

# Optimizing the Photodetector/Analog Front-End Interface in Optical Communication Receivers

Bahaa Radi, *Member, IEEE*, Zonghao Li, *Member, IEEE*, Dhruv Patel, *Member, IEEE*, and Anthony Chan Carusone, *Fellow, IEEE*

**Abstract**—This paper addresses optimizing the interface between the photodetector (PD) and the analog front-end (AFE) in high-speed, high-density optical communication receivers. Specifically, the paper focuses on optimizing design elements in the interface, including the interconnecting transmission line, the T-coil, the transimpedance amplifier (TIA), and digital equalization tap weights. To optimize the optical link, we use a combination of analytical models, electromagnetic simulations (EM), and machine learning (ML) techniques to describe different interface elements as most appropriate for each. Finally, we use the genetic algorithm to obtain optimal design parameters. The proposed optimization approach leads to a quick design time and reveals insights into best design practices. For example, we use the proposed method to investigate the relationship between optimal transmission line width and the amount of equalization available on the receiver. These conclusions are further supported by measurements of an assembled prototype with various PD-to-TIA interconnect lengths.

**Index Terms**—Circuit noise, decision-feedback equalizer, feed-forward equalizer, optical receivers, pulse amplitude modulation, sensitivity, machine learning.

## I. INTRODUCTION

To support the demand for the current 400G and emerging 800G and 1.6T Ethernet standards in data centers, the per-lane data rate and the number of lanes must be increased. Higher-order modulation implementations such as PAM-6 and PAM-8 are in active research to improve the per-lane data rate. Moreover, as the limited bandwidth of the analog front-end (AFE) has an increasingly detrimental effect on intersymbol interference (ISI) for higher-order modulation schemes, equalization techniques are used to account for the limited bandwidth. On the other hand, increasing the number of lanes presents packaging challenges on the receiver side.

Many integrated CMOS optical receivers were developed on the receiver side, allowing the AFE and the SerDes circuits to coexist on one chip, such as the 100 Gb/s 4-PAM optical receiver in [1], the linear transimpedance amplifier (TIA) in 16nm FinFET in [2], the linear TIA in 28nm CMOS in [3], and the linear TIA co-packaged with the photodiode in [4]. However, the photodetector (PD) remains a discrete

Manuscript received XXXX XX, 2022; revised XXXX XX, 2022; accepted XXXX X, 2022. Date of publication XXXX X, 2022; date of current version XXXX X, 2022. This brief was recommended by Associate Editor X. XXXXXX. (Corresponding author: Bahaa Radi.)

Bahaa Radi, Zonghao Li, Dhruv Patel, and Anthony Chan Carusone are with the Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G4, Canada (e-mail: bahaa.radi@isl.utoronto.ca; zonghao.li@isl.utoronto.ca; dhruv.patel@isl.utoronto.ca; tony.chan.carusone@isl.utoronto.ca).

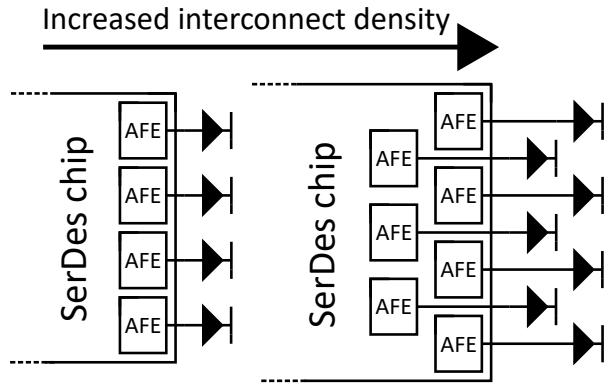


Fig. 1. Illustration of increased interconnect density leading to longer and potentially different interconnect lengths.

component. Since silicon-based CMOS technologies are not optimized for efficient light absorption [5], PDs are typically made from germanium or compound semiconductors (e.g., InGaAs), that offer better sensitivity and responsivity to light [6]. Such PDs can be designed and optimized independently or in an array to achieve a desired combination of high responsivity, low noise, and wide bandwidth. Alternatively, they may be integrated into a silicon photonic platform alongside other optical components. In either case, the PD is generally not monolithically integrated with a DSP-based equalizing front-end, implying package-level heterogeneous integration of the PD and front-end. With the eventual increase of the number of data lanes in the near future, the spacing between the discrete PD and their corresponding front-ends will inevitably increase, as demonstrated in Fig. 1. Moreover, this leads to different interconnect lengths between the PD and the AFE. Consequently, more parasitics will be present at the optical receiver's input. Signal integrity impairments such as reflections will manifest at the PD and AFE interface. To mitigate these impairments and the impact of the added parasitics on AFE performance, the package and the AFE should be co-designed for optimal performance. Moreover, the optimal AFE design is different for various interconnect lengths. This necessitates developing an automated and fast optimization flow that considers interconnect design.

System-level high-speed data link modelling and optimization have been studied intensively in recent years. Prior work such as [7], [8], and [9] primarily focused on modelling equalizers without detailed consideration of the proceeding AFE. In particular, [8] and [9] do not consider the noise. In [10], the

authors presented using machine learning (ML) techniques to model SerDes systems without providing much design insight. Reference [11] presents the results of an IBIS-AMI holistic model but needs to describe many key implementation details.

This paper demonstrates the holistic modelling and optimization of the packaging interface and the AFE of an optical communication receiver. We have the following contributions to make our work distinct from others. First, we discuss how each AFE block is modelled in detail and make them open-sourced<sup>1</sup> so that readers can reuse the information provided in this work. We use foundry-provided models to capture their impact on the design accurately. Second, we consider both jitter and noise so that their degradation of channel performance can be investigated. Third, we have the T-coil included in our link model, and we apply some novel ML techniques to accelerate its modelling.

The structure of this paper is organized as follows. Section II describes the modelling of the parts of the interface under consideration. Section III discusses the optimization procedure. Section IV presents the optimization results and discussion. Section V presents the experimental validations. Finally, Section VI concludes the paper.

## II. MODELING THE ANALOG FRONT-END

Our model considers an optical receiver using 4-PAM modulation with a baud rate of 64 Gbaud (128 Gb/s). The interface is shown in Fig. 2, along with corresponding models, comprising a discrete photodetector connected to the optical receiver through some packaging interconnect. A low-cost organic substrate is assumed here. Electrostatic discharge (ESD) protection circuits are required to prevent damage during activities such as manufacturing and assembly. The ESD circuits introduce parasitic capacitances that could harm performance. To ameliorate this, a bridged T-coil circuit is introduced to extend bandwidth. Then a transimpedance amplifier (TIA) is followed by additional variable-gain amplification (VGA) stages and an analog-to-digital (ADC) converter. In the following optimization, we assume the noise and impairments of the VGA and ADC are negligible compared to the noise and bandwidth limitations of the TIA. Finally, digital equalization is used to remove ISI at the output of the TIA. A model of the complete front end is formed by combining analytical (two-port linear) models, electromagnetic (EM) simulations, and ML techniques for each element in the model as appropriate. Each component in Fig. 2 will be described in detail in the following subsections.

### A. Modeling the Photodetector and the Interconnect

The PD is modelled as an ideal current source in parallel with a junction capacitance,  $C_{PD}$ , of  $10.7\text{ fF}$ , and a series resistance,  $R_{PD}$ , of  $87\Omega$ . These values are based on the GlobalFoundries (GF) 45SPCLO CMOS (silicon photonics) process but are also consistent with standalone germanium PDs, and other silicon photonics [12]. The rise and fall times of the signal are assumed to be  $6\text{ ps}$ , corresponding to around

<sup>1</sup>Source code: [https://github.com/ChrisZonghaoLi/optical\\_receiver\\_optimization](https://github.com/ChrisZonghaoLi/optical_receiver_optimization)

0.4 unit interval (UI) at 64 Gbaud. We also assume that the PD generates a peak-to-peak current  $I_{pp} = 100\mu\text{A}$ . The PD source impedance is analytically modelled using the following ABCD matrix:

$$T_{PD} = \begin{bmatrix} 1 & R_{PD} \\ sC_{PD} & sR_{PD}C_{PD} + 1 \end{bmatrix} \quad (1)$$

where  $s$  is the complex frequency. In terms of packaging, we assume flip-chip packaging of the PD die, and the AFE die onto an organic substrate [4]. The solder bump introduces parasitics and discontinuity. It is modeled as  $10\text{ fF}$  shunt capacitance,  $C_{bump}$ , and a  $20\text{ pH}$  series inductance,  $L_{bump}$  [13]. The ABCD matrix of the bump on the PD side,  $T_{bump}$ , is given by:

$$T_{bump} = \begin{bmatrix} 1 + s^2 L_{bump} C_{bump} & sL_{bump} \\ sC_{bump} & 1 \end{bmatrix} \quad (2)$$

As the distance between PDs and AFEs increases, the trace interconnecting the two should be designed as a transmission line to alleviate reflections and signal degradation. Thus, we consider a transmission line connecting the PD to the AFE here. We consider transmission line lengths,  $TL$ , of  $250\mu\text{m}$  and  $500\mu\text{m}$ , typical values between the PD and the AFE. We also consider a hypothetical transmission line length of  $5\text{ mm}$ . Such a long-length transmission line may be needed to support future high-density interconnects where many PDs are arrayed around and connected to the receiver IC. For a given length  $TL$ , the width of the transmission line  $TW$  is a design parameter, and it is assumed to be bounded between  $15\mu\text{m}$  and  $100\mu\text{m}$  with  $5\mu\text{m}$  steps. A lower limit for the trace width is typically set to ensure the minimum trace width allowed by the organic substrate is manufacturable, while an upper limit is assumed to permit high interconnect density. The width of the transmission line controls the characteristic impedance. The microstrip transmission line was simulated using Ansys HFSS over the design space to obtain its ABCD matrix,  $T_{TL}$ , as a function of frequency. We note that the EM simulations consider losses in the transmission line. The organic substrate stack shown in Fig. 3 was used in these EM simulations. The model assumes an epoxy-based substrate dielectric material developed for high-speed and low-dielectric loss applications [14]. It has a relative dielectric constant of 3.3, dielectric loss at 5.8 GHz of 0.0044, and a surface roughness of 200 nm. A  $15\mu\text{m}$  transmission line results in 0.1 dB loss for  $250\mu\text{m}$ , 0.3 dB for  $500\mu\text{m}$ , and 1.1 dB loss for 5 mm at the Nyquist frequency of 32 GHz.

Similar to the solder bump on the PD side, a solder bump connects the transmission line to the receiver IC,  $T_{bump,rx}$ . The ABCD matrix of this bump is given by:

$$T_{bump,rx} = \begin{bmatrix} 1 & sL_{bump} \\ sC_{bump} & 1 + sL_{bump}C_{bump} \end{bmatrix} \quad (3)$$

The pad on the receiver side introduces a relatively large capacitance that creates a discontinuity at the interface and introduces a pole at the input of the AFE, limiting bandwidth. Here, we assume a fixed pad size regardless of the transmission line width. This assumption is made considering that a large

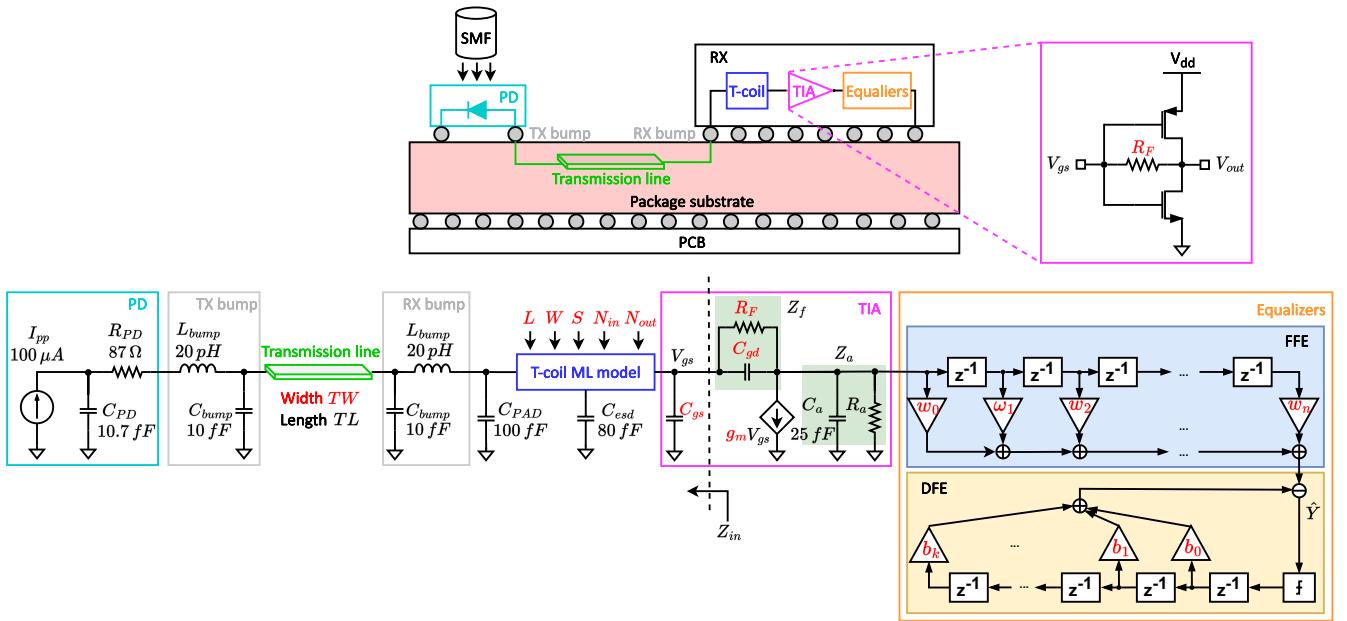


Fig. 2. Interface packaging (top) and the corresponding models used for the optimization (bottom). Colors are used to delineate which model correspond to which component. Design parameters are annotated in red, which are: transmission line width  $TW$ , T-coil geometric parameters  $L$ ,  $W$ ,  $S$ ,  $N_{in}$  and  $N_{out}$ , TIA modeling parameters  $R_F$ ,  $C_{gs}$ ,  $C_{gd}$ , and  $g_m$ , FFE tap weights  $w$ , and DFE tap weights  $b$ .

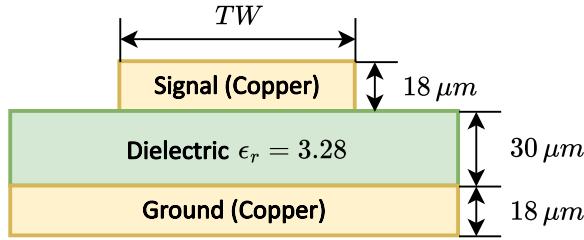


Fig. 3. Stack of the organic substrate used for transmission line simulations.

bonding pad is necessary. A typical capacitance of  $C_{PAD} = 100 fF$  is modeled with the following ABCD matrix:

$$T_{PAD} = \begin{bmatrix} 1 & 0 \\ sC_{PAD} & 1 \end{bmatrix} \quad (4)$$

#### B. Machine Learning Model for T-coil S-parameters Predictions

A bridged T-coil is often incorporated at the input of the AFE to offset the impact of the ESD capacitance  $C_{esd}$ , which is assumed to be  $80 fF$  at the receiver's input, as shown in Fig. 2. The capacitor  $C_{esd}$  is necessary to protect the circuit from ESD. However,  $C_{esd}$  introduces a low-frequency pole, which decreases the front-end bandwidth. The T-coil ameliorates the impact of  $C_{esd}$ . Intuitively, the T-coil essentially introduces inductance on either side of  $C_{esd}$ , creating an artificial LC transmission line that increases the front-end bandwidth while introducing a small delay [15]. The T-coil can be modelled as two mutually coupled inductors with a bridge capacitance [16]. Fig. 4(a) shows the T-coil-enhanced ESD circuit, and Fig. 4(b) shows the layout of a T-coil.

We leverage the up-sampling CNN proposed in [17] to promptly predict each T-coil's S-parameters over a wide

frequency range to bypass time-consuming EM simulations. This is done by taking the T-coil's geometric parameters as inputs to an NN that quickly predicts S-parameters, allowing for accelerated optimization iterations. The design geometric parameters are the T-coil length,  $L$ , width,  $W$ , metal spacing,  $S$ , inner number of turns,  $N_{in}$ , and outer number of turns,  $N_{out}$ . The NN is trained with S-parameters (DC to 256 GHz) from 2920 T-coils. They are simulated with Cadence EMX using 32 Intel Xeon Gold 6242R CPU cores. It takes about 30 hours to prepare these training data. Training the proposed up-sampling CNN on an Nvidia RTX A4000 GPU requires only approximately 10 minutes. Note that these were one-time efforts for this technology. We used K-fold cross-validation to evaluate the NN performance. The loss function is given by Eq. 5, and the EVM in 7 is used as the metric to evaluate the accuracy of the trained model versus the truth. Fig. 7 shows the mean S-parameters EVM of the NN output over 584 test cases. It can be seen that the error increases with frequency. However, given that the Nyquist rate of 4-PAM here is about 32 GHz. According to Fig. 7, for 584 test T-coils, their mean EVM is at most 0.01-0.02 below the Nyquist frequency, which is about -30 dB to -40 dB error.

To demonstrate the idea of the proposed NN<sup>2</sup> and its feasibility, we used a GF 22nm FD-SOI CMOS process as the targeted technology node here since its design kit has built-in T-coil layouts. However, designers can apply this NN to any other technology nodes. Table I lists the T-coil geometric parameter inputs. The NN outputs the T-coil S-parameters' real and imaginary parts as a frequency function. Since the number of input geometric parameters is significantly smaller than the number of output S-parameters, a series of up-sampling layers are required in the NN. A single-input-multi-channel

<sup>2</sup>Source code: <https://github.com/ChrisZonghaoLi/upcnn>

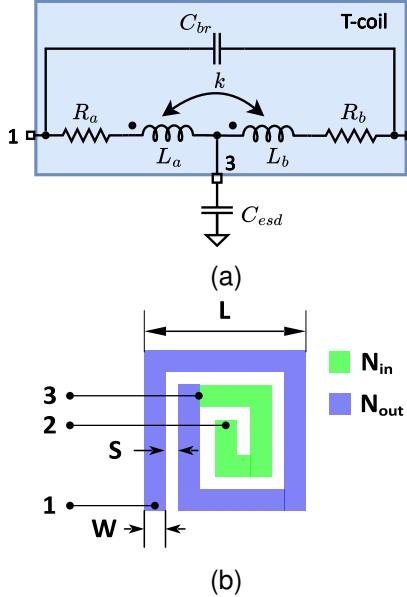


Fig. 4. (a) T-coil-enhanced ESD circuit; (b) T-coil layout, where in this case  $N_{in} = 4$  and  $N_{out} = 5$  [17].

TABLE I  
GEOMETRIC PARAMETERS OF T-COIL IN GF 22NM FD-SOI.

Parameter	Unit	Min	Max
Outer Diameter $L$	$\mu\text{m}$	32	80
Metal Width $W$	$\mu\text{m}$	2.4	5
Metal Spacing $S$	$\mu\text{m}$	1.2	1.44
Inner Segments $N_{in}$	-	5	25
Outer Segments $N_{out}$	-	4	12

deconvolutional layer (DeConv, shown in Fig. 5 (a)) and upsampling convolutional NN (UpCNN, shown in Fig. 5 (b)) are employed to achieve this objective. The upsampling layer can use different algorithms, such as the nearest neighbour and linear interpolation. For simplicity, this work applies the former. Fig. 6 shows the entire structure of the UpCNN. The T-coil's geometric parameters are first mapped to some high-level abstract representation through a multi-layer perceptron (MLP), which is then passed to the DeConv and a series of UpCNNs. The predicted T-coil S-parameters will be the final output of the NN. These S-parameters are converted to the ABCD parameters,  $T_{Tcoil}$ , to represent the T-coil network.

The NN is trained with S-parameters (DC to 256 GHz) from 2920 T-coils. They are simulated with Cadence® EMX® using 32 cores of Intel® Xeon® Gold 6242R CPU. It takes about 30 hours to prepare these training data. Training the NN on an Nvidia RTX A4000 GPU required approximately 10 minutes. Note that these were one-time efforts for this technology. We used K-fold cross-validation to evaluate the NN performance. The loss function used to train and test the proposed model is a modified mean squared error, as in [18]:

$$L_{freq} = \frac{1}{N} \sum_{n=1}^N \sqrt{\frac{1}{K} \sum_{k=1}^K (S_{n,k} - \hat{S}_{n,k})^2} \quad (5)$$

where  $N$  is the number of elements in the training set,  $K$  is

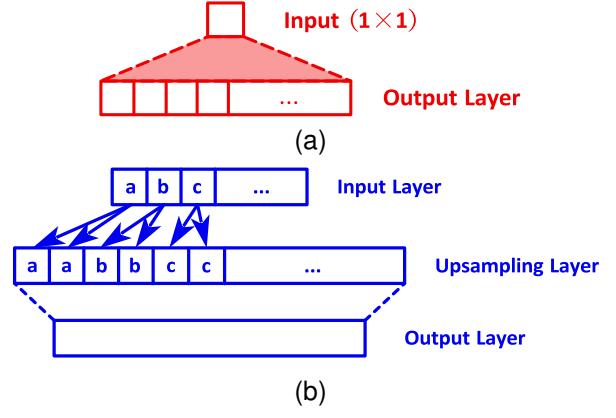


Fig. 5. (a) The structure of DeConv layer; (b) the structure of UpCNN, which consists of a pure upsampling layer and a convolutional layer. The upsampling algorithm used here is the nearest neighbor [17].

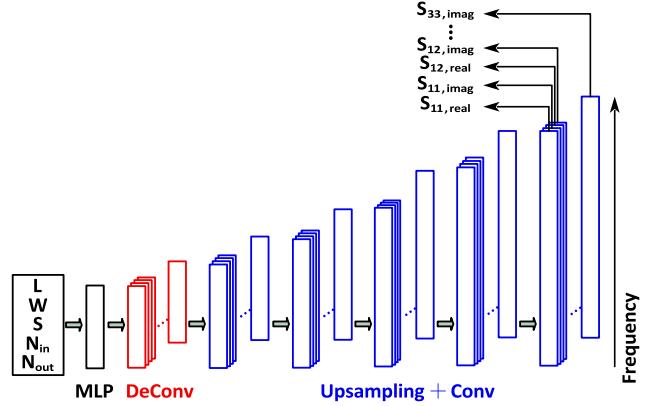


Fig. 6. The structure of the proposed NN for predicting T-coil S-parameters [17].

the number of frequency points,  $S_{n,k}$  is the true S-parameters (obtained by EM simulation) at frequency point  $k$  for the  $n^{th}$  T-coil, and  $\hat{S}_{n,k}$  is the corresponding prediction. This loss function trains the model to minimize the error across all frequencies.

One way to evaluate the accuracy is to evaluate the error vector magnitude between the NN output and the true EM-simulated S-parameters given by:

$$EVM_{n,k} = \sqrt{(S_{n,k} - \hat{S}_{n,k})^2} \quad (6)$$

where  $EVM_{n,k}$  is the error vector magnitude of the S-parameters at the frequency point  $k$  for the  $n^{th}$  T-coil. Fig. 7 shows the mean S-parameters EVM of the NN output over 584 test cases. It can be seen that the error increases with frequency. However, given the Nyquist rate of 4-PAM here is about 32 GHz. According to Fig. 7, for 584 test t-coils, their mean EVM is at most 0.01-0.02 below the Nyquist frequency, which is about -30 dB to -40 dB error.

We have also investigated the performance of our proposed NN in the time domain by examining its derived pulse response. We terminate the middle tap of the T-coil with the  $C_{esd} = 80 \text{ fF}$  as well as both input and output with a  $50 \Omega$

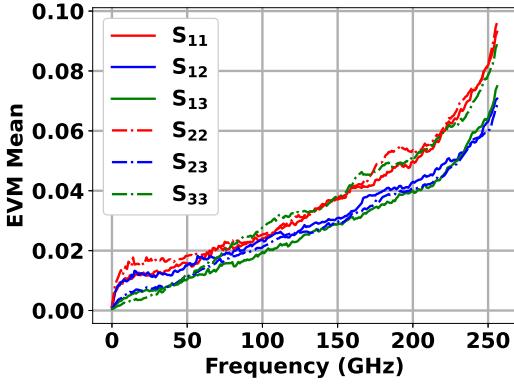
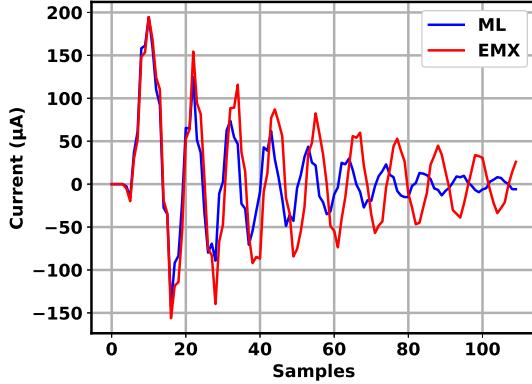


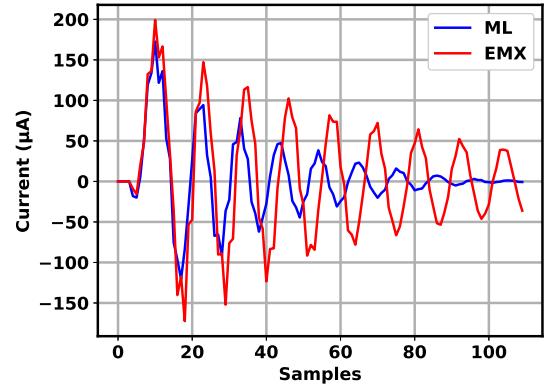
Fig. 7. S-parameter EVM mean with 584 test cases.

Fig. 8. T-coil pulse response for  $TL = 250 \mu m$  with  $[L, W, N_{in}, N_{out}] = [43, 4.2, 11, 6]$ .

resistor. We convolve its impulse response with a current pulse  $I_{PP}$  to generate the pulse response, which is then compared to the one generated from the EMX® simulation. For example, the optimal T-coils for  $TL = 250 \mu m$  and  $500 \mu m$  have been examined, with their geometric parameters shown in Table III in Section IV. Note that these T-coils do not necessarily match  $50 \Omega$ . Fig. 8 and 9 show the pulse response comparison for these two T-coils. Our model's predictions toward the main cursor are reasonably accurate but over-optimistic on the post-cursor reflections. This is possibly due to larger EVM in the high-frequency domain, as shown in Fig. 7. The output of the NN is the S-parameter, and the cost function during the training process only evaluates the accuracy of the predicted S-parameters, not the time-domain pulse responses [19]. This is acceptable here since the motivation of our ML model is to replace the EM simulation by promptly predicting the S-parameters of a given T-coil so that the design space can be quickly narrowed down [17].

### C. Modeling the Inverter-Based Shunt-Feedback TIA

The input of the AFE is a TIA that follows the T-coil and converts the input photocurrent into voltage. A commonly used TIA architecture is inverter-based shunt feedback, such as the 128 Gb/s PAM-4 linear TIA in [20], the 64 Gb/s PAM-4 TIA in [21], the 53 Gb/s TIA in [22], and the 64 Gb/s NRZ TIA in [23]. Thus, we use it here. Inverter-based TIAs consist of

Fig. 9. T-coil pulse response for  $TL = 500 \mu m$  with  $[L, W, N_{in}, N_{out}] = [44, 4.2, 11, 5]$ .TABLE II  
TIA SIMULATION PARAMETERS

Parameter	Description	Value
$f_{baud}$	Baud rate	64 Gbaud
$f_t$	Technology transient frequency	$5 \times f_{baud}$
$g_m$	TIA combined transconductance	10-120 mS
$R_F$	TIA Feedback resistor	100-4000 $\Omega$
$C_g$	TIA Gate capacitance	$C_g = g_m/(2\pi f_t)$
$C_{gs}^a$	TIA Gate-to-source capacitance	$2/3 C_g$
$C_{gd}^a$	TIA Gate-to-drain capacitance	$1/3 C_g$
$C_a$	TIA Output capacitance	25 fF
$A$	TIA Inverter gain	6 V/V
$R_a$	TIA output resistance	$A/g_m$

<sup>a</sup> Assumed based on simulations reported in [23].

an inverter with a shunt-feedback resistor converting the input current to output voltage, as illustrated in Fig. 2. Inverter-based TIAs are simple to implement using CMOS technology, allowing them to be integrated alongside DSP equalizers, and have been used in optical receivers at 100 Gb/s and beyond (e.g., [1]). Fig. 10 shows the small signal model of a shunt-feedback TIA with noise sources highlighted in red, where the thermal noise of  $R_F$  is divided into the gate and drain terminal for more straightforward calculations.

The design parameters of the TIA are transistor transconductance  $g_m$  and feedback resistance  $R_F$  considered in our model. Other TIA parameters are dependent on these two. For instance, the value of  $g_m$  is related to the gate capacitance,  $C_g$ , by the cut-off frequency of the technology node. We assume that the transistors are in deep inversion and that the ratio of  $C_{gd}/C_{gs} = 2$  (i.e.,  $C_{gs} = 2/3 C_g$ , and  $C_{gd} = 1/3 C_g$ ) based on [23]. We assume this because the transistor threshold voltage  $V_t$  values are usually significantly below  $V_{gs} = 0.5 V$  in our inverter-based TIA feedback configuration. However, adjusting the  $C_{gs}/C_{gd}$  ratio is important if  $V_t$  approaches  $V_{gs}$ . Table II-C summarizes the numerical values used in this study alongside the relationships between coupled parameters. We note that  $C_a$  represents the combination of the TIA's output capacitance and the following stage's input capacitance. Moreover, considering the limited input current swing we assumed and considering one-stage inverter-based TIA, which is characterized by having good linearity, we ignore linearity issues.

The parameters of the ABCD matrix of the TIA,  $T_{TIA}$ , are

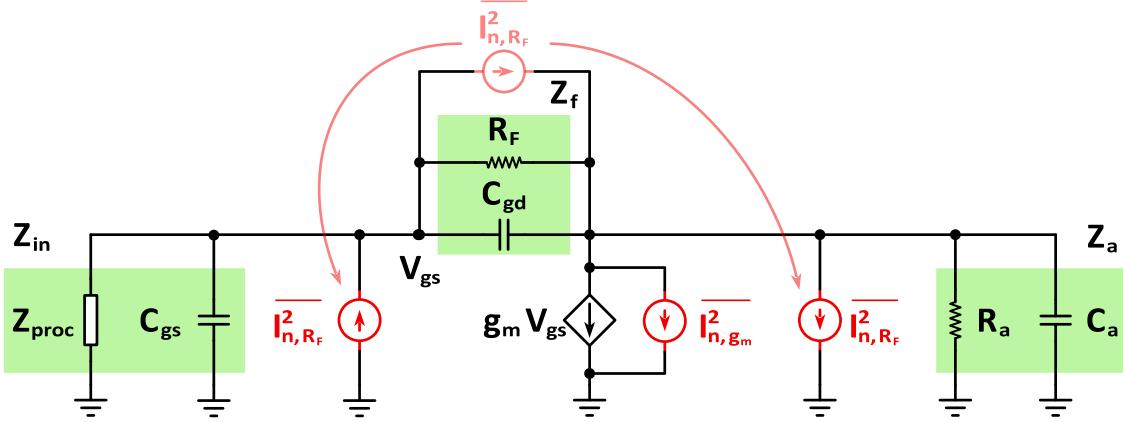


Fig. 10. Small signal model of a shunt-feedback TIA with noise sources highlighted in red. The picture is adapted from [24].

expressed by the following set of equations:

$$A_{TIA} = \frac{R_a + R_F + sR_aR_F(C_a + C_{gd})}{R_a(sC_{gd}R_F - R_Fg_m + 1)} \quad (7a)$$

$$B_{TIA} = \frac{-R_F}{g_mR_F - sC_{gd}R_F - 1} \quad (7b)$$

$$C_{TIA} = \frac{(sC_{gd}R_F + 1)(R_ag_m + sC_aR_a + 1)}{R_a(sC_{gd}R_F - R_Fg_m + 1)} \quad (7c)$$

$$+ \frac{sC_{gs}(R_a + R_F + sC_aR_aR_F + sC_{gd}R_aR_F)}{R_a(sC_{gd}R_F - R_Fg_m + 1)} \quad (7d)$$

$$D_{TIA} = \frac{sC_{gd}R_F + sC_{gs}R_F + 1}{sC_{gd}R_F - R_Fg_m + 1}. \quad (7e)$$

The ABCD matrix of the series connection of all the elements from the PD to the output of TIA is given by:

$$T_{link} = T_{PD}T_{bump}T_{TL}T_{bump,rx}T_{PAD}T_{coil}T_{TIA}. \quad (8)$$

From the  $T_{link}$ , we are interested in the transimpedance from the PD current to the voltage output of the TIA. This transfer function is given by:

$$H(f) = \frac{1}{C_{link}}, \quad (9)$$

where  $C_{link}$  is the  $C$  parameter of the  $T_{link}$  matrix. The impulse response,  $h$ , is obtained by taking the inverse Fourier transform of  $H(f)$ :

$$h = \mathcal{F}^{-1}\{H(f)\}. \quad (10)$$

Finally, the pulse response  $h_{pulse}$ , is obtained by convolving the impulse response  $h$  with the input current pulse  $\Pi$  of 1 UI in duration with 6ps rise- and fall-time:

$$h_{pulse} = h * \Pi. \quad (11)$$

#### D. Noise Analysis of the Inverter-Based Shunt-Feedback TIA

The pulse response captures the time-domain behaviour of the system, such as reflections. However, it does not consider other signal impairments, such as noise and jitter. We will

describe how jitter is taken into account in the later subsection. Highlighted in Fig. 10, the noise contributions arise from the feedback resistance  $R_F$  whose power spectrum density (PSD) is denoted by  $I_{n,R_F}^2$ , and the MOS channel thermal noise of the TIA transistors, whose PSD is denoted as  $I_{n,g_m}^2$ :

$$\overline{I_{n,R_F}^2} = \frac{4kT}{R_f} \quad (12)$$

$$\overline{I_{n,g_m}^2} = 4kT\gamma g_m, \quad (13)$$

where  $k = 1.38 \times 10^{-23} \text{ J/K}$  is the Boltzmann constant,  $T = 300 \text{ K}$  and the semi-empirical factor  $\gamma = 2$ . We can refer the PSD in Eq. 12 and 15 to the output of TIA in terms of voltage:

$$\overline{v_{n,R_F}^2} = \overline{I_{n,R_F}^2} \left| \frac{Z_a \times Z_f \times (1 + g_m \times Z_{in})}{Z_f + Z_{in} + (1 + g_m \times Z_{in}) \times Z_a} \right|^2 \quad (14)$$

$$\overline{v_{n,g_m}^2} = \overline{I_{n,g_m}^2} \left| \frac{Z_a \times (Z_f + Z_{in})}{Z_f + Z_{in} + (1 + g_m \times Z_{in}) \times Z_a} \right|^2. \quad (15)$$

In these expressions,  $Z_{in}$  refers to the impedance looking into the T-coil  $Z_{proc}$  in shunt with the impedance of  $C_{gs}$  of the TIA ( $Z_{in}$  in Fig. 2 and 10),  $Z_f$  is the parallel combination of the feedback resistor  $R_F$  and the gate-to-drain capacitance  $C_{gd}$  ( $Z_f$  in Fig. 2 and 10), and  $Z_a$  is the parallel combination of the equivalent output impedance of the TIA ( $Z_a$  in Fig. 2 and 10). To find  $Z_{proc}$ , we first find out the ABCD matrix of the cascading stages before the TIA as shown in Fig. 11, which can be denoted as:

$$T_{proc} = T_{PD}T_{bump}T_{TL}T_{bump,rx}T_{PAD}T_{coil}. \quad (16)$$

According to the relationship between the ABCD matrix and Z-matrix, we can find out  $Z_{proc}$  by (which is essentially  $Z_{22,proc}$  of the two-port network in Fig. 11) [25]:

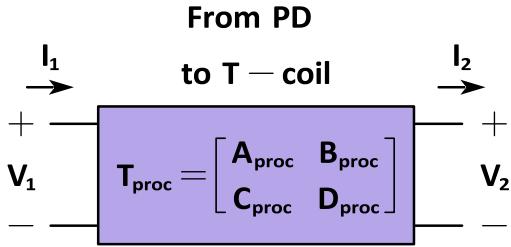


Fig. 11. A two-port model of the cascading stages from PD to T-coil proceeding the TIA model.

$$C_{proc} = \frac{1}{Z_{21,proc}} \quad (17a)$$

$$D_{proc} = \frac{Z_{22,proc}}{Z_{21,proc}} \quad (17b)$$

$$Z_{proc} = Z_{22,proc} = \frac{D_{proc}}{C_{proc}}. \quad (17c)$$

Finally, the total noise variance at the output of the TIA  $\sigma_{n,TIA}^2$  is the integral of the summation of Eq. 14 and 15 assuming they are independent to each other.

$$S_{n,TIA} = \overline{v_{n,R_F}^2} + \overline{v_{n,g_m}^2} \quad (18a)$$

$$\sigma_{n,TIA}^2 = \int_0^\infty S_{n,TIA} df. \quad (18b)$$

### III. EQUALIZATION

In wireline communication, data are usually transmitted successively. Due to the low-pass bandlimited nature of a communication channel, successive transmitted data will likely interfere with each other, especially when the data rate is high. This phenomenon is called intersymbol interference, or ISI. Intuitively, if the high-frequency component of a signal can be recovered, the low-pass exponential rise and fall behaviour can be mitigated, thereby reducing ISI. Equalization methods are frequently used to reduce ISI through the boosting of high-frequency content by the application of circuits such as continuous-time linear equalizer (CTLE), feed-forward equalizer (FFE), and decision-feedback equalizer (DFE). Since our optical receiver model in Fig. 2 does not include CTLE, our discussions will be limited to the latter two. In addition, our discussion of FFE will be confined only to the receiver side. We will then briefly overview the algorithms to obtain the optimal tap coefficients for FFE and DFE under the additive white Gaussian noise (AWGN) channel.

#### A. Feed-Forward Equalizer

FFE is a linear finite impulse response filter (FIR) that boosts the high-frequency content of a signal. In general, the output of a discrete-time FFE can be expressed as:

$$Y[n] = \sum_{i=0}^N b_i X[n-i]. \quad (19)$$

We assume one delay  $z^{-1}$  introduces a time shift of 1 unit interval (UI) for simplicity, although fractional UI delay is also possible. The scaling of the input signal is done by the tap coefficients  $w$ . Applying such a method can reduce both pre-cursor and post-cursor ISI.

From a frequency perspective, FFE has a high-pass function shape that boosts a signal's high-frequency content, compensating for the high-frequency channel loss, but it also increases the total output noise power. To find out the total noise power at the output of FFE due to the noise generated by TIA in Eq. 18, we can simply do the following by exploiting the fact that FFE is a linear time-invariant (LTI):

$$S_{n,FFE}(f) = S_{n,TIA}(f)|H(f)_{FFE}|^2 \quad (20a)$$

$$\sigma_{n,FFE}^2 = \int_0^\infty S_{n,FFE}(f) df. \quad (20b)$$

where  $H_{FFE}(f)$  is the transfer function of FFE. Alternatively, one can find out the total FFE output noise power by exploiting the Wiener–Khintchin theorem:

$$R_{n,FFE}(\tau) = \int_0^\infty S_{n,FFE}(f) e^{j2\pi f \tau} df \quad (21a)$$

$$\sigma_{n,FFE}^2 = \sum_{i=1}^M \sum_{j=1}^M w_i w_k R \left( \frac{|i-k|}{f_{baud}} \right). \quad (21b)$$

where  $w$  is the tap coefficient of FFE, and  $f_{baud}$  is the baud rate of the signal (64 Gband in our model).

#### B. Decision Feedback Equalizer

Unlike FFE, DFE is a non-linear equalizer. One big advantage of FFE versus DFE is that it does not enhance the noise while amplifying the high-frequency content of the signal as the feedback values are digitized by the slicer. The traditional DFE architecture has the exponential scaling of the area and power as a function of tap length, limiting their deployment and ISI removal ability [26]. Thus, a common practice is to combine the use of both FFE and DFE. For instance, FFE can focus on removing the pre-cursor ISI, and DFE can concentrate on mitigating the post-cursor ISI. [27].

#### C. Receiver Equalization Techniques: ZF, MMSE, and MMSE-DFE

This subsection will briefly discuss three different equalization techniques: zero-forcing (ZF), minimum mean square error (MMSE), and minimum mean square error decision-feedback equalizer (MMSE-DFE). The latter is essentially the scenario where FFE and DFE are jointly used to optimize the channel response. The former two cases are used when only FIR equalizers are used.

Fig. 12 depicts a simplified receiver channel for our later discussions, where  $P(t)$  (shorthand as  $P$ ) is the pulse response of the channel.  $N(t)$  (shorthand as  $N$ ) denotes the AWGN due to the noise generated by AFE. The final equalized signal is denoted as  $Z(kT)$  (shorthand as  $Z$ ). Assuming input  $X(t)$  (shorthand as  $X$ ),  $N$ ,  $Y$ , and  $Z$  are all column vectors,

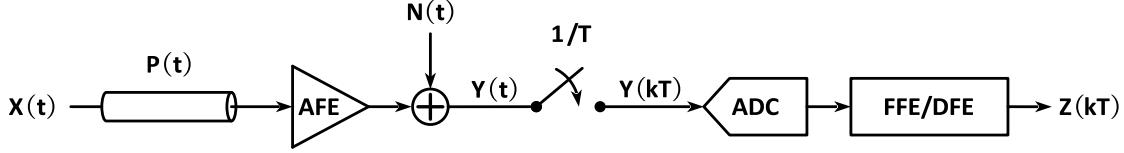


Fig. 12. Simplified receiver channel model.

the signal  $Y(kT)$  (shorthand as  $\mathbf{Y}$ ) before equalized can be written as:

$$\mathbf{Y} = \mathbf{P} * \mathbf{X} + \mathbf{N} = \mathbf{H}\mathbf{X} + \mathbf{N}, \quad (22)$$

where  $\mathbf{H}$  is the Toeplitz matrix of  $\mathbf{P}$ . Ideally, the input signal  $\hat{\mathbf{X}}$  should only carry signal power at the main cursor and zero power elsewhere if it is ISI-free. If we assume  $\hat{\mathbf{X}}$  has  $a$  taps pre-cursor and  $b$  taps post-cursor:

$$\hat{\mathbf{X}} = [\underbrace{0 0 0 \dots 0}_{a \text{ zeros}} \underbrace{x_k 0 0 \dots 0}_{b \text{ zeros}}]. \quad (23)$$

Therefore, we would like to compare the  $k^{th}$  entry of the final equalized output  $Z$  with respect to  $x_k$ . Assuming the equalizer coefficient is a row vector  $\mathbf{w}$  (for simpler notation of avoiding transpose), the equalized output  $z_k$  is defined as the inner product of FIR tap coefficient and  $\mathbf{Y}$ :

$$z_k = \mathbf{w} \cdot \mathbf{Y}. \quad (24)$$

We define the mean squared error (MSE) between the equalized and desired signal as:

$$e = x_k - z_k \quad (25a)$$

$$E[|e|^2] = E[|x_k - z_k|^2] = E[|x_k - \mathbf{w} \cdot \mathbf{Y}|^2]. \quad (25b)$$

Therefore, our goal is to minimize  $|e_k|^2$ , which gives the name minimum mean square error (MMSE). We can take the derivative of it with respect to  $\mathbf{w}$  and set it to zero.

$$E\left[\frac{\partial |e_k|^2}{\partial \mathbf{w}}\right] = E[2(x_k - \mathbf{w} \cdot \mathbf{Y})(-\mathbf{Y}^T)] = -2E[e_k \mathbf{Y}^T] = 0, \quad (26)$$

which is called the orthogonality principle. Now we can substitute Eq. 24 and 25(a) back to Eq. 26, and arriving at the equation below:

$$E[e \mathbf{Y}^T] = E[x_k \mathbf{Y}^T] - \mathbf{w} E[\mathbf{Y} \mathbf{Y}^T] = \mathbf{0} \quad (27)$$

That is, we need to find  $\mathbf{w}$  such that the derivative in Eq. 26, and  $\mathbf{Z}$  becomes the optimal estimator for  $\mathbf{X}$ :

$$\mathbf{w} = E[x_k \mathbf{Y}^T] E[\mathbf{Y} \mathbf{Y}^T]^{-1} = \mathbf{R}_{x\mathbf{Y}} \mathbf{R}_{\mathbf{YY}}^{-1} \quad (28)$$

where  $\mathbf{R}_{x\mathbf{Y}} = E[x_k \mathbf{Y}^T]$  is the cross-correlation row vector between  $x_k$  and  $\mathbf{Y}$ , and  $\mathbf{R}_{\mathbf{YY}} = E[\mathbf{Y} \mathbf{Y}^T]$  is the auto-correlation matrix of  $\mathbf{Y}$ .

We could further expand the expression of  $\mathbf{R}_{x\mathbf{Y}}$  and  $\mathbf{R}_{\mathbf{YY}}$  in terms of the input  $\mathbf{X}$ , channel response Toeplitz matrix  $\mathbf{H}$ , and noise  $\mathbf{N}$ :

$$\begin{aligned} \mathbf{R}_{x\mathbf{Y}} &= E[x_k \mathbf{Y}^T] = E[x_k (\mathbf{H}\mathbf{X} + \mathbf{N})^T] \\ &= \underbrace{E[x_k \mathbf{X}^T]}_{=[0 0 \dots 0 \dots 0]} \mathbf{H}^T + \underbrace{E[x_k \mathbf{N}^T]}_{=0} \\ &= \sigma_x^2 \mathbf{I}_x \end{aligned} \quad (29a)$$

$$\begin{aligned} \mathbf{R}_{\mathbf{YY}} &= E[\mathbf{Y} \mathbf{Y}^T] = \mathbf{H} \underbrace{E[\mathbf{X} \mathbf{X}^T]}_{=\sigma_x^2 \mathbf{I}_x} \mathbf{H}^T + \mathbf{H} \underbrace{E[\mathbf{X} \mathbf{N}^T]}_{=0} \\ &\quad + \underbrace{E[\mathbf{N} \mathbf{X}^T] \mathbf{H}^T}_{=0} + \underbrace{E[\mathbf{N} \mathbf{N}^T]}_{=\sigma_n^2 \mathbf{I}_N}, \end{aligned} \quad (29b)$$

where  $\mathbf{H}$  has been pulled out of the expectation operation as it behaves as a constant scaling. We can make a few observations here. First, notice the last term in Eq. 29(a), second term, and third term in Eq. 29(b) will be zero as we assume  $\mathbf{N}$  is AWGN whose mean is zero,  $\mathbf{X}$  is just a scaling to the amplitude of the noise. Second, the term  $E[\mathbf{X} \mathbf{X}^T]$  in Eq. 29(b) becomes the signal power of  $\mathbf{X}$  times the identity matrix  $\mathbf{I}_x$ . If we assume each entry in  $\mathbf{X}$  is independent of each other, and the mean of each entry is zero, then  $E[\mathbf{X} \mathbf{X}^T] = \sigma_x^2 \mathbf{I}$  where  $\sigma_x^2$  represent the signal power. This goes to the first term of 29(a) similarly, where  $E[x_k \mathbf{X}^T]$  results in a vector whose only non-zero entry represents the main-cursor signal power. Third,  $E[\mathbf{N} \mathbf{N}^T]$  in Eq. 29(b) is the auto-correlation of the AWGN noise  $\mathbf{N}$ , which becomes the noise power  $\sigma_n^2$  times the identity matrix  $\mathbf{I}_N$ , since the mean of AWGN noise and each entry is independent with each other in  $\mathbf{N}$ . In the context of our modelling,  $\sigma_n^2$  is the AFE noise generated by the TIA  $\sigma_{TIA}^2$  in Eq. 18.

Thus, we can further simplify Eq. 29 based on the above observations and obtain the following:

$$\mathbf{R}_{x\mathbf{Y}} = \sigma_x^2 [\underbrace{0 0 \dots 0}_{a \text{ zeros}} \underbrace{1 0 0 \dots 0}_{b \text{ zeros}}] \mathbf{H}^T = \sigma_x^2 \mathbf{1}_a \mathbf{H}^T \quad (30a)$$

$$\mathbf{R}_{\mathbf{YY}} = \sigma_x^2 \mathbf{H} \mathbf{H}^T + \sigma_n^2 \mathbf{I}_N = \sigma_x^2 \mathbf{H} \mathbf{H}^T + \frac{1}{SNR} \mathbf{I}_N. \quad (30b)$$

Sometimes people like to use  $SNR = \frac{\sigma_x^2}{\sigma_n^2}$  to substitute the noise power  $\sigma_n^2$  in Eq. 30(b).  $\mathbf{1}_a$  is just a shorthand to express the row vector  $[\underbrace{0 0 \dots 0}_{a \text{ zeros}} \underbrace{1 0 0 \dots 0}_{b \text{ zeros}}]$ . In reality, the number of

pre-cursor coefficients  $a$  sets the delay of the received signal and can be a design variable. To see this, we plug Eq. 28 and 24 back to Eq 25(b), we obtain the final expression of the MSE:

$$\begin{aligned}
E[|e|^2] &= E[|x_k - z_k|^2] \\
&= E[(x_k - z_k)^T(x_k - z_k)] = E[(x_k - \mathbf{w}\mathbf{Y})^T(x_k - \mathbf{w}\mathbf{Y})] \\
&= E[x_k^T x_k - x_k^T \mathbf{w}\mathbf{Y} - \mathbf{Y}^T \mathbf{w}^T x_k + \mathbf{Y}^T \mathbf{w}^T \mathbf{w}\mathbf{Y}] \\
&= \sigma_x^2 - 2\mathbf{R}_{x\mathbf{Y}}\mathbf{w}^T + \mathbf{w}\mathbf{R}_{\mathbf{Y}\mathbf{Y}}\mathbf{w}^T \\
&= \sigma_x^2 - 2\mathbf{R}_{x\mathbf{Y}}(\mathbf{R}_{\mathbf{Y}\mathbf{Y}}^{-1})^T \mathbf{R}_{x\mathbf{Y}}^T \\
&\quad + \mathbf{R}_{x\mathbf{Y}} \underbrace{\mathbf{R}_{\mathbf{Y}\mathbf{Y}}^{-1} \mathbf{R}_{\mathbf{Y}\mathbf{Y}} (\mathbf{R}_{\mathbf{Y}\mathbf{Y}}^{-1})^T}_{=\mathbf{I}} \mathbf{R}_{x\mathbf{Y}}^{-1} \\
&= \sigma_x^2 - 2\mathbf{R}_{x\mathbf{Y}}(\mathbf{R}_{\mathbf{Y}\mathbf{Y}}^{-1})^T \mathbf{R}_{x\mathbf{Y}}^T + \mathbf{R}_{x\mathbf{Y}}(\mathbf{R}_{\mathbf{Y}\mathbf{Y}}^{-1})^T \mathbf{R}_{x\mathbf{Y}}^T \\
&= \sigma_x^2 - \mathbf{R}_{x\mathbf{Y}}(\mathbf{R}_{\mathbf{Y}\mathbf{Y}}^{-1})^T \mathbf{R}_{x\mathbf{Y}}^T \\
&= \sigma_x^2 - \mathbf{R}_{xy}\mathbf{w}^T. \tag{31}
\end{aligned}$$

Thus, the MSE is dependent on the tap delay  $a$  ( $\mathbf{R}_{x\mathbf{Y}}$  is dependent on  $\mathbf{1}_a$ ). One could sweep tap delay to find the optimal  $a$  for the smallest MSE. Interestingly, the MSE is independent with  $\mathbf{R}_{\mathbf{Y}\mathbf{Y}}$ .

When we let  $SNR = \infty$  in Eq. 30, which suggests no noise in the channel, MMSE degenerates to the zero-forcing (ZF) equalization. In reality, ZF enhances the high-frequency noise since it purely focuses on removing the ISI at the cost of ignoring noise. ZF behaves like a matched filter of the channel response, whereas MMSE does not amplify the high-frequency noise in contrast to ZF. Since MMSE strikes a balance between noise enhancement and ISI removal, plus its implementation is of the same complexity as ZF, MMSE is always preferred, and it performs equal or better than ZF.

The situation becomes slightly different when both FFE and DFE are used. People usually will use DFE to remove post-cursor ISI and FFE to remove pre-cursor ISI. However, if the post-cursor ISI has a very long tail, merely DFE equalization might be insufficient as its length is limited by power and area. FFE can still be used to remove post-cursor ISI after the DFE taps. Under this scenario, a common misconception is that one may just simply zero out the entries in the channel response matrix  $\mathbf{H}$  that correspond to the post-cursors, which DFE will compensate, and then do the MMSE or ZF for FFE tap weights calculation. We will show that FFE and DFE need to be jointly optimized using MMSE, which is named MMSE-DFE.

Recall Eq. 24 where vector  $\mathbf{Y}$  is equalized by the FFE only. When both FFE and DFE are employed, as shown in Fig. 2, we can write the new equalized output  $z_k$  as:

$$z_k = \mathbf{w}\mathbf{Y} - \mathbf{b}\hat{\mathbf{Y}}, \tag{32}$$

where  $\mathbf{b}$  denotes the DFE coefficients as a row vector (for simpler notation of avoiding transpose) and  $\hat{\mathbf{Y}}$  is the feedback signal as a column vector after the slicer. So, DFE behaves similarly to an FIR filter, except the slicer quantizes its feedback signals. Now we can plug Eq. 32 back to Eq. 25 and define the MSE as:

$$E[|e|^2] = E[|x_k - z_k|^2] = E[|x_k - (\mathbf{w}\mathbf{Y} - \mathbf{b}\hat{\mathbf{Y}})|^2]. \tag{33}$$

We would like to find both  $\mathbf{w}$  and  $\mathbf{b}$  such that  $E[|e|^2]$  is minimized. It will be easier to do the derivative afterward if we can lump both  $\mathbf{w}$  and  $\mathbf{b}$  together and declare it as a new matrix  $\mathbf{W}$ , and put  $\mathbf{Y}$  and  $\hat{\mathbf{Y}}$  together as  $\tilde{\mathbf{Y}}$ :

$$\mathbf{W} = [\mathbf{w} \ -\mathbf{b}] \tag{34a}$$

$$\tilde{\mathbf{Y}} = [\mathbf{Y} \ \hat{\mathbf{Y}}]^T. \tag{34b}$$

Therefore, Eq. 33 becomes:

$$E[|e|^2] = E[|x_k - \mathbf{W}\tilde{\mathbf{Y}}|^2]. \tag{35}$$

which has the identical look as Eq. 25(b) except the variables are changed. Therefore, the derivation will be similar to MMSE, and the expression of tap weights  $\mathbf{W}$  should be the same as Eq. 28 except the change of variables:

$$\mathbf{W} = \mathbf{R}_{x\tilde{\mathbf{Y}}} \mathbf{R}_{\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}}^{-1}. \tag{36}$$

where  $\mathbf{R}_{x\tilde{\mathbf{Y}}}$  is similar to Eq. 29(a) defined as the cross-correlation of  $x_k$  and  $\tilde{\mathbf{Y}}$ , and  $\mathbf{R}_{\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}}$  is similar to Eq. 29(b) defined as the auto-correlation of  $\tilde{\mathbf{Y}}$ . If we assume each entry in  $\tilde{\mathbf{Y}}$  is independent of each other whose mean is zero, and  $x_k$  is independent of  $\tilde{\mathbf{Y}}$ :

$$\mathbf{R}_{x\tilde{\mathbf{Y}}} = E[x_k \tilde{\mathbf{Y}}^T] \tag{37a}$$

$$\begin{aligned}
&= [E[x_k \mathbf{Y}] \ E[x_k \hat{\mathbf{Y}}]] = \underbrace{[E[x_k (\mathbf{H}\mathbf{X} + \mathbf{N})^T]]}_{\mathbf{R}_{x\mathbf{Y}} \text{ from Eq. 29(a)}} \underbrace{[E[x_k \hat{\mathbf{Y}}]]}_{\mathbf{0}} \\
&= [\mathbf{R}_{xy} \ \mathbf{0}] = [\sigma_x^2 \mathbf{1}_a \mathbf{H}^T \ \mathbf{0}] \tag{37b}
\end{aligned}$$

$$\mathbf{R}_{\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}} = E[\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^T] = E[[\mathbf{Y} \ \hat{\mathbf{Y}}]^T [\mathbf{Y}^T \ \hat{\mathbf{Y}}^T]] = \tag{37c}$$

$$\begin{aligned}
&= \left[ \begin{array}{cc} \underbrace{E[\mathbf{Y}\mathbf{Y}^T]}_{\mathbf{R}_{\mathbf{Y}\mathbf{Y}} \text{ from Eq. 29(b)}} & E[\mathbf{Y}\hat{\mathbf{Y}}^T] \\ E[\hat{\mathbf{Y}}\mathbf{T}^T] & E[\hat{\mathbf{Y}}\hat{\mathbf{Y}}^T] \end{array} \right] \\
&= \left[ \begin{array}{cc} \mathbf{R}_{\mathbf{Y}\mathbf{Y}} & E[\mathbf{H}\mathbf{X}\hat{\mathbf{Y}}^T] + \underbrace{E[\mathbf{N}\hat{\mathbf{Y}}^T]}_{=0} \\ E[\hat{\mathbf{Y}}\mathbf{X}^T \mathbf{H}^T] + \underbrace{E[\hat{\mathbf{Y}}\mathbf{N}^T]}_{=0} & \sigma_x^2 \mathbf{I}_{N_b} \end{array} \right] \\
&= \left[ \begin{array}{cc} \sigma_x^2 \mathbf{H}\mathbf{H}^T + \frac{1}{SNR} \mathbf{I}_N & \sigma_x^2 \mathbf{H}\mathbf{J} \\ \sigma_x^2 \mathbf{J}^T \mathbf{H}^T & \sigma_x^2 \mathbf{I}_{N_b} \end{array} \right] \tag{37d}
\end{aligned}$$

$$\mathbf{J} = \begin{bmatrix} \mathbf{0}_{(a+1) \times N_b} \\ \mathbf{I}_{N_b \times N_b} \end{bmatrix}, \tag{37e}$$

where  $N_b$  is the number of DFE taps. The matrix  $\mathbf{J}$  is introduced here for mathematical convenience.  $J$  consists of a  $a+1$  by  $N_b$  sub-matrix of zeros, which basically says the pre-cursor and main cursor of  $X$  and DFE feedback signal  $\hat{\mathbf{Y}}$

does not correlate. The identity matrix has a dimension of  $N_b$  by  $N_b$ , which suggests that DFE is only applied to the post-cursor of  $\mathbf{X}$ . Finally, we can plug Eq. 37a back to Eq. 34a and obtain the following:

$$\begin{aligned} \mathbf{W} &= \mathbf{R}_{\mathbf{x}\tilde{\mathbf{Y}}} \mathbf{R}_{\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}}^{-1} \\ [\mathbf{w} \cdot \mathbf{b}] &= [\sigma_x^2 \mathbf{1}_a \mathbf{H}^T \mathbf{0}] \begin{bmatrix} \sigma_x^2 \mathbf{H} \mathbf{H}^T + \frac{1}{SNR} \mathbf{I}_N & \sigma_x^2 \mathbf{H} \mathbf{J} \\ \sigma_x^2 \mathbf{J}^T \mathbf{H}^T & \sigma_x^2 \mathbf{I}_{N_b} \end{bmatrix}^{-1} \\ [\mathbf{w} \cdot \mathbf{b}] \begin{bmatrix} \sigma_x^2 \mathbf{H} \mathbf{H}^T + \frac{1}{SNR} \mathbf{I}_N & \sigma_x^2 \mathbf{H} \mathbf{J} \\ \sigma_x^2 \mathbf{J}^T \mathbf{H}^T & \sigma_x^2 \mathbf{I}_{N_b} \end{bmatrix} &= [\sigma_x^2 \mathbf{1}_a \mathbf{H}^T \mathbf{0}], \end{aligned} \quad (38a)$$

which can be further expanded as the following:

$$\begin{aligned} \sigma_x^2 \mathbf{w} \mathbf{H} \mathbf{J} - \sigma_x^2 \mathbf{b} \mathbf{I}_{N_b} &= \mathbf{0} \\ \mathbf{b} &= \mathbf{w} \mathbf{H} \mathbf{J} \end{aligned} \quad (39a)$$

$$\begin{aligned} \mathbf{w} (\sigma_x^2 \mathbf{H} \mathbf{H}^T + \frac{1}{SNR} \mathbf{I}_N) - \sigma_x^2 \mathbf{b} \mathbf{J}^T \mathbf{H}^T &= \sigma_x^2 \mathbf{1}_a \mathbf{H}^T \\ \mathbf{w} (\mathbf{H} \mathbf{H}^T + \frac{1}{SNR} \mathbf{I}_N) - \mathbf{w} \mathbf{H} \mathbf{J} \mathbf{J}^T \mathbf{H}^T &= \mathbf{1}_a \mathbf{H}^T \\ \mathbf{w} (\mathbf{H} \mathbf{H}^T + \frac{1}{SNR} \mathbf{I}_N - \mathbf{H} \mathbf{J} \mathbf{J}^T \mathbf{H}^T) &= \mathbf{1}_a \mathbf{H}^T. \end{aligned} \quad (39b)$$

We finally arrive at the expression for  $\mathbf{w}$ , the tap coefficients of FFE:

$$\mathbf{w} = \mathbf{1}_a \mathbf{H}^T (\mathbf{H} \mathbf{H}^T + \frac{1}{SNR} \mathbf{I}_N - \mathbf{H} \mathbf{J} \mathbf{J}^T \mathbf{H}^T)^{-1}, \quad (40)$$

and the expression of the tap coefficients of DFE  $\mathbf{b}$ , by substitute Eq. 40 back to Eq. 39(a):

$$\mathbf{b} = \mathbf{1}_a \mathbf{H}^T (\mathbf{H} \mathbf{H}^T + \frac{1}{SNR} \mathbf{I}_N - \mathbf{H} \mathbf{J} \mathbf{J}^T \mathbf{H}^T)^{-1} \mathbf{H} \mathbf{J}. \quad (41)$$

#### IV. OPTIMIZATION FORMULATION

To optimize the interface, signal integrity criteria that reflect the quality of the signal at the receiver's output and take impairments such as reflections, jitter, and noise into account has to be defined. Therefore, we define a signal integrity figure of merit (FoM) that can be calculated statistically and correlates with BER. Statistical analysis of high-speed serial links provides an efficient way to evaluate performance since it relies on the pulse response of the channel rather than relying on a large amount of randomly generated bit patterns. A statistically calculated FoM with the proposed optimization approach allows it to be calculated rapidly, allowing for faster convergence on optimal design.

Specifically, the FoM is defined as:

$$FoM = 10 \log_{10} \frac{A_{signal}^2}{\sigma_{ISI}^2 + \sigma_{n,FFE}^2 + \sigma_{jitter}^2}, \quad (42)$$

where  $\sigma_{ISI}$  represents the residual ISI at the output of the equalizer,  $\sigma_{n,FFE}$  is the RMS voltage noise at the output of FFE, and  $\sigma_{jitter}$  represents the RMS jitter-to-amplitude voltage conversion. The term  $A_{signal}$  is calculated from the pulse response as follows: assuming the equalized pulse response  $h_{pulse,eq}$  is baud rate sampled with  $O$  samples,

and that the index of the main (max) cursor is zero, then  $A_{signal} = 1/3 \times h_{pulse}[0]$  for PAM-4 signalling. The peak of the pulse response,  $A_{signal}$ , represents the peak-to-peak amplitude of the modulated and equalized signal. This FoM is a signal-to-noise ratio, and a higher FoM corresponds to a better bit error rate.

The residual ISI power,  $\sigma_{ISI}^2$ , is also calculated from the pulse response:

$$\sigma_{ISI}^2 = \frac{5}{9} \sum_{n \neq 0}^O h_{pulse,eq}^2[n], \quad (43)$$

where the factor 5/9 takes into account the differing ISI contributed by different 4-PAM symbol amplitudes, obtained by averaging the power of four signal levels as  $((-1)^2 + (-\frac{1}{3})^2 + (\frac{1}{3})^2 + 1^2)/4 = \frac{5}{9}$ .

Noise variance at the output of the equalizer is calculated using the calculated FFE tap weights,  $w_0, w_1, \dots, w_M$ , which has been given by Eq. 20 and 21. Again, DFE is noiseless since the feedback signals are digitized.

Jitter causes eye height to fluctuate around the sampling point. Thus, jitter translates into amplitude noise, reducing signal integrity. In other words, when the signal is jittery, the location of the peak of the signal changes with respect to the sampling time. If the sampling phase is fixed, the signal will be sampled off-peak when there is jitter. This leads to eye height degradation. The jitter-to-amplitude conversion variance,  $\sigma_{jitter}^2$  is [28]:

$$\sigma_{jitter}^2 = \frac{5}{9} \sigma_j^2 \sum_n^O \mu[n]^2, \quad (44)$$

where  $\sigma_j$  is the RMS jitter, and  $\mu$  is the slope of equalized pulse response at the sampling points, which can be found by taking the derivative of the sampled signal with respect to its time. The value of  $\sigma_j$  is assumed to be 0.015 UI. Similar to Eq. 43, the factor 5/9 here accounts for the density of 4-PAM transitions. The amount of eye height variation equals the amount of time variation times the slope. The quantities are squared to make them power quantities.

We use the genetic algorithm (GA) with the flow summarized in Algorithm 1 for this optimization. In this flow, an initial population of 200 sets of design parameters ( $TW$ ,  $W, L$ ,  $N_{in}$ ,  $N_{out}$ ,  $g_m$ , and  $R_F$ ) are randomly generated. Pulse responses  $h_{pulse}$  and noise variances at the output of the TIA  $\sigma_{n,TIA}^2$  are calculated for each set of design parameters. Using this information, the FFE and DFE tap weights are calculated using MMSE-DFE algorithm. The equalized pulse responses can then be calculated using the tap weights. Noise variances are also referred to as the output of the equalizer denoted  $\sigma_{n,FFE}$ . The FoM for each design parameter set is calculated. A new generation is created by selecting, crossover, and mutation of the best current design parameter sets. The process repeats until we hit the maximum allowable generation number. We also used the same seed for optimization to ensure repeatable results. Finally, the parameter values set corresponding to the best achieved FoM is selected as the

**Algorithm 1** Optical receiver optimization using GA

**Require:** Initialize population of randomly generated optical receiver design sets  $[TW \ W \ L \ N_{in} \ N_{out} \ g_m \ R_F]$ , GA total generation number  $G$ , GA current generation number  $i$ .

- 0:  $G \leftarrow 100$
- 0: **while**  $i < G$  **do**
- 0:    $i \leftarrow i + 1$
- 0:   Calculate pulse response  $h_{pulse}$  by Eq. 11 and TIA noise  $\sigma_{TIA}$  by Eq. 18 for each set
- 0:   Calculate FFE tap weights  $w$  by Eq. 40 and DFE tap weights  $b$  by Eq. 41
- 0:   Equalize  $h_{pulse}$  by  $w$  and  $b$
- 0:   Calculate  $\sigma_{ISI}$ ,  $\sigma_{n,FFE}$ , and  $\sigma_{jitter}$  using Eq. 43, 20, and 44, respectively
- 0:   Calculate FoM of each set using Eq. 42
- 0:   Create the next generation with a population of 100 by selecting the elite design sets and doing mutation and crossover
- 0: **end while**
- 0: Pick the design set with the best  $FoM = 0$

optimal design. In this optimization scheme, the power consumption can be controlled by limiting the range of  $g_m$  and the number of equalizer taps. To ensure the practical utility of the optimizer, it is necessary to model each component in the link accurately. Foundry-provided models are used for the photodetector and to train the T-coil modelling agent. The overall optimization process takes around 29 minutes on a computer with the following specifications: Intel Core i7-8750H @ 2.20 GHz CPU, 2666 MHz 16 GB SDRAM, and a 256 GB PCIe SSD.

## V. OPTIMIZATION RESULTS

The design was optimized for three transmission line lengths:  $250 \mu m$ ,  $500 \mu m$ , and  $5 mm$ . Table III shows optimal design values, assuming 32 FFE taps and 4 DFE taps, and the corresponding FoM. Table IV shows the results assuming 6 FFE taps and 2 DFE taps, while Table V shows results with no equalization. We note that while the optimal transistor sizes in Table III were the largest permitted in our analysis for all three cases considered here (corresponding to largest  $g_m$ ), this was not the case in trials where there were fewer taps of equalization (Table IV and Table V). The likely reason why  $g_m$  is optimal is that when  $g_m$  value is high, the value of  $R_a$  is low, resulting in a high output frequency pole and allowing for a high  $R_F$ , which results in a higher gain. Although large  $g_m$  results in a larger  $C_g$  lowering the input pole, the T-coil offsets this negative impact. Thus, a large  $g_m$  is more favourable overall. Fig. 13 shows the eye diagrams obtained with this optimization (Table III), including impairments. As these figures show, there is a good eye opening in all three cases. The eye diagram also shows a contour corresponding to a  $BER = 2.4 \times 10^{-4}$ . This validates that the proposed optimization approach converges on designs with a good eye opening and low BER.

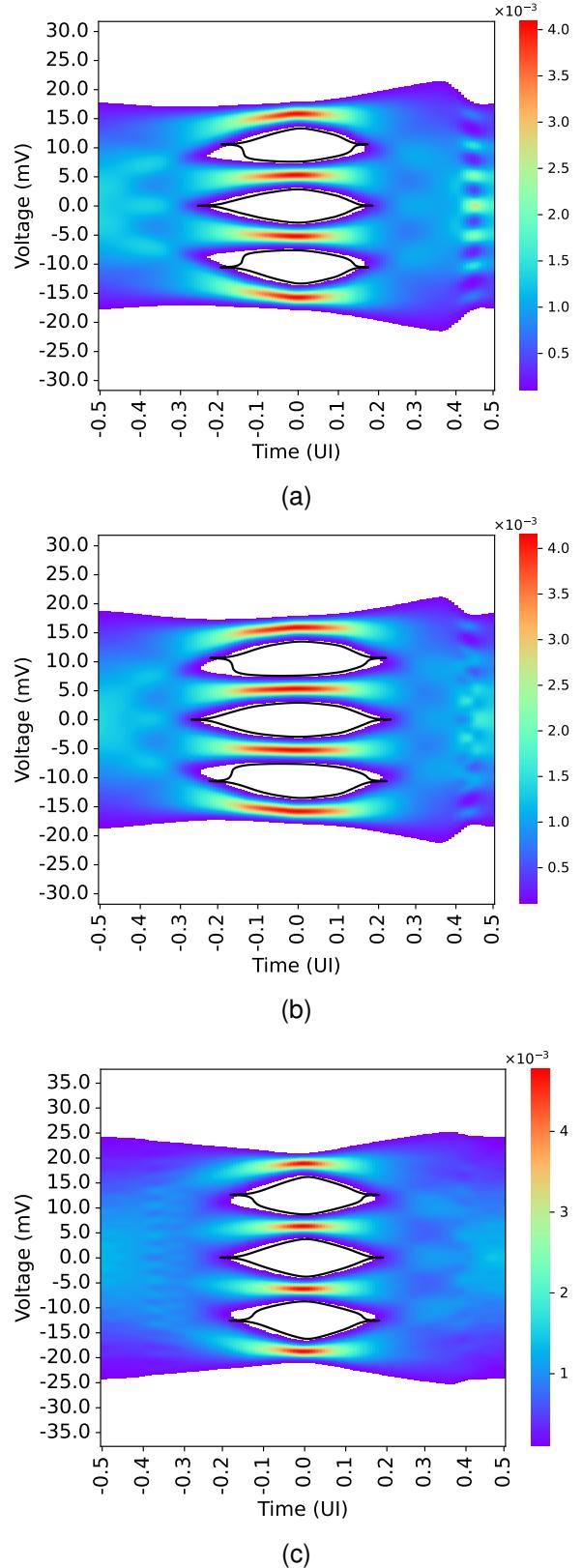


Fig. 13. Eye diagrams at the output of the equalizer for different length transmission lines. They are obtained by using the optimal design values assuming 32 FFE taps and 4 DFE taps. (a)  $TL = 250 \mu m$  (b)  $TL = 500 \mu m$ , and (c)  $TL = 5 mm$ . The contour shown corresponds to a  $BER = 2.4 \times 10^{-4}$ .

TABLE III  
OPTIMIZATION RESULTS ASSUMING 32 FFE TAPS AND 4 DFE TAPS

$TL(\mu m)$	$TW(\mu m)$	$L(\mu m)$	$W(\mu m)$	$N_{IN}$	$N_{OUT}$	$g_m(mS)$	$R_F(\Omega)$	$\sigma_{ISI}(mV_{rms})$	$\sigma_{n,output}(mV_{rms})$	$\sigma_{jitter}(mV_{rms})$	$FoM(dB)$
250	15	43	4.2	11	6	120	2514	0.257	0.738	0.198	22.34
500	15	44	4.2	11	5	120	2508	0.236	0.741	0.125	22.57
5000	15	43	4.2	11	6	120	1143	0.226	0.741	0.37	23.32

TABLE IV  
OPTIMIZATION RESULTS ASSUMING 6 FFE TAPS AND 2 DFE TAPS

$TL(\mu m)$	$TW(\mu m)$	$L(\mu m)$	$W(\mu m)$	$N_{IN}$	$N_{OUT}$	$g_m(mS)$	$R_F(\Omega)$	$\sigma_{ISI}(mV_{rms})$	$\sigma_{n,FFE}(mV_{rms})$	$\sigma_{jitter}(mV_{rms})$	$FoM(dB)$
250	15	42	4.2	12	7	114	1762	0.297	0.755	0.2	21.89
500	15	34	2.4	10	5	116	2215	0.317	0.762	0.128	22.06
5000	80	37	5	18	6	108	609	0.382	0.728	0.228	19.67

TABLE V  
OPTIMIZATION RESULTS ASSUMING NO EQUALIZATION

$TL(\mu m)$	$TW(\mu m)$	$L(\mu m)$	$W(\mu m)$	$N_{IN}$	$N_{OUT}$	$g_m(mS)$	$R_F(\Omega)$	$\sigma_{ISI}(mV_{rms})$	$\sigma_{n,FFE}(mV_{rms})$	$\sigma_{jitter}(mV_{rms})$	$FoM(dB)$
250	70	43	4.2	10	7	112	176	0.219	0.604	0.135	16.99
500	15	49	5	11	5	104	191	0.215	0.63	0.089	17.50
5000	100	32	4.2	6	11	80	142	0.649	0.697	0.069	12.38

Fig. 14 shows a plot of the number of FFE taps versus the value of FoM for all three lengths of transmission lines. As can be seen, FoM increases steadily with the number of FFE taps. Moreover, we notice that the first few taps significantly improve FoM with diminishing returns as the number of FFE taps increases beyond about 6. This is particularly true for the long 5 mm transmission line, which benefits significantly from a few equalization taps. With sufficient equalization, the FoM for the 5 mm transmission line is on par with the 250  $\mu m$  and 500  $\mu m$  transmission lines.

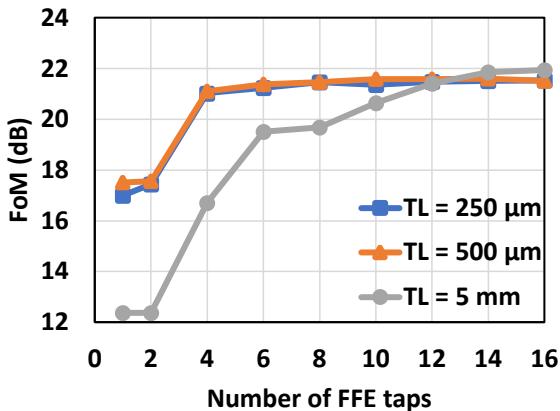


Fig. 14. FoM versus the number of FFE taps for different length transmission lines.

In addition to optimizing designs, the proposed approach can be used to gain insight into optimal design guidelines. Here, we explore the relationship between the optimal transmission line width,  $TW$ , which controls the characteristic impedance of the transmission line, and the amount of available equalization in the case of the long 5 mm transmission line that can exhibit strong reflections.

We use the optimization platform to obtain optimal design values for various numbers of FFE taps with no DFE taps. For instance, Fig. 15 shows the optimal transmission line

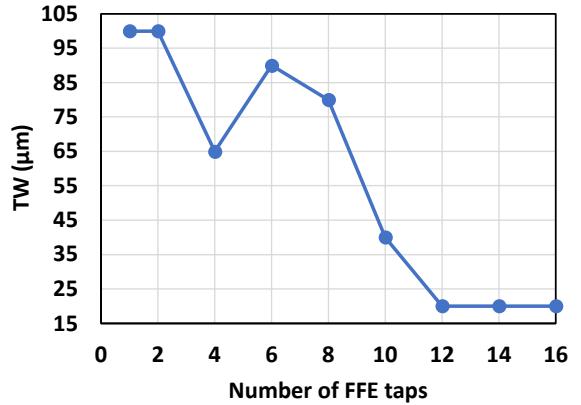


Fig. 15. Optimal  $TW$  versus the number of FFE taps for the  $TL = 5 \text{ mm}$  transmission line.

width versus the number of FFE taps for the  $TL = 5 \text{ mm}$  transmission line. With little or no equalization, the optimal transmission line width is wide and tends to narrow with increasing FFE taps. To explain this behaviour, we look at the pulse response  $h_{pulse}$  of the  $TL = 5 \text{ mm}$  transmission line shown in Fig. 16 in two scenarios: without equalization and with 16 FFE taps, where for each case we use optimal AFE design parameters. With no equalization, the pulse response shown in Fig. 16(a) shows little to no reflections. In this case, a wide transmission line is preferred to avoid reflections manifesting as ISI. In other words, the optimizer chooses to achieve impedance matching between the transmission line and the input of the TIA to avoid reflections. To prove this, we inspected the input impedance of the TIA and compared it to the transmission line's characteristic impedance. The value of the transmission line's characteristic impedance is around  $37 \Omega$ , while the input impedance value is  $32 \Omega$ , confirming the close matching.

On the other hand, a narrow transmission line is preferred in the case of 16 FFE taps. To explain this, we look at

the pulse response at the output of the TIA shown in Fig. 16 (b). Here, we see many reflections due to a significant impedance mismatch between the characteristic impedance of the transmission line and the input of the TIA. However, when inspecting the pulse response at the output of the FFE (Fig. 16 (c)), we see that reflections are significantly reduced, particularly at the sampling points. This makes it unnecessary to do the impedance matching since the FFE is taking care of the reflections. The optimizer chooses a narrow transmission line, likely to reduce its introduced capacitance at the chip's input.

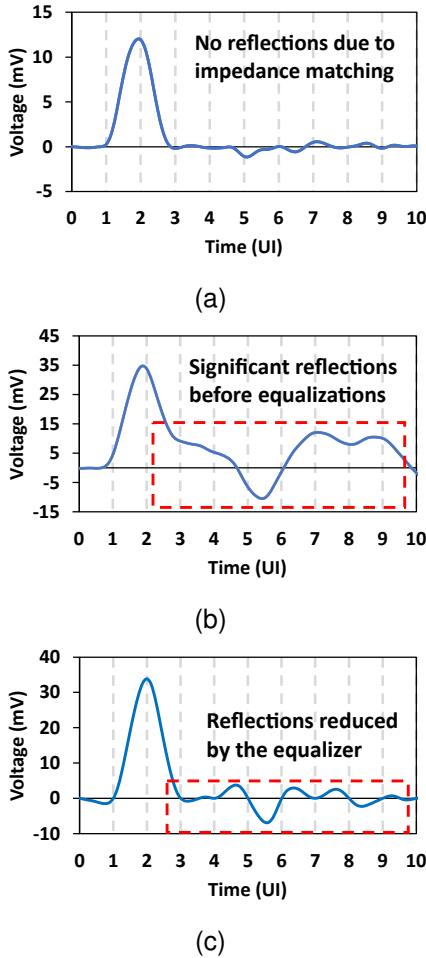


Fig. 16. Pulse responses for  $TL = 5\text{ mm}$ . (a) Pulse response was obtained with optimal design values, assuming no equalization. (b) Pulse responses at the output of the TIA are obtained with optimal design values, assuming there is a 16-tap FFE. (c) Pulse responses at the output of the FFE equalizer, obtained with optimal design values assuming there is a 16-tap FFE.

Based on the preceding analysis, narrow transmission lines are preferred with sufficient equalization, along with a lower bandwidth, lower noise, and higher gain front end. Such a design affords a lower power consumption in the AFE but higher power in the DSP equalizer. With less equalization, a wider transmission line is preferable to ensure smaller reflections and better signal integrity. Note that the pitch of neighbouring receiver lanes is typically limited by practical considerations such as the pitch of mating fibre arrays, typically 100's of  $\mu\text{m}$ , and is unaffected by trace width optimizations. Of course,

the optimization could be constrained to accommodate narrow channel pitches, especially if and as required. Therefore, we conclude the following design guideline: sufficient equalization to cancel reflections results in a narrow transmission line for the optimal design; otherwise, impedance matching is needed, which requires a wider transmission line. These simulations highlight the importance of equalization in counteracting reflections.

## VI. EXPERIMENTAL VALIDATION

This section presents measurement results that illustrate the trends and tradeoffs elucidated by the automated optimization approach. This is courtesy of Dhruv Patel for generously providing measuring results from his TIA work published in [4]. Measurements were performed on a TIA prototype fabricated in 16nm FinFET CMOS and flip-chip co-packaged along with commercial PDs. Two co-packaging arrangements were optimized for the same TIA, as shown in Fig. 17 with  $TL = 250\text{ }\mu\text{m}$  and  $TL = 500\text{ }\mu\text{m}$ . The complete front-end design details are presented in [4], [29].

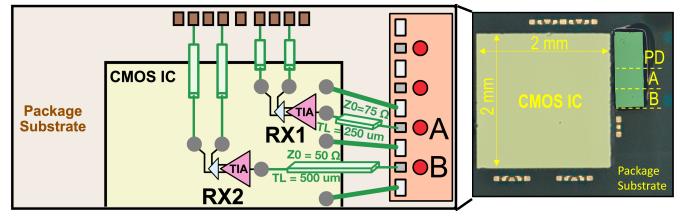


Fig. 17. Co-packaged prototype with TIA in 16-nm FinFET CMOS co-packaged with arrayed PD with  $TL = 250\text{ }\mu\text{m}$  and  $TL = 500\text{ }\mu\text{m}$ .

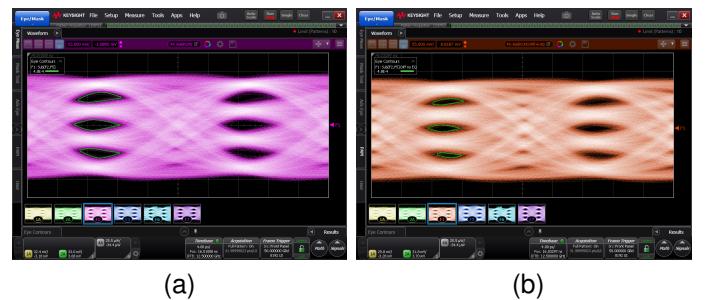


Fig. 18. Measured eye diagrams at 100 Gb/s 4-PAM with (a)  $TL = 250\text{ }\mu\text{m}$ ,  $TW = 22\text{ }\mu\text{m}$  and  $Z_0 = 75\Omega$  (b)  $TL = 500\text{ }\mu\text{m}$ ,  $TW = 60\text{ }\mu\text{m}$  and  $Z_0 = 50\Omega$ .

Although the prototype TIA's design parameters differ somewhat from the simulation model presented in Fig. 2, the same trends and tradeoffs are evident in the measured results. As predicted by the ML-assisted genetic optimizer in this work, the optimized interconnect is wider with  $TW = 60\text{ }\mu\text{m}$  and a characteristic impedance  $Z_0 = 50\Omega$  for the longer trace and narrower with  $TW = 22\text{ }\mu\text{m}$  and a higher characteristic impedance  $Z_0 = 75\Omega$  for the shorter trace. This allows both co-packaging arrangements to maintain comparable 4-PAM signal integrity up to 100 Gb/s, as illustrated by the unequalized TIA output eye diagrams in Fig. 18.

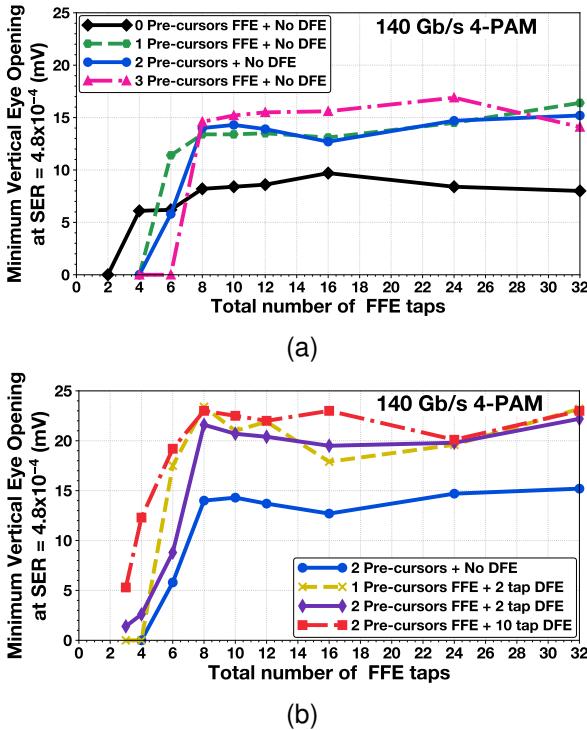


Fig. 19. Measured vertical eye opening at 140 Gb/s 4-PAM with  $TL = 250 \mu\text{m}$  (a) FFE only (b) FFE + DFE.

Furthermore, as in the ML-assisted genetic optimization, we see a dramatic improvement in signal integrity (quantified by the vertical eye opening after equalization measured on the oscilloscope) once the span of the equalizers is sufficient to compensate for reflections and ringing induced in the package. Results incorporating FFE and DFE equalizers and varying the number of taps are shown in Fig. 19 at 140 Gb/s. An 8-tap FFE with one pre-cursor tap equalizes the package-induced ISI and TIA bandwidth limitations, with additional taps providing little benefit. Including a 2-tap DFE provides a noticeable improvement, with little benefit from increasing the DFE length to 10 taps. The TIA has 32 GHz bandwidth, 45% of the baud rate in these experiments, comparable to the ML-assisted genetic optimization results.

## VII. CONCLUSION

The paper presented the optimization of the interface between the PD and the analog front-end in high-speed, high-density optical receivers. We used the proposed framework to optimize transmission line width, the geometry of the T-coil, the inverter-based TIA, and FFE and DFE tap weights. We have applied a hybrid modelling methodology consisting of analytical models, an electromagnetic simulation, and a NN model to describe the interface and effectively optimize parameters. The framework is also used to draw insight into optimal design practices. For example, we have shown trends highlighting the relationship between the amount of equalization and the width of the transmission line. We showed that narrow transmission lines are favoured when there is enough equalization. However, it should be noted that this could lead

to high power consumption because of the increased number of taps required to counteract reflections. Therefore, a wider transmission line may be favoured in power-efficient designs with limited equalization. These trends are further validated with measurements performed on a fabricated and assembled TIA prototype with various PD-to-TIA interface lengths at 100 Gb/s.

## ACKNOWLEDGEMENT

The authors would like to thank Dr. Hossein Shakiba from Huawei Technologies for his valuable discussions throughout this project.

## REFERENCES

- [1] H. Li, C.-M. Hsu, J. Sharma, J. Jaussi, and G. Balamurugan, "A 100-Gb/s PAM-4 Optical Receiver With 2-Tap FFE and 2-Tap Direct-Feedback DFE in 28-nm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 57, no. 1, pp. 44–53, 2022.
- [2] K. R. Lakshmkumar, A. Kurylak, M. Nagaraju, R. Booth, R. K. Nandwana, J. Pampanin, and V. Bocuzzi, "A Process and Temperature Insensitive CMOS Linear TIA for 100 Gb/s/λ PAM-4 Optical Links," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 11, pp. 3180–3190, 2019.
- [3] H. Li, G. Balamurugan, J. Jaussi, and B. Casper, "A 112 Gb/s PAM4 Linear TIA with 0.96 pJ/bit Energy Efficiency in 28 nm CMOS," in *ESSCIRC 2018 - IEEE 44th European Solid State Circuits Conference (ESSCIRC)*, pp. 238–241, 2018.
- [4] D. Patel, A. Sharif-Bakhtiar, and T. C. Carusone, "A 112-gb/s —8.2-dbm sensitivity 4-pam linear tia in 16-nm cmos with co-packaged photodiodes," *IEEE Journal of Solid-State Circuits*, vol. 58, no. 3, pp. 771–784, 2023.
- [5] F.-P. Chou, G.-Y. Chen, C.-W. Wang, Y.-C. Liu, W.-K. Huang, and Y.-M. Hsin, "Silicon Photodiodes in Standard CMOS Technology," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 17, no. 3, pp. 730–740, 2011.
- [6] Z. Sheng, L. Liu, J. Brouckaert, S. He, and D. V. Thourhout, "InGaAs PIN photodetectors integrated on silicon-on-insulator waveguides," *Opt. Express*, vol. 18, pp. 1756–1761, Jan 2010.
- [7] S. Song and Y. Sui, "System level optimization for high-speed serdes: Background and the road towards machine learning assisted design frameworks," *Electronics*, vol. 8, no. 11, 2019.
- [8] S. Katare, "Novel framework for modelling high speed interface using python for architecture evaluation," in *2020 IEEE REGION 10 CONFERENCE (TENCON)*, pp. 556–560, 2020.
- [9] C. Xu, M. Lai, F. Lv, L. Xiao, Z. Luo, X. Qi, and Y. Ou, "A low ber adaptive sequence detection method for high-speed nrz data transmission," in *2022 7th International Conference on Integrated Circuits and Microsystems (ICICM)*, pp. 692–697, 2022.
- [10] A. Manukovsky, Z. Khasidashvili, A. J. Norman, Y. Juniman, and R. Bloch, "Designcon 2019 machine learning applications for simulation and modeling of 56 and 112 gb serdes systems,"
- [11] D. Yang, Y. Gan, V. Telang, M. Valliappan, and F. S. Tang, "Designcon 2014 improving ibis-ami model accuracy: Model-to-model and model-to-lab correlation case studies,"
- [12] A. Novack, M. Gould, Y. Yang, Z. Xuan, M. Streshinsky, Y. Liu, G. Capellini, A. E.-J. Lim, G.-Q. Lo, T. Baehr-Jones, and M. Hochberg, "Germanium photodetector with 60 ghz bandwidth using inductive gain peaking," *Opt. Express*, vol. 21, pp. 28387–28393, Nov 2013.
- [13] B. Dehlaghi, N. Wary, and T. C. Carusone, "Ultra-Short-Reach Interconnects for Die-to-Die Links: Global Bandwidth Demands in Microcosm," *IEEE Solid-State Circuits Magazine*, vol. 11, no. 2, pp. 42–53, 2019.
- [14] S. Lee, W. Jung, C. Ma, D. Lee, Y. Jung, D. Lee, S. Han, E.-C. Ahn, Y. Shin, H. Lee, H. Lim, and I. Hwang, "Development of FCBGA substrate with low Dk/Df material based on automotive reliability conditions," in *2019 IEEE 21st Electronics Packaging Technology Conference (EPTC)*, pp. 271–275, 2019.
- [15] J. Kim, J.-K. Kim, B.-J. Lee, and D.-K. Jeong, "Design Optimization of On-Chip Inductive Peaking Structures for 0.13-μm CMOS 40-Gb/s Transmitter Circuits," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 56, no. 12, pp. 2544–2555, 2009.

- [16] B. Razavi, "The strongarm latch [a circuit for all seasons]," *IEEE Solid-State Circuits Magazine*, vol. 7, no. 2, pp. 12–17, 2015.
- [17] Z. Li and A. C. Carusone, "Design and optimization of t-coil-enhanced esd circuit with upsampling convolutional neural network," in 2022 *IEEE/MTT-S International Microwave Symposium - IMS 2022*, pp. 495–497, 2022.
- [18] H. M. Torun, H. Yu, N. Dasari, V. C. K. Chekuri, A. Singh, J. Kim, S. K. Lim, S. Mukhopadhyay, and M. Swaminathan, "A spectral convolutional net for co-optimization of integrated voltage regulators and embedded inductors," in 2019 *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, pp. 1–8, 2019.
- [19] M. Swaminathan, H. M. Torun, H. Yu, J. A. Hejase, and W. D. Becker, "Demystifying machine learning for signal and power integrity problems in packaging," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 10, no. 8, pp. 1276–1295, 2020.
- [20] S. Daneshgar, H. Li, T. Kim, and G. Balamurugan, "A 128 Gb/s PAM4 Linear TIA with 12.6 pA/ $\sqrt{\text{Hz}}$  Noise Density in 22nm FinFET CMOS," in 2021 *IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, pp. 135–138, 2021.
- [21] K.-L. Fu and S.-I. Liu, "A 64-Gb/s PAM-4 Optical Receiver With Amplitude/Phase Correction and Threshold Voltage/Data Level Calibration," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 28, no. 7, pp. 1726–1735, 2020.
- [22] L. Szilagyi, J. Pliva, R. Henker, D. Schoeniger, J. P. Turkiewicz, and F. Ellinger, "A 53-Gbit/s Optical Receiver Frontend With 0.65 pJ/bit in 28-nm Bulk-CMOS," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 3, pp. 845–855, 2019.
- [23] I. Ozkaya, A. Cebrero, 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 64-Gb/s 1.4-pJ/b NRZ Optical Receiver Data-Path in 14-nm CMOS FinFET," *IEEE Journal of Solid-State Circuits*, vol. 52, no. 12, pp. 3458–3473, 2017.
- [24] B. Radi, D. Abdelrahman, O. Liboiron-Ladouceur, G. Cowan, and T. C. Carusone, "Optimal optical receivers in nanoscale cmos: A tutorial," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 69, no. 6, pp. 2604–2609, 2022.
- [25] D. M. Pozar, *Microwave Engineering*. John Wiley & Sons, 2012.
- [26] J. Bailey, H. Shakiba, E. Nir, G. Marderfeld, P. Krotnev, M.-A. LaCroix, D. Cassan, and D. Tonietto, "A 112-gb/s pam-4 low-power nine-tap sliding-block dfe in a 7-nm finfet wireline receiver," *IEEE Journal of Solid-State Circuits*, vol. 57, no. 1, pp. 32–43, 2022.
- [27] R. Payne, P. Landman, B. Bhakta, S. Ramaswamy, S. Wu, J. Powers, M. Erdogan, A.-L. Yee, R. Gu, L. Wu, Y. Xie, B. Parthasarathy, K. Brouse, W. Mohammed, K. Heragu, V. Gupta, L. Dyson, and W. Lee, "A 6.25-gb/s binary transceiver in 0.13-spl mu/m cmos for serial data transmission across high loss legacy backplane channels," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 12, pp. 2646–2657, 2005.
- [28] "Measuring channel operating margin." <https://dlcdn-anritsu.com/en-us/test-measurement/files/Technical-Notes/White-Paper/11410-00989A.pdf>, 2016. Accessed: April 18, 2023.
- [29] D. Patel, A. Sharif-Bakhtiar, and A. C. Carusone, "A 112 Gb/s -8.2 dBm Sensitivity 4-PAM Linear TIA in 16nm CMOS with Co-Packaged Photodiodes," in 2022 *IEEE Custom Integrated Circuits Conference (CICC)*, pp. 1–2, 2022.



**Bahaa Radi** (Member, IEEE) received the B.S. degree in electrical engineering from The Hashemite University, Zarqa, Jordan, in 2012, the M.S. degree in microsystems engineering from Khalifa University, Abu Dhabi, United Arab Emirates, in 2015, and the Ph.D. degree in electrical engineering from McGill University, Montreal, QC, Canada, in 2021. He is currently a Postdoctoral Fellow with the Integrated Systems Laboratory, Department of Electrical and Computer Engineering, University of Toronto, Canada. His current research interests include the design and optimization of energy-efficient optical systems for communication and computing applications, and the co-design of electronic and photonic integrated circuits.



**Zonghao Li** (Student Member, IEEE) Zonghao Li received his B.A.Sc. degree from the University of British Columbia in 2017 and his M.Eng. degree from McGill University in 2019, both in electrical engineering. Currently, he is pursuing a Ph.D. degree with the Edward S. Rogers Sr. Department of Electrical & Computer Engineering, University of Toronto. His current research interests focus on applying machine learning techniques to analog integrated circuit design and system-level high-speed SerDes modeling.



**Dhruv Patel** received BASc. and MASc. degrees in Electrical Engineering from the University of Waterloo and the University of Toronto in 2016 and 2020, respectively. He was involved with variation tolerant sub-threshold SRAM circuits research during undergraduate studies. Since 2019, he is pursuing his PhD. at the University of Toronto with Integrated Systems Laboratory working towards Optical Communication Links in CMOS. He was the recipient of the outstanding student paper award at Custom Integrated Circuits Conference 2022. He has received Ontario Graduate Scholarship and NSERC scholarship for his doctoral studies.



**Anthony Chan Carusone** (Fellow, IEEE) received his Ph.D. from the University of Toronto in 2002 and has since been a professor with the Department of Electrical and Computer Engineering at the University of Toronto. He has also been a consultant to industry in the areas of integrated circuit design and digital communication since 1997. He is currently the Chief Technology Officer of Alphawave IP Group in Toronto, Canada. Prof. Chan Carusone co-authored Best Student Papers at the 2007, 2008, 2011, and 2022 Custom Integrated Circuits Conferences, Best Invited Paper at the 2010 Custom Integrated Circuits Conference, Best Paper at the 2005 Compound Semiconductor Integrated Circuits Symposium, Best Young Scientist Paper at the 2014 European Solid-State Circuits Conference, and Best Paper at DesignCon 2021. He co-authored the popular textbooks "Analog Integrated Circuit Design" (along with D. Johns and K. Martin) and "Microelectronic Circuits," 8th edition (along with A. Sedra, K.C. Smith and V. Gaudet). He was Editor-in-Chief of the *IEEE Transactions on Circuits and Systems II: Express Briefs* in 2009, and an Associate Editor for the *IEEE Journal of Solid-State Circuits* 2010–2017. He was a Distinguished Lecturer for the IEEE Solid-State Circuits Society 2015–2017 and has served on the Technical Program Committee of several IEEE conferences including the International Solid-State Circuits Conference 2016–2021. He is currently the Editor-in-Chief of the *IEEE Solid-State Circuits Letters*.