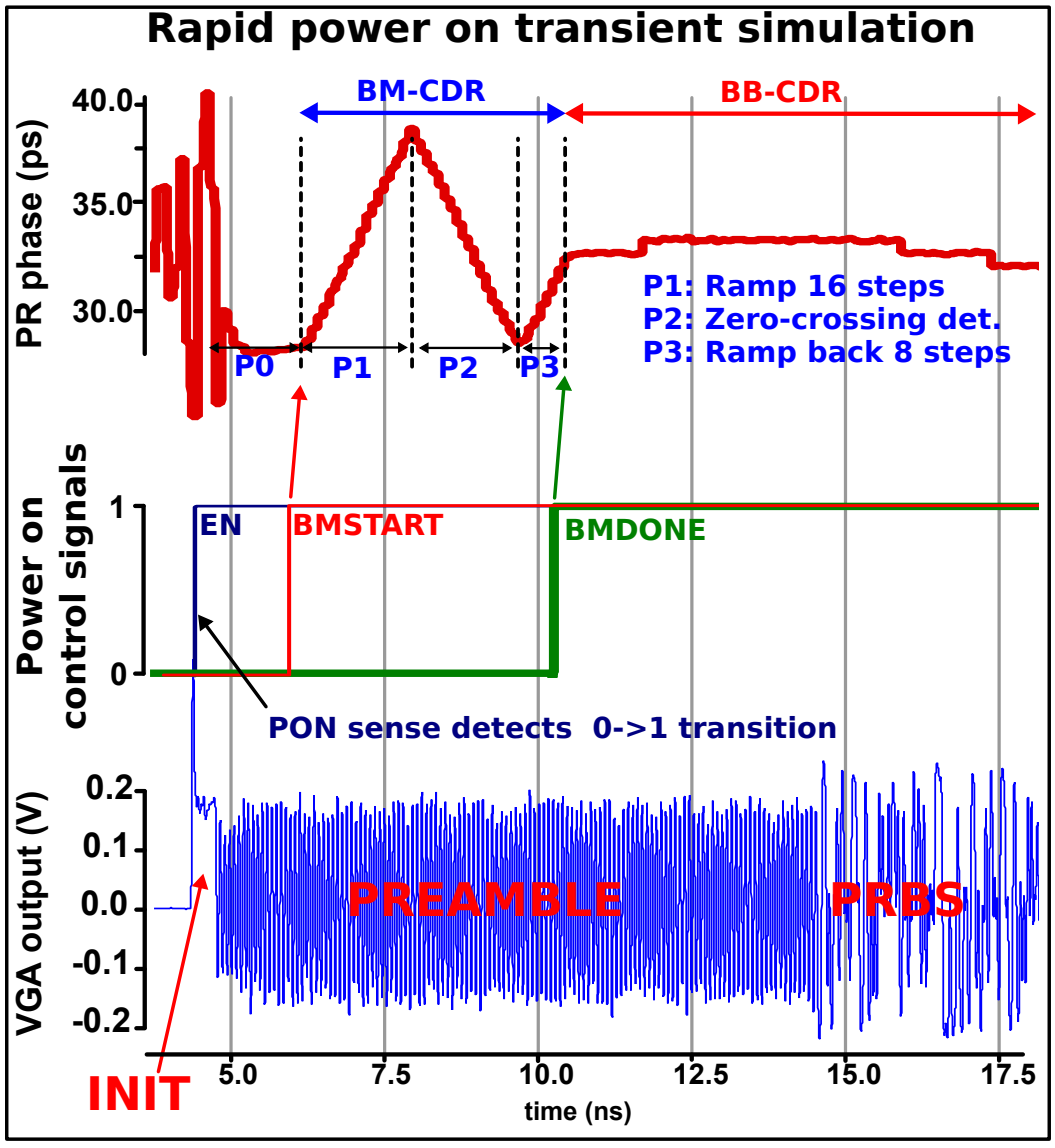

Figure 4.27: Transient simulation of a complete power-on cycle

Chapter 5

Measurement Results

Following the fabrication schedule, 3 main versions of the RX were fabricated. The first version included the data path and a simple clock path without CDR functionality. The purpose of this chip was to investigate the speed and sensitivity performance of the data path. In the second version, CDR was added to measure the jitter tolerance and frequency offset tracking performance. Finally, the last version was modified to include adaptivity function.

The RXs, implemented in 14nm bulk finFET technology, were wire-bonded to custom PCBs for testing. The surface illuminated GaAs PIN diode used to convert the incoming optical signal into current has a diameter of 16 μm, parasitic capacitance of 69 fF, and responsivity of 0.52 A/W. The bandwidth of the PD is around 25 GHz ([54]).

The measurements were conducted on the setup given in Fig. 5.1. A 56 Gb/s bit pattern generator overclocked up to 64 Gb/s drives a SiGe driver with 2-tap FFE (1 precursor and 1 main cursor with a ratio of around 0.45). The SiGe chip modulates a high speed 850 nm VCSEL which is connected to a 7 m OM2 multimode fiber. The performance characteristics of the VCSEL and SiGe driver can be found in [16]. An optical attenuator is connected in the signal path for sensitivity measurements. The optical modulation amplitude (OMA) used to report sensitivity performance is calculated as follows:

$$OMA = 2\frac{Av_{cur}}{Res}\frac{ER-1}{ER+1} \tag{5.1}$$

where ER is the extinction ratio (measured: $ER = 1.8$), Res is the responsivity of the PD ($Res = 0.52$), and Av_{cur} is the average current of the PD.

All 3 chips have a an integrated correlator that can perform BER check and more complicated tasks such as extracting the speculative eye. Measurement results stored on on-chip registers are then transmitted off-chip via a three wire serial interface. The correlator is digitally synthesized and runs at 1/32 of the data rate (2 GHz at 64 Gb/s data rate).

The remainder of this chapter is organized as follows: Section 5.1 provides the measurement results of the first version of the RX and the focus is data-path performance; Section 5.2 provides the measurement results of the second

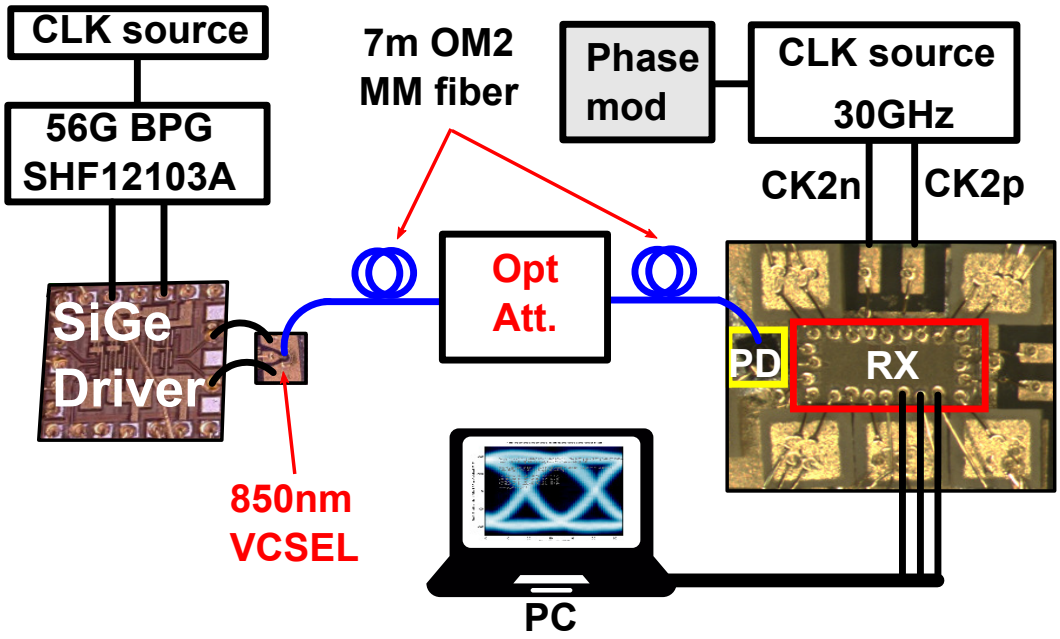

Figure 5.1: Measurement setup.

RX version and the focus is CDR performance; and finally Section 5.3 provides the measurement results of the final version and the focus is rapid power-on performance.

5.1 Data Path Measurement Results

This section provides the measurement results of the data path of the RX. The measurements were performed on the first version of the RX whose block diagram is given in Fig. 5.2. It has a simplified clock path consisting of a CML buffer, a CML-to-CMOS converter and an injection locked oscillator based CMOS IQ generator. The reference clock was at a quarter of the data rate. Also, this RX version does not include the offset cancellation loop in the analog front end. The DC current of the PD was canceled by fine-tuning the 12 bit current DAC. The chip micrograph and layout of the active area, in which the building blocks are marked, are given in Fig. 5.3. The active area of the RX is 150 μm by 190 μm.

The pulse response of the link at the output of AFE is measured by applying a repeated 8-bit pattern of '10 000 000' to the optical TX. This pulse response includes all the bandwidth limitations of VCSEL, PD, and AFE of the RX as well as the FFE at the TX. The correlator is configured to count the ratio of '1's to the total number of bits received from a single comparator. The measurement is repeated by changing the threshold of the comparator via VDAC and the ratios of '1's are recorded for each point. When combined together, the recorded points

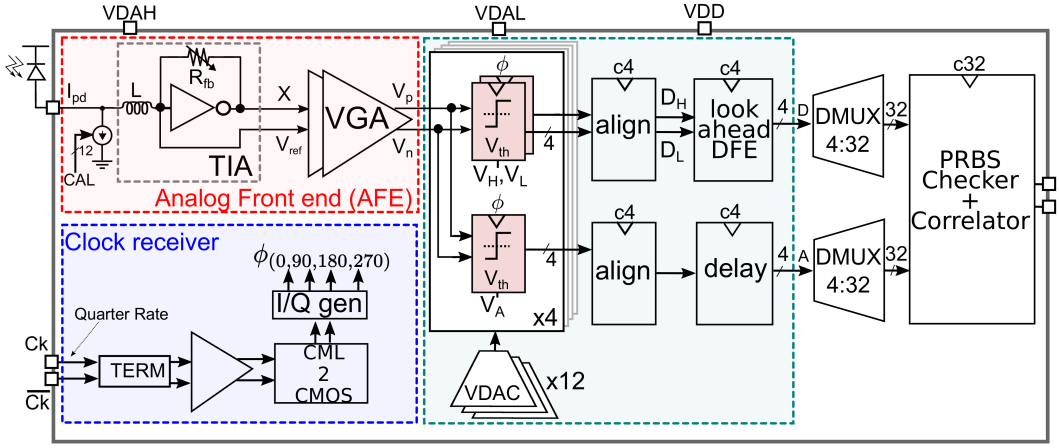

Figure 5.2: Block diagram of the 1^{st} RX chip

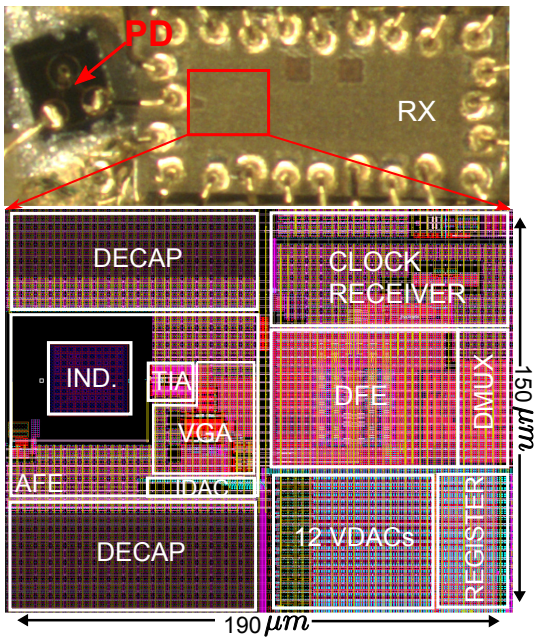

Figure 5.3: Layout and chip micrograph of the 1^{st} RX chip

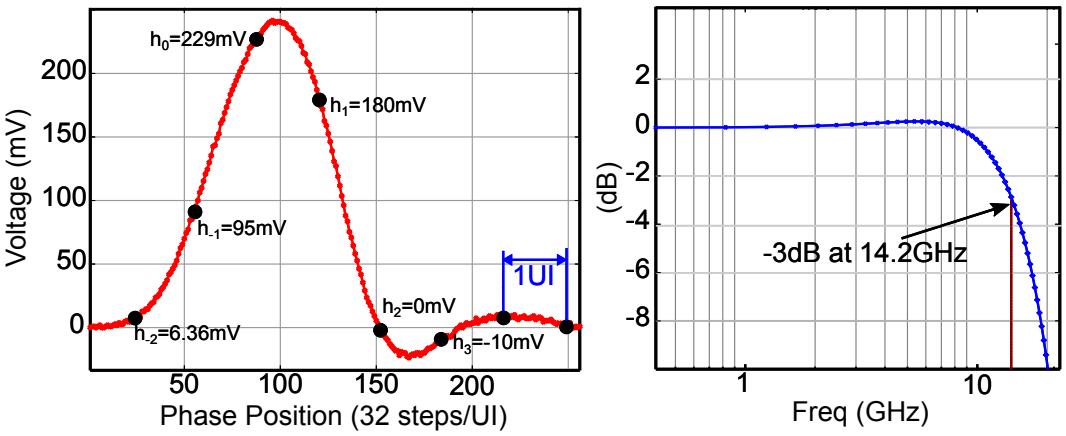

Figure 5.4: Transient pulse response at 64 Gb/s and the corresponding frequency response

gives a cumulative distribution function (cdf) with respect to the threshold voltage at a certain phase. Taking the derivative of the cdf will give a probability distribution function (pdf), whose mean value corresponds to the signal level and whose standard deviation is the rms noise at that point. The pdf measurement is repeated by stepping the phase of the RX external clock source. Fig. 5.4 shows the single bit response of the link at 64 Gb/s and the corresponding transfer function. The 3-dB bandwidth matches very well with the optimum value shown in Fig. 3.9 for 1-tap DFE. It must be noted that this pulse response is the combined response of all elements in the data-path, including SiGe driver, VCSEL, PD, and RX-AFE.

The eye diagrams are found by applying the same procedure described above with a PRBS-7 input. The results are given in Fig. 5.5. At 36 Gb/s the signal is not affected by ISI and the RX operates error free without DFE (Fig. 5.5a). As the data-rate increases, ISI starts to close the unequalized eye but the DFE speculative eyes are still open up to 64 Gb/s (Fig. 5.5c and Fig. 5.5d). The measured eye diagram matches very well to the *RC* extracted simulation as illustrated in Fig. 5.5b.

Figure 5.6 shows BER contour plot at BER 10^{-9} for 60 Gb/s data-rate. This measurement is done by sweeping the two variables clock phase and DFE slicing level and measuring BER at each point. The plot clearly shows that without DFE (DFE value is set to 0) the BER is around 10^{-2}. And the optimal eye opening is achieved with a DFE value of 200 mV, which is very close to the h_1 value found in the pulse response given in the Fig. 5.4.

Figure 5.7a and Fig. 5.7b show BER bathtub curves measured with a PRBS-7 sequence at 56 Gb/s and 64 Gb/s, respectively at -5 and -8dBm OMA. The sensitivity of the RX, which is defined as the minimum optical power which

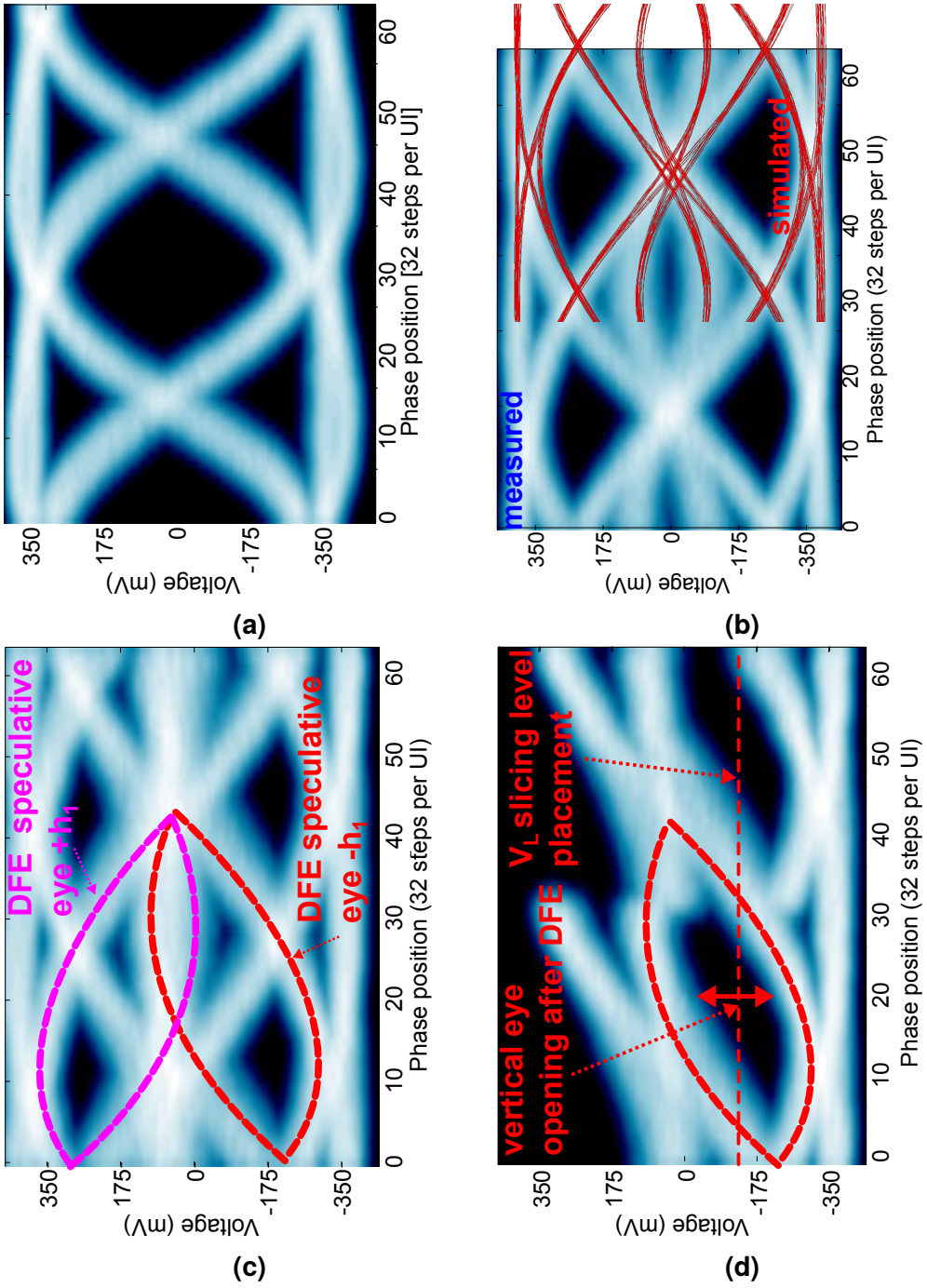

Figure 5.5: Eye Diagrams (a)36 Gb/s (b)56 Gb/s (c)64 Gb/s (d)DFE equivalent eye @64 Gb/s.

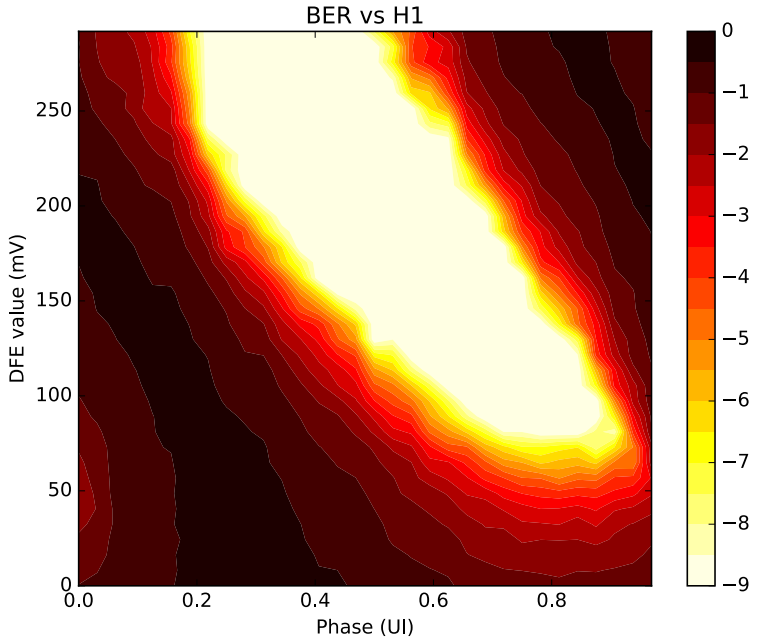

Figure 5.6: BER vs. DFE slicing level and phase contour plot @60 Gb/s.

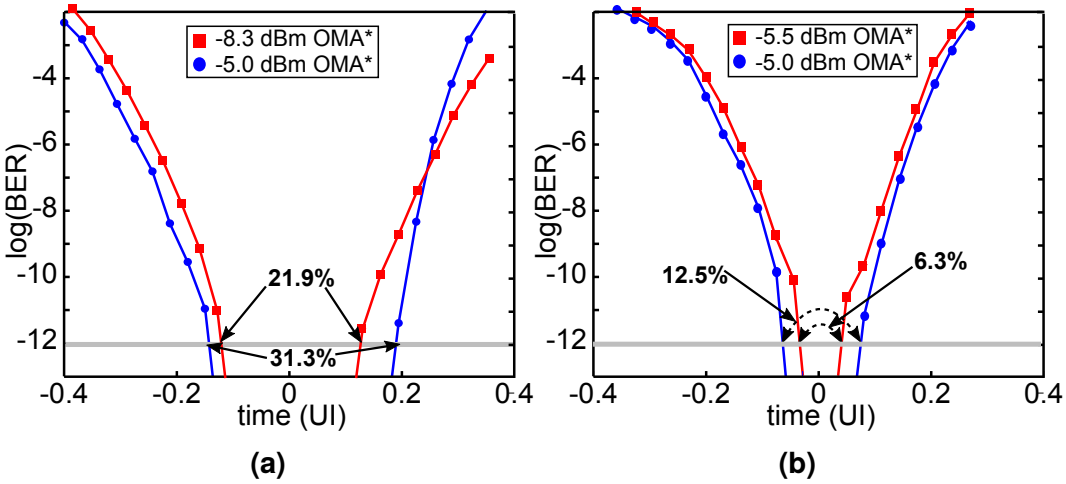

Figure 5.7: Bathtub curves (a)56 Gb/s (b)64 Gb/s.

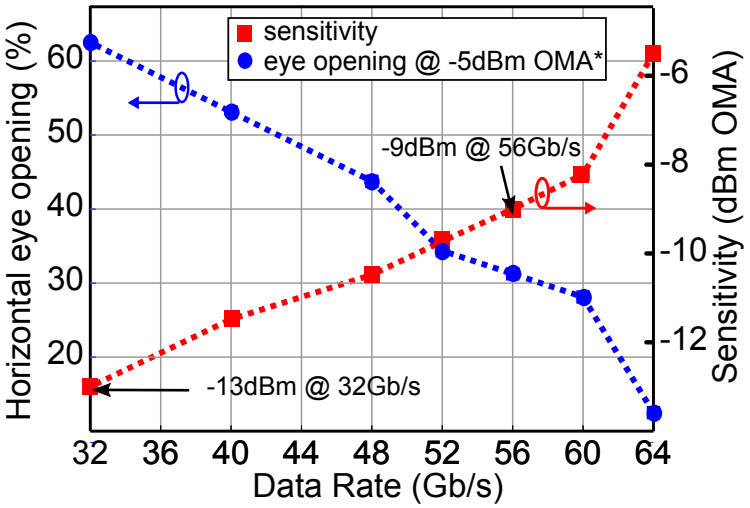

Figure 5.8: Sensitivity and eye opening at -5 dBm OMA vs data-rate.

satisfies a BER of $< 10^{-12}$ at least in one point on the bathtub curve, is found for the data-rates between 32 and 64 Gb/s as presented in Fig. 5.8 together with the eye opening at -5 dBm OMA. The sensitivity at 32 Gb/s, 56 Gb/s, and 64 Gb/s are -13 dBm, -9 dBm, and -5.5 dBm, respectively.

5.2 CDR Measurement Results

This section provides the measurement results of the second version of the RX. The block diagram of this version is given in Fig. 5.9. It has a fully functional CDR and receives a half rate reference clock. The chip micrograph and layout of the active area, in which the building blocks are marked, are given in Fig. 5.10. The active are of the Rx is 125 μm by 150 μm.

The bathtub and BER contour plots obtained at 60 Gb/s is given in Fig. 5.11. The bathtub is 28% open and the maximum eye opening is achieved with an $h(1)$ coefficient of 175 mV at the speculative slicer inputs.

The jitter tolerance curves (for BER $> 10^{-12}$) at 60 Gb/s and 30 Gb/s data rates are given in Fig. 5.12 together with the frequency tracking range measurement of the RX at 60 Gb/s. The jitter tolerance corner frequency is around 80 MHz at 60 Gb/s which is quite close to the theoretical value calculated in Section 3.3.1. Also the frequency tracking range is close to the theoretical value of 780 ppm, which comes from the 1 step update limitation due to PR architecture and the maximum output rate of the loop filter, which is 1 update every 10 c4 cycles.

The maximum error-free operation speed dropped from 64 Gb/s in the first chip down to 60 Gb/s in the second chip. This is an expected reduction since

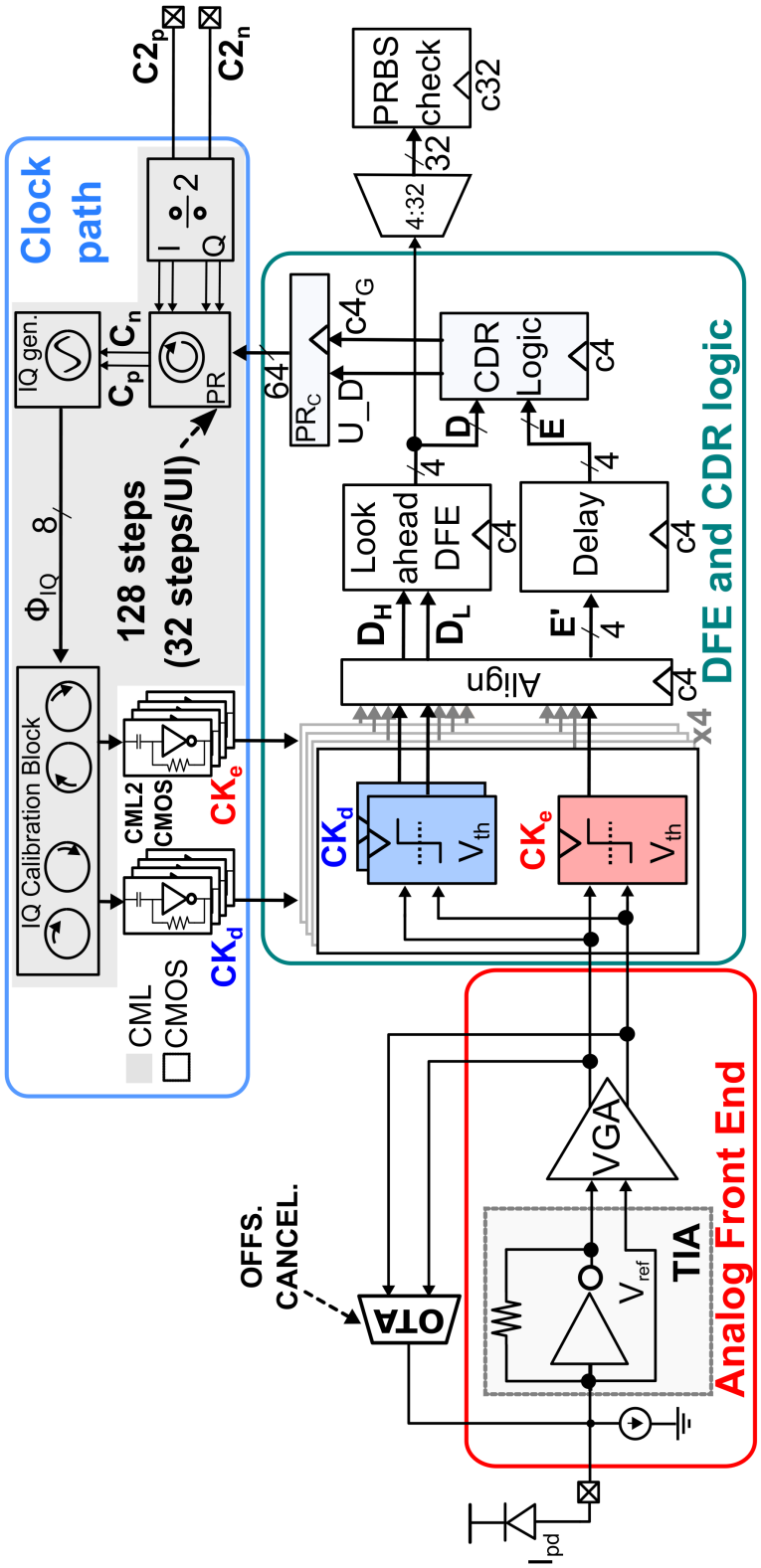

Figure 5.9: RX top level block diagram (2^{nd} version, with CDR)

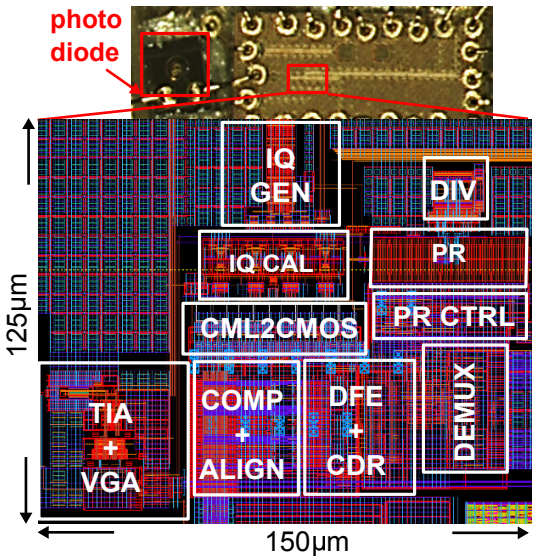

Figure 5.10: Layout and chip micrograph of the 2^{nd} RX chip

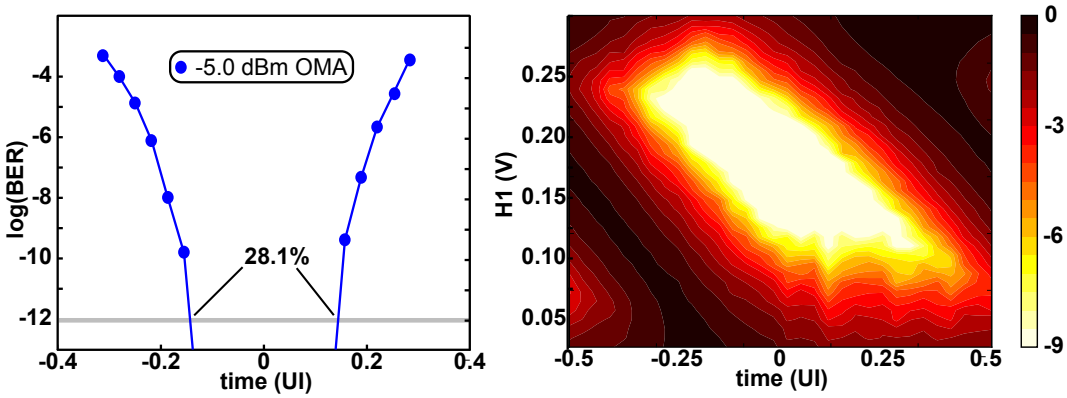

Figure 5.11: Bathtub and BER contour plots at 60 Gb/s and -5 dBm OMA

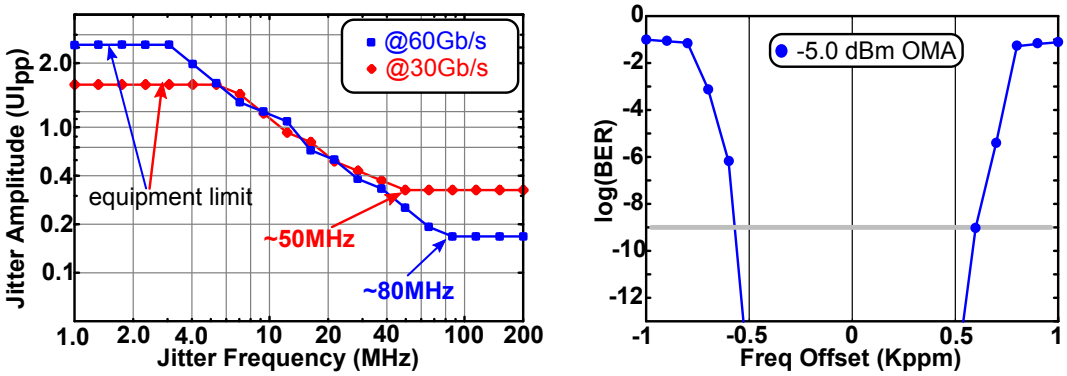

Figure 5.12: Jitter tolerance and frequency tracking range of the RX

all the additional blocks on the clock path such as the CML divider, the phase rotator, and the I/Q corrector contribute to jitter be it random or supply ripple induced.

5.3 Adaptive RX Measurements

The third RX version was modified to include the adaptive functionality, as shown in Fig. 4.8. The chip micrograph and layout of the active area, in which the building blocks are marked, are given in Fig. 5.13. The active are of the Rx is 200 μm by 300 μm including the RX digital, whose size is significantly affected by the measurement related functionality.

The RX digital block in this version was custom designed to help rapid power-on and off algorithm by sensing the beginning of an idle period and asserting PWROFF signal. Moreover it was modified such that it can run BER checks across multiple power-on cycles, accumulating all the errors.

In order to observe the power-on functionality of the adaptive RX, the 1/8 clock signal of the RX was transmitted off chip. This signal is observed on a real time oscilloscope together with the inverted TX output signal. The real time oscilloscope options were set to high speed acquisition and infinite persistence modes. At the end of an hour a screen-shot of the oscilloscope was saved, which is given in Fig. 5.14. Different parts of the data protocol are marked on the TX output data. This plot does not provide any information about the BER however it clearly indicates a correct power cycling which is based on 2 observations: First, the C8 clock is turned off for all idle periods. Second, the C8 clock is turned on for all data burst periods. If that was not the case, due to infinite persistence setting of the oscilloscope, there would have been overlaps of active and idle regions on C8 output. The reason why C8 seems to come earlier than the INIT sequence is the difference between the delay paths of the two signals on the measurement setup.

In order to measure the precise power-on time, data bursts with different preamble lengths were applied and BER was measured by the RX digital block. The OMA was kept at -4 dBm for rapid power-on measurements.

The RX digital block detects the start sequence (STR) and aligns the data with a barrel-shifter such that the PRBS register is seeded with the first 32 PRBS bits right after the STR and the PRBS check is completed with this seed for the whole data burst the seed belongs to. This ensures that even the first PRBS bit is included in the BER check. The error count is accumulated across multiple power cycles. Fig. 5.15 shows the BER versus preamble length in terms of UI for 10^{10} power on cycles at 56 Gb/s and 60 Gb/s data rates. The corresponding

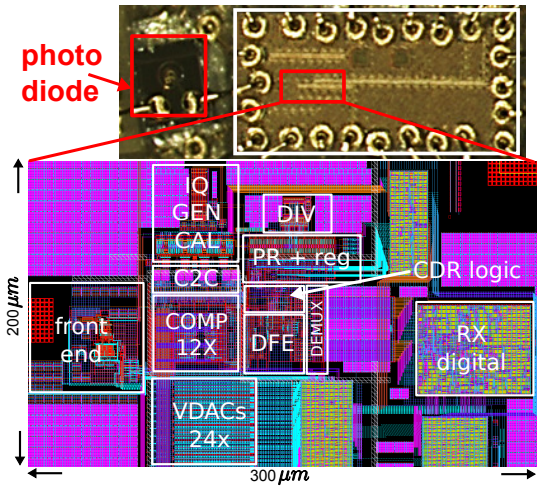

Figure 5.13: Layout and chip micrograph of the 3^{rd} RX chip (adaptive RX)

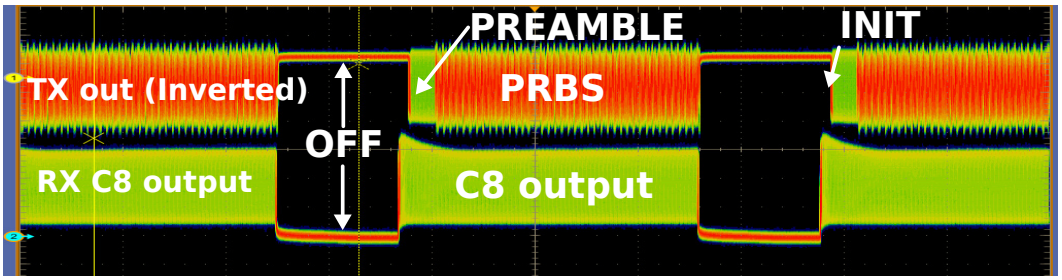

Figure 5.14: Inverted TX output signal (top) together with the C8 clock of the RX

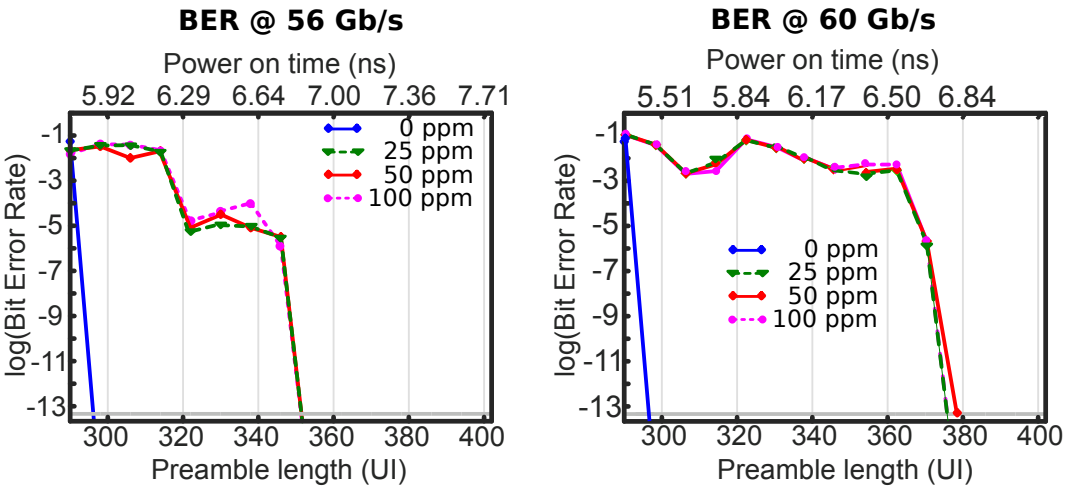

Figure 5.15: Power-on time measurement results for 56 and 60 Gb/s

power-on time, including the initialization (INIT, 32 UI) and start (STR, 4 UI) sequences, is given on the top x-axis for both date-rates in terms of ns.

The power-on time for the implemented RX is measured to be less than 7 ns for both 56 Gb/s and 60 Gb/s data-rates although the required preamble length is approximately 30 UI longer at 60 Gb/s. This performance also covers a frequency offset of up to 100 ppm. For source synchronous systems (0 ppm frequency offset) the power-on time is less than 6 ns.

The RX is equipped with a snapshot register, which is activated automatically at the power-up to store the first PR control bit history for 70 early or late decisions starting synchronously with the ramp phase (P1) of the BM-CDR algorithm. The recorded PR transient during rapid-lock at 56Gb/s is shown in Fig. 5.16 demonstrating correct CDR operation. The 3 phases (P1-3) of BM-CDR algorithm are marked on the PR position figure. At the end of the BM-CDR the PR position is only 4 steps away from the final locking point around 14. The recorded PR position is in good agreement with the simulated PR position given in Fig. 4.27.

It must be noted that the x-axis is not a linear time scale: during the burst mode the PR is clocked every c4 cycle, however during the normal operation the PR clock is gated by the loop filter and the maximum output rate is once every 10 c4 cycles, which can be much lower than this value depending on the noise parameters. Thus, the BB-CDR portion of the PR position is actually much longer in real time compared to BM-CDR portion.

Measured on-power is 127 mW at 56 Gb/s, which corresponds to an energy efficiency of 2.2 pJ/b, and off-power is 8 mW, which mostly consists of TIA and bias generation circuitry. The power breakdown is given in Table 5.1.

Both measured and estimated power consumption vs. utilization ratio are

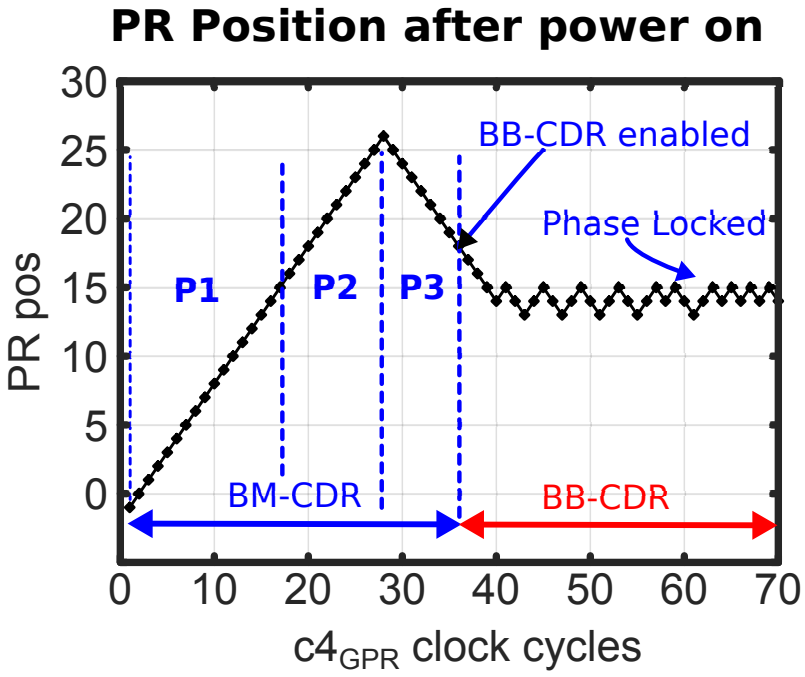

Figure 5.16: PR position recorded on snapshot register

Table 5.1: Power breakdown of the RX at 56 Gb/s

		VDAH = 1V	VDAL=0.9 V	VDD=0.9 V	
State	Power (mW)	AFE + Clk Path	Slicers + Aligner + CMOS Clk	DFE + CDR Logic + DMUX	RX Digital
on	126	59	38	19	10
off	8	7	0.5	0.3	0.2

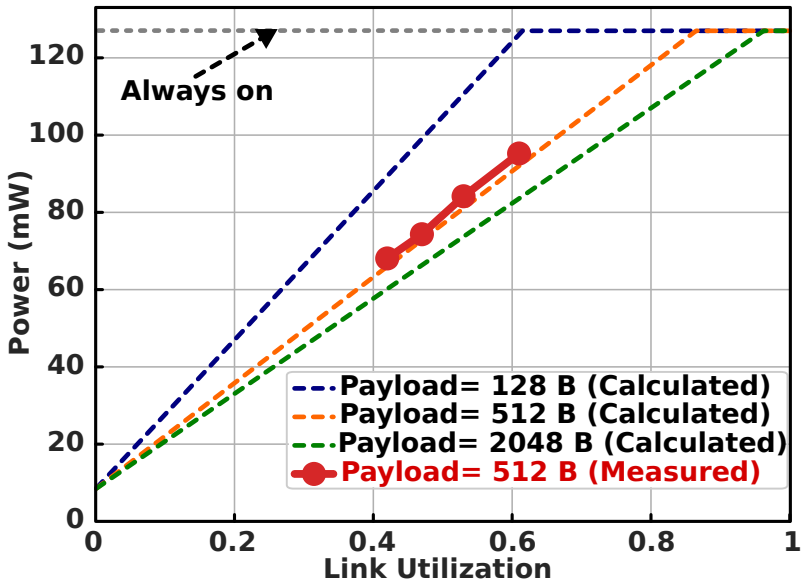

Figure 5.17: Estimated and measured power consumption vs. utilization ratio @ 56 Gb/s

given in Fig. 5.17. The estimation is based on Eq. (4.5). The parameter values P_{ON}, P_{OFF}, and T_{sync} are taken from the measurement results. The measurement results are in good agreement with the estimated power consumption.

The reason why lower utilization ratios could not be measured was the AC coupling between the pattern generator and the VCSEL driver on the transmitter. As the utilization ratio drops, due to AC coupling, the common mode of the effective signal at the VCSEL driver input moves away from the actual common, resulting in significant signal degradation at the optical output.

By modification of bias currents and supply voltages, an almost constant energy efficiency was measured down to 10 Gb/s data-rate for always on state RX. The power consumption and the supply voltages used for the measured data-rates are given in Fig. 5.18. The results indicate that the implemented RX can be integrated into many applications without an energy-efficiency penalty if the supply voltage can be adjusted as shown.

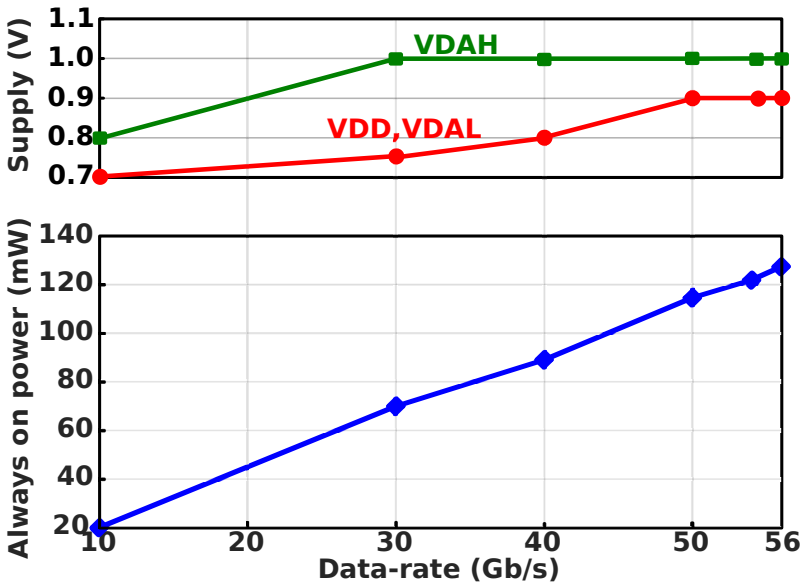

Figure 5.18: Always on power and supply voltages vs data-rate

Chapter 6

Conclusion

This thesis presents an adaptive optical receiver for high-speed optical communication that powers itself up or down depending on the bandwidth demand on the link. The adaptive receiver is fully autonomous and does not require any external enable or disable signal other than the incoming data itself. The link protocol introduced to perform the rapid power-on algorithm is very simple: it is just an additional piece of information attached at the beginning of each data packet. The protocol does not require any calculation or feedback loop at the transmitting side that may complicate the data generation. The total power-on time is measured to be 7 ns, which includes the settling of the bias voltages of the analog stages, common-mode cancellation of the photo-diode current, and CDR lock. The power-on performance of this study is compared to [15], [55], [56], and [29] in Table 6.1. To the best of authors knowledge, this is by far (almost 10X data rate) the fastest receiver with rapid power-on functionality, up to date.

The presented receiver, not only proves the adaptive receiver concept for high speed optical communication but also achieves state of the art performance in terms of data-rate, sensitivity, and jitter tolerance.

In Fig. 6.1, the data-rate and energy efficiency of the previously published high-speed optical receivers are compared with this work. At maximum speed, this work stands alone in the high-speed and low-power corner. The maximum data-rate of the presented optical receiver is almost double the fastest CMOS

Table 6.1: Power-on/off comparison with prior art

	[15]	[55]	[56]	[29]	This W.
Technology	65 nm	40 nm	40 nm	32 nm	14 nm
Data-rate (Gb/s)	7	5.6	4.3	25	60
Link Type	Elec.	Elec.	Elec.	Opt.	Opt.
Lock time	< 20 ns	8 ns	242 ns	18.5 ns	6.8 ns
Incl. Power-on/off	Yes	Yes	Yes	No	Yes
Enr. Eff. (pJ/bit)	9.1	2.4	3.3	4.4	2.2
On-State Pow. (mW)	63.7	13.4	14.2	110	126
Off-State Pow. (mW)	0.74	0	0.05	NA	8

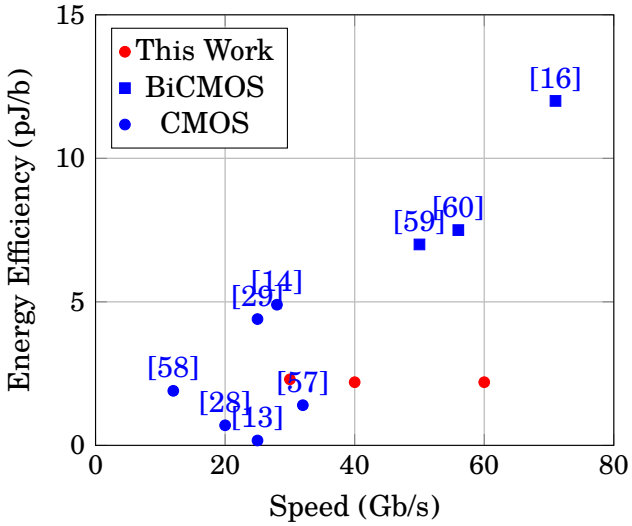

Figure 6.1: Speed and Power

optical receiver ([57]), and close to the maximum speed achieved with BiCMOS processes ([16]). Moreover, its energy efficiency is more than 3 times better compared to BiCMOS optical receivers and is on par with its CMOS counterparts.

In Fig. 6.2, the sensitivity and data-rate of the previously published optical receivers are compared with this work. At high data-rate the sensitivity achieved in this work matches the state of the art BiCMOS optical receiver ([16]), and is better than other CMOS optical receivers at comparable data-rates. A -16.8 dBm OMA was reported in [58]. However, it was optimized for a maximum data-rate of 12 Gb/s.

The techniques used to maximize the performance are explained in detail as well as the circuitry to realize them. The challenge to meet the tough specifications targeted at the beginning of the project lead to noble circuit topologies such as the polarity-switchless phase rotator and the phase accumulator that drives the phase rotator directly. The proposed architecture for the phase rotator and its control logic may also be used in conventional links. It does not have a significant disadvantage over the conventional one, whereas it can be run at higher speed and ensures glitch-free operation.

In 14 nm FinFET technology, layout of the performance critical blocks is an inseparable aspect of the circuit design. The expected parasitics have to be included in the simulations from the very beginning of the design. In order to be able to estimate the parasitic values realistically, the designer needs to be deeply involved in the layout, ideally doing the layout himself/herself. All the performance critical blocks have to be resimulated after parasitic resistance and capacitance extraction to make sure the performance is not compromised. During the implementation of the receiver, some blocks such as the TIA and

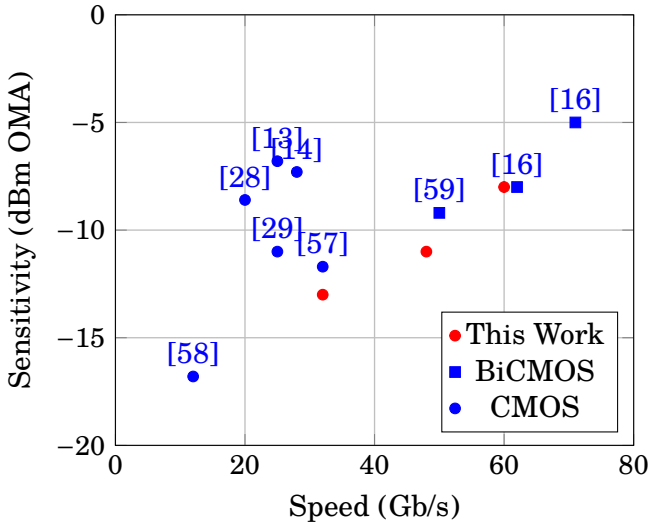

Figure 6.2: Speed and Sensitivity

phase rotator, required few layout iterations to satisfy the target specifications.

Considering the number of design-rule-checks (DRC) to be satisfied and the number of layers to be dealt with, it becomes quite a cumbersome (but unavoidable) task to be done by the designer. During the implementation of this project, as much time was spent on layout as on schematic design in this implementation, if not more. All the effort is spent just to fill-up an active area less than 0.06 mm^2!

6.1 Future Work

This study proves that a burst mode RX can be implemented with a power-on time of less than 10 ns. The RX is the key component for a complete adaptive IO link, including both transmitter and receiver. The achieved performance results are very motivating to pursue the use of adaptive IO links in industry as products. There are various ways to extend the potential application area of the techniques and circuit architectures presented in this study. Some of them may be listed as follows:

- Adaptive Transmitter: This study focuses on the design of the adaptive RX since CDR lock takes most of the time in receiver power-on. As the transmitter is physically placed at the source of the data, in most applications it is source synchronous and does not need a CDR loop. Moreover it does not necessarily have to sense the incoming data as an extra enable pin from the system can control the power-on and power-off periods directly. In that sense, the adaptivity function requires less modifications on transmitter side. Nevertheless, in order to get the complete IO link an adaptive TX must

be desgined.

- Electrical links: Although the specific implementation in this study was an optical RX, the proposed rapid power on and CDR lock algorithm is applicable to electrical links as well, expanding the possible application areas significantly. Furthermore, the receiver was implemented in a 14 nm FinFET technology allowing it to be integrated into digital chips such as processors.
- Long distance links: As the rapid power-on link protocol does not require any acknowledgment to be sent back to the transmitter, the adaptivity concept can be extended to long distance optical applications as well.
- Support for other encoding formats: The rapid power-on technique can also be used with other encoding formats, such as PAM4. Nevertheless, some minor modification on the circuit and data protocol may be required depending on the encoding type and other specifications.
- Standardization: Standardization of the data protocol would allow communication between the products of different vendors broadening the application.

References

[1] R. H. Dennard, F. H. Gaensslen, V. L. Rideout, E. Bassous, and A. R. LeBlanc. "Design of ion-implanted MOSFET's with very small physical dimensions". *IEEE Journal of Solid-State Circuits* 9.5 (Oct. 1974), pp. 256–268. ISSN: 0018-9200. DOI: 10.1109/JSSC.1974.1050511 (see p. 1).

[2] E. P. DeBenedictis. "It's Time to Redefine Moore's Law Again". *Computer* 50.2 (Feb. 2017), pp. 72–75. ISSN: 0018-9162. DOI: 10.1109/MC.2017.34 (see p. 1).

[3] A. Raghavan, Y. Luo, A. Chandawalla, M. Papaefthymiou, K. P. Pipe, T. F. Wenisch, and M. M. K. Martin. "Computational sprinting". *IEEE International Symposium on High-Performance Comp Architecture*. Feb. 2012, pp. 1–12. DOI: 10.1109/HPCA.2012.6169031 (see p. 1).

[4] "Why Energy Is A Big And Rapidly Growing Problem For Data Centers" (2017). Accessed: 2018-02-13. URL: https://www.forbes.com/sites/forbestechcouncil/2017/12/15/why-energy-is-a-big-and-rapidly-growing-problem-for-data-centers/#335b43bd5a30 (see p. 1).

[5] C. Kachris, K. Bergman, and I. Tomkos. *Optical Interconnects for Future Data Center Networks*. Springer, 2013 (see p. 1).

[6] C. Gonzalez and M. Floyd. "The 24-Core POWER9 Processor With Adaptive Clocking, 25-Gb/s Accelerator Links, and 16-Gb/s PCIe Gen4". *IEEE Journal of Solid-State Circuits* 53.1 (Jan. 2018), pp. 91–101. ISSN: 0018-9200 (see p. 2).

[7] A. Roy, H. Zeng, J. Bagga, G. Porter, and A. C. Snoeren. "Inside the Social Network's (Datacenter) Network". *SIGCOMM Comput. Commun. Rev.* 45.4 (Aug. 2015), pp. 123–137. ISSN: 0146-4833. DOI: 10.1145/2829988.2787472. URL: http://doi.acm.org/10.1145/2829988.2787472 (see p. 2).

[8] Q. Zhang, V. Liu, H. Zeng, and A. Krishnamurthy. "High-resolution Measurement of Data Center Microbursts". *Proceedings of the 2017 Internet Measurement Conference*. IMC '17. London, United Kingdom: ACM, 2017, pp. 78–85. ISBN: 978-1-4503-5118-8. DOI: 10.1145/3131365.3131375. URL: http://doi.acm.org/10.1145/3131365.3131375 (see p. 2).

[9] P. M., A. Z., G. S., and V. G. *Handbook on Data Centers. Techniques to Achieve Energy Proportionality in Data Centers: A Survey*. Springer, 2015 (see p. 2).

[10] S. Nedevschi, L. Popa, G. Iannaccone, S. Ratnasamy, and D. Wetherall. "Reducing Network Energy Consumption via Sleeping and Rate-adaptation". *Proceedings of the 5th USENIX Symposium on Networked Systems Design and Implementation*. NSDI'08. San Francisco, California: USENIX Association, 2008, pp. 323–336. ISBN: 111-999-5555-22-1. URL: http://dl.acm.org/citation.cfm?id=1387589.1387612 (see p. 2).

[11] M. Gupta and S. Singh. "Using Low-Power Modes for Energy Conservation in Ethernet LANs". *IEEE INFOCOM 2007 - 26th IEEE International Conference on Computer Communications*. May 2007, pp. 2451–2455. DOI: 10.1109/INFCOM.2007.299 (see pp. 2, 3).

[12] B. Heller, S. Seetharaman, P. Mahadevan, Y. Yiakoumis, P. Sharma, S. Banerjee, and N. McKeown. "ElasticTree: Saving Energy in Data Center Networks". *Proceedings of the 7th USENIX Conference on Networked Systems Design and Implementation*. NSDI'10. San Jose, California: USENIX Association, 2010, pp. 17–17. URL: http://dl.acm.org/citation.cfm?id=1855711.1855728 (see p. 3).

[13] S. Saeedi and A. Emami. "A 25Gb/s 170 μ W/Gb/s optical receiver in 28nm CMOS for chip-to-chip optical communication". *2014 IEEE Radio Frequency Integrated Circuits Symposium*. June 2014, pp. 283–286. DOI: 10.1109/RFIC.2014.6851720 (see pp. 4, 27, 114, 115).

[14] T. Takemoto, H. Yamashita, T. Yazaki, N. Chujo, Y. Lee, and Y. Matsuoka. "A 25-to-28 Gb/s High-Sensitivity (−9.7 dBm) 65 nm CMOS Optical Receiver for Board-to-Board Interconnects". *IEEE Journal of Solid-State Circuits* 49.10 (Oct. 2014), pp. 2259–2276. ISSN: 0018-9200. DOI: 10.1109/JSSC.2014.2349976 (see pp. 4, 114, 115).

[15] T. Anand, M. Talegaonkar, A. Elkholy, S. Saxena, A. Elshazly, and P. K. Hanumolu. "3.7 A 7Gb/s rapid on/off embedded-clock serial-link transceiver with 20ns power-on time, 740 μW off-state power for energy-proportional links in 65nm CMOS". *2015 IEEE International Solid-State Circuits Conference - (ISSCC) Digest of Technical Papers*. Feb. 2015, pp. 1–3. DOI: 10.1109/ISSCC.2015.7062927 (see pp. 4, 113).

[16] D. M. Kuchta, A. V. Rylyakov, F. E. Doany, C. L. Schow, J. E. Proesel, C. W. Baks, P. Westbergh, J. S. Gustavsson, and A. Larsson. "A 71-Gb/s NRZ Modulated 850-nm VCSEL-Based Optical Link". *IEEE Photonics Technology Letters* 27.6 (Mar. 2015), pp. 577–580. ISSN: 1041-1135. DOI: 10.1109/LPT.2014.2385671 (see pp. 5, 97, 114, 115).

[17] S. Gondi and B. Razavi. "Equalization and Clock and Data Recovery Techniques for 10-Gb/s CMOS Serial-Link Receivers". *IEEE Journal of Solid-State Circuits* 42.9 (Sept. 2007), pp. 1999–2011. ISSN: 0018-9200. DOI: 10.1109/JSSC.2007.903076 (see p. 11).

[18] C. A. Belfiore and J. H. Park. "Decision feedback equalization". *Proceedings of the IEEE* 67.8 (Aug. 1979), pp. 1143–1156. ISSN: 0018-9219. DOI: 10.1109/PROC.1979.11409 (see p. 12).

[19] S. Ibrahim and B. Razavi. "Low-Power CMOS Equalizer Design for 20-Gb/s Systems". *IEEE Journal of Solid-State Circuits* 46.6 (June 2011), pp. 1321–1336. ISSN: 0018-9200. DOI: 10.1109/JSSC.2011.2134450 (see p. 13).

[20] D. Duttweiler, J. Mazo, and D. Messerschmitt. "An upper bound on the error probability in decision-feedback equalization". *IEEE Transactions on Information Theory* 20.4 (July 1974), pp. 490–497. ISSN: 0018-9448. DOI: 10.1109/TIT.1974.1055246 (see p. 14).

[21] T. Beukema, M. Sorna, K. Selander, S. Zier, B. L. Ji, P. Murfet, J. Mason, W. Rhee, H. Ainspan, B. Parker, and M. Beakes. "A 6.4-Gb/s CMOS SerDes core with feed-forward and decision-feedback equalization". *IEEE Journal of Solid-State Circuits* 40.12 (Dec. 2005), pp. 2633–2645. ISSN: 0018-9200. DOI: 10.1109/JSSC.2005.856584 (see p. 14).

[22] A. Sharif-Bakhtiar and A. C. Carusone. "A 20 Gb/s CMOS Optical Receiver With Limited-Bandwidth Front End and Local Feedback IIR-DFE". *IEEE Journal of Solid-State Circuits* 51.11 (Nov. 2016), pp. 2679–2689. ISSN: 0018-9200. DOI: 10.1109/JSSC.2016.2602224 (see pp. 20, 28).

[23] B. Razavi. *Design of Integrated Circuits for Optical Communications*. 1st ed. New York, NY, USA: McGraw-Hill, Inc., 2003. ISBN: 0072822589, 9780072822588 (see p. 20).

[24] F. Y. Liu, D. Patil, J. Lexau, P. Amberg, M. Dayringer, J. Gainsley, H. F. Moghadam, X. Zheng, J. E. Cunningham, A. V. Krishnamoorthy, E. Alon, and R. Ho. "10-Gbps, 5.3-mW Optical Transmitter and Receiver Circuits in 40-nm CMOS". *IEEE Journal of Solid-State Circuits* 47.9 (Sept. 2012), pp. 2049–2067. ISSN: 0018-9200. DOI: 10.1109/JSSC.2012.2197234 (see p. 24).

[25] K. Yu, C. H. Chen, C. Li, H. Li, A. Titriku, B. Wang, A. Shafik, Z. Wang, M. Fiorentino, P. Y. Chiang, and S. Palermo. "25Gb/s hybrid-integrated silicon photonic receiver with microring wavelength stabilization". *2015 Optical Fiber Communications Conference and Exhibition (OFC)*. Mar. 2015, pp. 1–3 (see p. 27).

[26] M. H. Nazari and A. Emami-Neyestanak. “A 24-Gb/s Double-Sampling Receiver for Ultra-Low-Power Optical Communication”. *IEEE Journal of Solid-State Circuits* 48.2 (Feb. 2013), pp. 344–357. ISSN: 0018-9200. DOI: 10.1109/JSSC.2012.2227612 (see p. 27).

[27] J. Proesel, A. Rylyakov, and C. Schow. “Optical receivers using DFE-IIR equalization”. *2013 IEEE International Solid-State Circuits Conference Digest of Technical Papers*. Feb. 2013, pp. 130–131. DOI: 10.1109/ISSCC.2013.6487668 (see p. 28).

[28] A. Sharif-Bakhtiar, M. G. Lee, and A. C. Carusone. “Low-Power CMOS Receivers For Short Reach Optical Communication”. *Proceedings of the IEEE 2017 Custom Integrated Circuits Conference*. 2017 (see pp. 28, 114, 115).

[29] A. Rylyakov, J. Proesel, S. Rylov, B. Lee, J. Bulzacchelli, A. Ardey, B. Parker, M. Beakes, C. Baks, C. Schow, and M. Meghelli. “22.1 A 25Gb/s burst-mode receiver for rapidly reconfigurable optical networks”. *2015 IEEE International Solid-State Circuits Conference - (ISSCC) Digest of Technical Papers*. Feb. 2015, pp. 1–3. DOI: 10.1109/ISSCC.2015.7063095 (see pp. 29, 113–115).

[30] E. M. Cherry and D. E. Hooper. “The design of wide-band transistor feedback amplifiers”. *Electrical Engineers, Proceedings of the Institution of* 110.2 (Feb. 1963), pp. 375–389. ISSN: 0020-3270. DOI: 10.1049/piee.1963.0050 (see p. 35).

[31] D. Z. Turker, A. Rylyakov, D. Friedman, S. Gowda, and E. Sanchez-Sinencio. “A 19Gb/s 38mW 1-tap speculative DFE receiver in 90nm CMOS”. *2009 Symposium on VLSI Circuits*. June 2009, pp. 216–217 (see p. 36).

[32] P. A. Francese, M. Brandli, C. Menolfi, M. Kossel, T. Morf, L. Kull, A. Cevrero, H. Yueksel, I. Oezkaya, D. Luu, and T. Toifl. “23.6 A 30Gb/s 0.8pJ/b 14nm FinFET receiver data-path”. *2016 IEEE International Solid-State Circuits Conference (ISSCC)*. Jan. 2016, pp. 408–409. DOI: 10.1109/ISSCC.2016.7418080 (see pp. 36, 39).

[33] K. K. Parhi. “Design of multigigabit multiplexer-loop-based decision feedback equalizers”. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 13.4 (Apr. 2005), pp. 489–493. ISSN: 1063-8210. DOI: 10.1109/TVLSI.2004.842935 (see p. 36).

[34] T. Shibasaki, W. Chaivipas, Y. Chen, Y. Doi, T. Hamada, H. Takauchi, T. Mori, Y. Koyanagi, and H. Tamura. “A 56-Gb/s receiver front-end with a CTLE and 1-tap DFE in 20-nm CMOS”. *2014 Symposium on VLSI Circuits*

Digest of Technical Papers. June 2014, pp. 1–2. DOI: 10.1109/VLSIC.2014.6858400 (see p. 36).

[35] T. B. Cho and P. R. Gray. "A 10 b, 20 Msample/s, 35 mW pipeline A/D converter". *IEEE Journal of Solid-State Circuits* 30.3 (Mar. 1995), pp. 166–172. ISSN: 0018-9200. DOI: 10.1109/4.364429 (see p. 39).

[36] B. Razavi. "Challenges in the design high-speed clock and data recovery circuits". *IEEE Communications Magazine* 40.8 (Aug. 2002), pp. 94–101. ISSN: 0163-6804. DOI: 10.1109/MCOM.2002.1024421 (see p. 40).

[37] J. L. Sonntag and J. Stonick. "A Digital Clock and Data Recovery Architecture for Multi-Gigabit/s Binary Links". *IEEE Journal of Solid-State Circuits* 41.8 (Aug. 2006), pp. 1867–1875. ISSN: 0018-9200. DOI: 10.1109/JSSC.2006.875292 (see pp. 40, 44, 49).

[38] C. R. Hogge. "A self correcting clock recovery circuit". *IEEE Transactions on Electron Devices* 32.12 (Dec. 1985), pp. 2704–2706. ISSN: 0018-9383. DOI: 10.1109/T-ED.1985.22402 (see p. 41).

[39] J. D. H. Alexander. "Clock recovery from random binary signals". *Electronics Letters* 11.22 (Oct. 1975), pp. 541–542. ISSN: 0013-5194. DOI: 10.1049/el:19750415 (see pp. 41, 49).

[40] F. A. Musa. "High-speed baud-rate clock recovery". PhD thesis. University of Toronto, 2008 (see p. 41).

[41] K. Mueller and M. Muller. "Timing Recovery in Digital Synchronous Data Receivers". *IEEE Transactions on Communications* 24.5 (May 1976), pp. 516–531. ISSN: 0090-6778. DOI: 10.1109/TCOM.1976.1093326 (see p. 43).

[42] G. R. Gangasani, C. M. Hsu, J. F. Bulzacchelli, T. Beukema, W. Kelly, H. H. Xu, D. Freitas, A. Prati, D. Gardellini, R. Reutemann, G. Cervelli, J. Hertle, M. Baecher, J. Garlett, P. A. Francese, J. F. Ewen, D. Hanson, D. W. Storaska, and M. Meghelli. "A 32 Gb/s Backplane Transceiver With On-Chip AC-Coupling and Low Latency CDR in 32 nm SOI CMOS Technology". *IEEE Journal of Solid-State Circuits* 49.11 (Nov. 2014), pp. 2474–2489. ISSN: 0018-9200. DOI: 10.1109/JSSC.2014.2340574 (see pp. 44, 51).

[43] H. Wang and A. Hajimiri. "A Wideband CMOS Linear Digital Phase Rotator". *2007 IEEE Custom Integrated Circuits Conference*. Sept. 2007, pp. 671–674. DOI: 10.1109/CICC.2007.4405821 (see p. 52).

[44] J. F. Bulzacchelli, M. Meghelli, S. V. Rylov, W. Rhee, A. V. Rylyakov, H. A. Ainspan, B. D. Parker, M. P. Beakes, A. Chung, T. J. Beukema, P. K. Pepeljugoski, L. Shan, Y. H. Kwark, S. Gowda, and D. J. Friedman. "A 10-Gb/s 5-Tap DFE/4-Tap FFE Transceiver in 90-nm CMOS Technology". *IEEE*

Journal of Solid-State Circuits 41.12 (Dec. 2006), pp. 2885–2900. ISSN: 0018-9200. DOI: 10.1109/JSSC.2006.884342 (see p. 54).

[45] G. R. Gangasani, C. M. Hsu, J. F. Bulzacchelli, S. Rylov, T. Beukema, D. Freitas, W. Kelly, M. Shannon, J. Qi, H. H. Xu, J. Natonio, T. Rasmus, J. R. Guo, M. Wielgos, J. Garlett, M. A. Sorna, and M. Meghelli. "A 16-Gb/s Backplane Transceiver With 12-Tap Current Integrating DFE and Dynamic Adaptation of Voltage Offset and Timing Drifts in 45-nm SOI CMOS Technology". *IEEE Journal of Solid-State Circuits* 47.8 (Aug. 2012), pp. 1828–1841. ISSN: 0018-9200. DOI: 10.1109/JSSC.2012.2196313 (see p. 54).

[46] P. A. Francese, T. Toifl, P. Buchmann, M. Brändli, C. Menolfi, M. Kossel, T. Morf, L. Kull, and T. M. Andersen. "A 16 Gb/s 3.7 mW/Gb/s 8-Tap DFE Receiver and Baud-Rate CDR With 31 kppm Tracking Bandwidth". *IEEE Journal of Solid-State Circuits* 49.11 (Nov. 2014), pp. 2490–2502. ISSN: 0018-9200. DOI: 10.1109/JSSC.2014.2344008 (see pp. 54–56).

[47] K.-h. Kim, Y.-S. Sohn, C.-K. Kim, M. Park, D.-J. Lee, W.-S. Kim, and C. Kim. "A 20-gb/s 256-mb DRAM with an inductorless quadrature PLL and a cascaded pre-emphasis transmitter". *IEEE Journal of Solid-State Circuits* 41.1 (Jan. 2006), pp. 127–134. ISSN: 0018-9200. DOI: 10.1109/JSSC.2005.859017 (see pp. 56, 58).

[48] T. Kusaga and T. Shima. "Four-stage ring oscillator for quadrature signal generation". *2008 International Conference on Signals and Electronic Systems*. Sept. 2008, pp. 89–92. DOI: 10.1109/ICSES.2008.4673365 (see p. 56).

[49] B. Casper and F. O'Mahony. "Clocking Analysis, Implementation and Measurement Techniques for High-Speed Data Links x2014;A Tutorial". *IEEE Transactions on Circuits and Systems I: Regular Papers* 56.1 (Jan. 2009), pp. 17–39. ISSN: 1549-8328. DOI: 10.1109/TCSI.2008.931647 (see p. 56).

[50] F. E. Doany, D. M. Kuchta, and J. A. Kash. "Measurement of picosecond transverse mode dynamics in data communication VCSELs beyond 10 Gbit/sec". *Conference on Lasers and Electro-Optics, 2004. (CLEO).* May 2004 (see p. 71).

[51] K. Petermann. *Laser Diode Modulation and Noise*. Jan. 1991 (see p. 71).

[52] T. Morf, M. Seifried, A. Cevrero, I. Ozkaya, C. Menolfi, D. Kuchta, M. Kossel, P. Francese, L. Kull, J. Kropp, and T. Toifl. "VCSEL-based optical links in burst-mode slow optical power ramp-up and how to achieve ultra-short wake-up times". *Electronics Letters* 53.19 (2017), pp. 1325–1327. ISSN: 0013-5194. DOI: 10.1049/el.2017.2513 (see pp. 71, 72).

[53] "Pulsed Operation of VCSELs for High Peak Powers" (2007). Accessed: 2017-12-28. URL: https://www.finisar.com/sites/default/files/downloads/application_note_pulsed_operation_of_vcsels_for_high_peak_powers.pdf (see p. 71).

[54] N. Dupuis, D. M. Kuchta, F. E. Doany, A. Rylyakov, J. Proesel, C. W. Baks, C. L. Schow, S. Luong, C. Xie, L. Wang, S. Huang, K. Jackson, and N. Y. Li. "Exploring the limits of high-speed receivers for multimode VCSEL-based optical links". *OFC 2014*. Mar. 2014, pp. 1–3. DOI: 10.1364/OFC.2014.M3G.5 (see p. 97).

[55] J. Zerbe, B. Daly, W. Dettloff, T. Stone, W. Stonecypher, P. Venkatesan, K. Prabhu, B. Su, J. Ren, B. Tsang, B. Leibowitz, D. Dunwell, A. C. Carusone, and J. Eble. "A 5.6Gb/s 2.4mW/Gb/s bidirectional link with 8ns power-on". *2011 Symposium on VLSI Circuits - Digest of Technical Papers*. June 2011, pp. 82–83 (see p. 113).

[56] B. Leibowitz, R. Palmer, J. Poulton, Y. Frans, S. Li, J. Wilson, M. Bucher, A. M. Fuller, J. Eyles, M. Aleksic, T. Greer, and N. M. Nguyen. "A 4.3 GB/s Mobile Memory Interface With Power-Efficient Bandwidth Scaling". *IEEE Journal of Solid-State Circuits* 45.4 (Apr. 2010), pp. 889–898. ISSN: 0018-9200. DOI: 10.1109/JSSC.2010.2040230 (see p. 113).

[57] J. E. Proesel, Z. Toprak-Deniz, A. Cevrero, I. Ozkaya, S. Kim, D. M. Kuchta, S. Lee, S. V. Rylov, H. Ainspan, T. O. Dickson, J. F. Bulzacchelli, and M. Meghelli. "A 32 Gb/s, 4.7 pJ/bit Optical Link With -11.7 dBm Sensitivity in 14-nm FinFET CMOS". *IEEE Journal of Solid-State Circuits* PP.99 (2017), pp. 1–13. ISSN: 0018-9200. DOI: 10.1109/JSSC.2017.2778280 (see pp. 114, 115).

[58] M. G. Ahmed, M. Talegaonkar, A. Elkholy, G. Shu, A. Elmallah, A. Rylyakov, and P. K. Hanumolu. "A 12-Gb/s -16.8-dBm OMA Sensitivity 23-mW Optical Receiver in 65-nm CMOS". *IEEE Journal of Solid-State Circuits* 53.2 (Feb. 2018), pp. 445–457. ISSN: 0018-9200 (see pp. 114, 115).

[59] T. Takemoto, Y. Matsuoka, H. Yamashita, Y. Lee, H. Arimoto, M. Kokubo, and T. Ido. "A 50-Gb/s High-Sensitivity (-9.2 dBm) Low-Power (7.9 pJ/bit) Optical Receiver Based on 0.18- μ m SiGe BiCMOS Technology". *IEEE Journal of Solid-State Circuits* PP.99 (2018), pp. 1–21. ISSN: 0018-9200. DOI: 10.1109/JSSC.2018.2791474 (see pp. 114, 115).

[60] T. Takemoto, Y. Matsuoka, H. Yamashita, Y. Lee, K. Akita, H. Arimoto, and M. Kokubo. "A jitter-reduction packaging structure for a 56-Gb/s NRZ modulated optical receiver". *2016 Optical Fiber Communications Conference and Exhibition (OFC)*. Mar. 2016, pp. 1–3 (see p. 114).

Ilter Ozkaya

Personal Information

Address	Rutistrasse 78, 8134, Adliswil-Switzerland
Phone	+41 (78) 827 1999
e-mail	ilterozkaya@gmail.com
Date of Birth	03/12/1983
Nationality	Turkish and Bulgarian

Work Experience

2014–2018 **IBM Zurich Research Laboratory**, Switzerland
Position Pre-Doctoral Researcher
High-Speed Optical Receiver Design in 14 nm FinFET:

2007–2014 **Mikroelektronik R&D Ltd.** (Subsidiary of ASELSAN), Turkey
Position Analog IC Design Engineer
Mixed Signal and Analog Circuit Design:

Education

2014–2018 **PhD., Ecole Polytechnique Federale de Lausanne**, Lausanne
Doctoral Program in Microsystems & Microelectronics

2007–2010 **M.S., Istanbul Technical University**, Istanbul
Electronics Engineering Master Program

2002–2007 **B.S., Middle East Technical University**, Ankara
Electrical and Electronics Engineering

Languages

Turkish (Native Language)
English (Fluent)

Research Interests

Mixed-Signal IC Design
High-Speed Electrical/Optical IO Link Design
Low-Power Low-Voltage Analog CMOS Circuit Design

Computer skills

Design Environment
Cadence: ADE, Virtuoso Layout
Synopsys: Galaxy Custom Designer

System Level Modeling:
Matlab, Python, VerilogA