(article starts on next page)

Article

Learning Gradient-Based Feed-Forward Equalizer for VCSELs

Muralikrishnan Srinivasan [1], Alireza Pourafzal [2], Stavros Giannakopoulos [3], Peter Andrekson [4], Christian Häger [2] and Henk Wymeersch [2,*]

[1] Department of Electronics Engineering, Indian Institute of Technology (BHU), Varanasi 221005, India; muralikrishnan.ece@iitbhu.ac.in
[2] Electrical Engineering, Chalmers University of Technology, 41296 Gothenburg, Sweden; alireza.pourafzal@chalmers.se (A.P.); christian.haeger@chalmers.se (C.H.)
[3] Computer Science and Engineering, Chalmers University of Technology, 41296 Gothenburg, Sweden; stagia@chalmers.se
[4] Photonics, Microtechnology and Nanoscience, Chalmers University of Technology, 41296 Gothenburg, Sweden; peter.andrekson@chalmers.se
* Correspondence: henkw@chalmers.se

Abstract: Vertical cavity surface-emitting laser (VCSEL)-based optical interconnects (OI) are crucial for high-speed data transmission in data centers, supercomputers, and vehicles, yet their performance is challenged by harsh and fluctuating thermal conditions. This paper addresses these challenges by integrating an ordinary differential equation (ODE) solver within the VCSEL communication chain, leveraging the adjoint method to enable effective gradient-based optimization of pre-equalizer weights. We propose a machine learning (ML) approach to optimize feed-forward equalizer (FFE) weights for VCSEL transceivers, which significantly enhances signal integrity by managing inter-symbol interference (ISI) and reducing the symbol error rate (SER).

Keywords: machine learning; optical communications; VCSEL-based optical interconnects; end-to-end learning

1. Introduction

Vertical cavity surface emitting laser (VCSEL)-based optical interconnects (OIs) serve as the primary connectivity solution in data centers, supercomputers, and vehicles, offering cost-effective and high-speed connections [1]. Given the harsh and dynamically changing environments in which these systems operate, VCSELs demand adaptive and resilient design strategies throughout the communication chain [2,3]. Among the many factors that influence the performance and reliability of VCSELs, temperature poses a particular challenge [4]. In short-range OIs, the optical links are positioned close to heat sources, which are typically the processing units, leading to rapid and substantial temperature variations. This scenario is common in data centers, where the compact and densely packed nature of systems often results in significant heat buildup [1]. Such temperature changes impact the operational characteristics of VCSELs in several ways, including increased threshold current, shifts in the emission wavelength of the VCSEL due to changes in the refractive index and the physical dimensions of the laser cavity, decreased output power due to decreased carrier density, and increased non-radiative recombination within the laser's active region [5].

The inherent nonlinear transfer characteristics of VCSELs, especially under significant temperature variations, necessitate sophisticated approaches to ensure optimal operation. For example, maintaining robust 100 Gbps links in such fluctuating environments requires implementing advanced equalization techniques [6]. Equalization can be implemented in two forms: post-equalization at the receiver, and pre-equalization at the transmitter. Pre-equalizers actively modify the signal before it encounters the distorting effects of the transmission medium and VCSEL nonlinearities [7,8]. This proactive approach allows for

the correction of impairments before they occur, making it more efficient than attempting to reverse these effects at the receiver end. Moreover, pre-equalization helps to reduce the complexity and computational load on the receiver, which is particularly advantageous in high-speed applications where minimizing processing delays is crucial.

Equalizers have traditionally been designed based on mathematical models [9–11]. However, considerations of cost, energy efficiency, and temperature variations significantly impact communication capacity. Modeling individual components is already highly challenging; even if a model is available, it tends to be complex, as these models often involve the concatenation of numerous nonlinear, frequency-selective, and noisy submodels, which in turn precludes the possibility of designing an optimal transmitter and the corresponding optimal receiver.

Machine learning (ML) provides an attractive alternative to traditional model-based approaches to overcome the three challenges of modeling, design, and adaptivity [12]. Classical models of components serve as a foundation for constructing neural network (NN) equivalents. In the context of optical communications, receiver-side algorithms encompassing equalization, synchronization, data detection, and decoding can be learned by mimicking conventional algorithms or utilizing deep neural networks (DNNs) from scratch [13–17]. Digital pre-equalization techniques have also gained popularity for enhancing performance the optical communication links [7,8,18,19]. The real-time deployment of NN-based digital equalizers hinges on computational complexity comparable to or lower than conventional digital signal processing (DSP) solutions. For instance, NN-based digital predistortion was designed using three convolutional layers in [20]. In the realm of pre-equalization methods with a view to reducing the complexity, the feed-forward equalizer (FFE) stands out as a prominent analog filter structure employed in transmitters. Operating as a finite impulse response (FIR) filter, the FFE optimally shapes the pulse response, aiming to eliminate inter-symbol interference (ISI) and reduce the link's symbol error rate (SER) performance.

Notably, no work has yet attempted to optimize FFE weights using ML in the context of VCSEL transceivers. Transmitter-side techniques such as pre-equalization introduce significant challenges due to the need for a corresponding adaptive receiver that must participate in the learning process [21]. One of the primary issues is learning the architecture on the transmit side, which often involves backpropagation through a mathematical model of the VCSEL. While differentiable models of VCSELs do exist, such as NN equivalents [6,20,22], the requirements for high-speed operation and precise control necessitate a more comprehensive modeling approach. Accurate modeling of VCSELs involves capturing both the small-signal and large-signal response, including thermal effects across varying temperatures and with limited training samples. However, these comprehensive models, which need to incorporate temperature dynamics explicitly, often lose their differentiability [23]. This complicates the application of standard ML approaches that require gradient calculations. This limitation presents an opportunity to explore novel representations of VCSELs that are both comprehensive and compatible with back-propagation.

Our work includes the integration of an ordinary differential equation (ODE) solver within the VCSEL-based OI chain framework [24,25]. This integration allows for simulating the dynamic behavior of VCSELs using the rate equations and ensures the availability of gradients at each step of the ODE. This gradient availability is essential for updating pre-equalizer weights in gradient-based learning methods.

The contributions of the paper are as follows:

1. **Integration of ODE-based ML for VCSEL Modeling:** We utilize the adjoint method [25,26] within an ML framework for backpropagation through the ODE solver. This approach directly integrates the VCSEL model and its dynamics, avoiding surrogate models and enabling optimization of transmitter components.
2. **Optimization of FFE Weights for VCSEL Transceivers:** Building on the ODE-based integration, we introduce an ML approach to optimize FFE weights for VCSEL transceivers. This method effectively manages ISI and SER, leading to improved overall performance.

The rest of this paper is organized as follows: Section 2 introduces the rate equations of VCSEL and discusses the intrinsic small-signal modulation response; Section 3 provides an overview of FFE; Section 4 describes the ML-based pipeline for training an FFE within an OI system; Section 5 discusses the numerical results; and Section 6 concludes the paper.

2. Rate Equations of VCSEL

The ability of VCSELs to effectively respond to current changes at the data rate is essential for ensuring dependable data transmission. To achieve this, a comprehensive understanding of the VCSEL's dynamic response is necessary. This dynamic response is governed by a set of rate equations that account for the intricate interactions between injected free carriers and photons within the cavity [27].

2.1. Parasitic Elements

Parasitic elements in VCSELs arise from their physical structure and manufacturing processes. These include imperfections at material interfaces such as the p–n junction and metallic contacts, which can lead to unwanted resistance. Parasitic capacitance formed at the interfaces between semiconductor layers and around the active region affects how quickly a VCSEL can respond to input signal changes, limiting the modulation speed. To account for these effects, a simple parallel RC circuit model with resistance (R_j) and capacitance (C_j) components is used in simulations, as shown in Figure 1. Here, I_{in} represents the VCSEL driving current, I is the injection current without any parasitic element, the transfer from I_{in} to I is the parasitic response, and the transfer from I to the optical output is determined by the rate equations of the VCSEL.
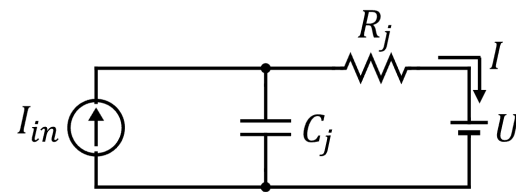

Figure 1. Schematic of the RC circuit model used to simulate the parasitic effects in VCSEL; I_{in} represents the VCSEL driving current, I is the injection current, and U is the device voltage.

2.2. Rate Equations

The laser's operation is modeled through the single-mode laser rate equations derived from a simplified VCSEL model [28], which provide a mathematical framework for understanding the interactions between carrier and photon dynamics within the laser cavity.

2.2.1. Carrier Dynamics Equation

The rate of change of carrier density N [m^{-3}] within the laser's active region is modeled by the following equation [27,28]:

$$\frac{dN}{dt} = \frac{I}{qV} - \frac{c}{n_{\text{geff}}}gS - \frac{N}{\tau_n}. \tag{1}$$

The rate of carrier injection $\frac{I}{qV}$ is driven by the injection current I [A], where q is the elementary charge and V [m^3] is the active volume. The term $\frac{c}{n_{\text{geff}}}gS$ represents the stimulated emission rate, where c denotes the speed of light in vacuum, n_{geff} is the effective modal refractive index, g is the optical gain per unit length, and S is the photon density; lastly, $\frac{N}{\tau_n}$ accounts for the carrier recombination losses, with τ_n [s] as the carrier lifetime, encompassing both radiative and non-radiative decay processes.

2.2.2. Photon Dynamics Equation

The photon density S [m^{-3}] that captures the dynamics of photon population within the laser cavity is provided by the following equation [27,28]:

$$\frac{dS}{dt} = \Gamma \frac{c}{n_{\text{geff}}} gS + \Gamma\beta\frac{N}{\tau_n} - \frac{S}{\tau_p}. \tag{2}$$

The stimulated emission rate is reduced by the internal quantum efficiency Γ. The second term $\Gamma\beta\frac{N}{\tau_n}$ introduces the contribution of spontaneous emission to the overall photon density, with β representing the spontaneous emission coupling factor. The photon losses are modeled by the final term $\frac{S}{\tau_p}$, where τ_p [s] is the the photon lifetime. It is important to note that we treat Γ, g, τ_n, and τ_p as temperature-dependent parameters.

2.2.3. Output Power Equation

The relationship between the output power P_o [W] and photon density S is expressed as [27,28]

$$P_o = S \cdot V \cdot h\nu \cdot \frac{\eta_{out}}{\tau_p\Gamma}. \tag{3}$$

The output power P_o of the VCSEL is directly proportional to the photon density S, and is calculated considering the active volume V and the energy per photon ($h\nu$). The efficiency of the laser output η_{out} expressed relative to the wavelength λ_{cav} [m] quantifies the energy conversion efficiency of the VCSEL, illustrating how the VCSEL converts electrical power into optical power at a specific wavelength.

2.3. Self-Heating

To analyze self-heating effects, an additional set of differential equations is used to monitor the internal temperature (T) of the VCSEL [23]:

$$\frac{\tau_{th}}{r_{th}}\frac{dT}{dt} = g_{gen} - g_{diss} \tag{4}$$

where g_{gen} [W] represents the rate of heat generation, calculated as

$$g_{gen} = U \cdot I_{in} - P_o, \tag{5}$$

where U [V] is the device voltage, I_{in} is the driving current (see Figure 1), and g_{diss} [W] denotes the rate of heat dissipation, provided by

$$g_{diss} = \frac{1}{r_{th}}(T - T_{amb}). \tag{6}$$

where τ_{th} denotes the thermal time constant, r_{th} [K/W] is the thermal impedance, and T_{amb} is the ambient temperature.

2.4. Dynamic Response of VCSEL

In this way, the rate equations establish a direct relationship between the excess carrier density in the active region and the photon density within the cavity when the current passes through the VCSEL. By perturbing these rate equations around a bias current I_b using first-order Taylor expansion and measuring the differential output power, we obtain the intrinsic small-signal modulation response, for which the two-pole transfer function is [27]

$$H_{\text{int}}(f) = \eta_d \frac{hc}{\lambda_0 q} \cdot \frac{f_r^2}{f_r^2 - f^2 + j\gamma\frac{f}{2\pi}}, \tag{7}$$

where η_d is the differential quantum efficiency, h is the Planck constant, c is the speed of light, λ_0 is the lasing wavelength in vacuum, q is the elementary charge, f_r is the resonance

frequency, and γ is the damping factor. The small-signal modulation response is measured by S_{21}

$$S_{21} = 20\log_{10}\frac{\mid H_{\text{int}}(f)}{\mid H_{\text{int}}(0)},\tag{8}$$

and is plotted in Figure 2 for increasing bias current $I_{b1} < I_{b2} < I_{b3}$ and two temperatures, 27 °C and 70 °C, showing the movement of f_r and that the VCSEL reaches a critically damped (flat) response at some current. The plot reveals shifts in resonance frequency and a nonlinear reduction in bandwidth, significantly influencing the frequency response and impacting data transmission capabilities. Pre-equalization at the transmitter, either analog or digital, is crucial to address impairments from the limited bandwidth of VCSELs.
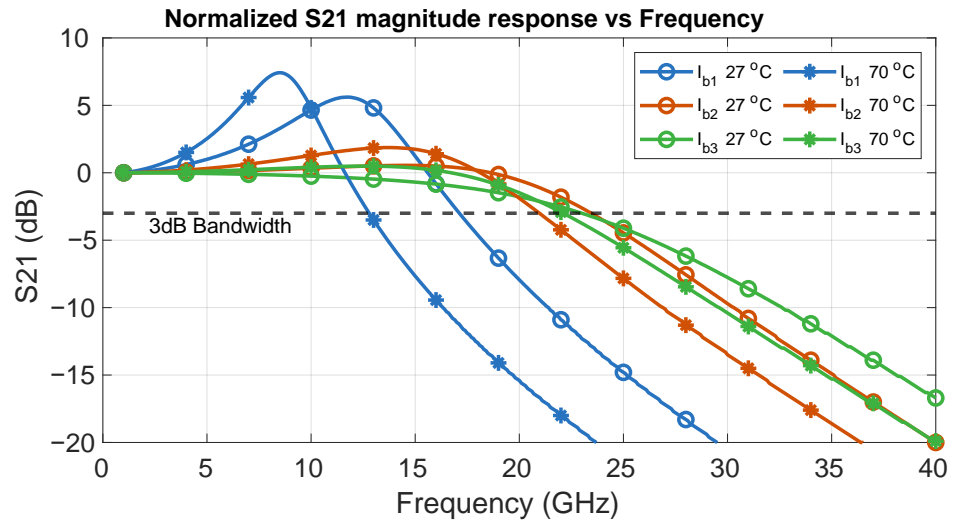

Figure 2. Simulated intrinsic VCSEL response (with parasitic effects neglected) for three representative bias currents ($I_{b1} < I_{b2} < I_{b3}$) and two temperatures (27 °C and 70 °C).

The rate equation is directly solved for the forward inference step using the ODE solver torchdiffeq in PyTorch, generating the output power for input current sequences and establishing the loss function [25]. This library not only facilitates the integration of rate equations but also supplies gradients at each ODE solver step for the back-propagation step. This capability is imperative for updating the transmitter FFE weights in the optimization process described in the next section.

3. FFEs Overview

FFEs use an FIR filter to shape the pulse response and ideally eliminate all ISI. The FFE consists of a series of weighted coefficients called taps. Each tap represents a particular weight applied to a delayed version of the input signal. The number of taps determines the complexity and effectiveness of the equalizer. The delay in an FFE refers to the time difference between the input signal and its delayed versions that are fed into the taps. This delay allows the FFE to capture and compensate for the effects of previous symbols on the current symbol. The output of the FFE at time instant t is provided by

$$I_p(t) = I_{in}(t) + \sum_{k=1}^{K} w_k I_{in}(t - t_k),\tag{9}$$

where $I_{in}(t)$ is the input current, w_k are the tap weights determining the contribution of each delayed input sample $I_{in}(t - t_k)$, and t_k are the corresponding delays.

A model block diagram indicating the position of the FFE in a VCSEL-based OI is shown in Figure 3. The FFE is placed after the digital-to-analog converter (DAC) but before the VCSEL and its parasitic elements. The delay elements in FFEs can be implemented with synchronously clocked flip-flops, transmission lines, or analog delay elements. Coefficient

summing and scaling can be achieved with scaled switched current sources either before or within the final driver stage.
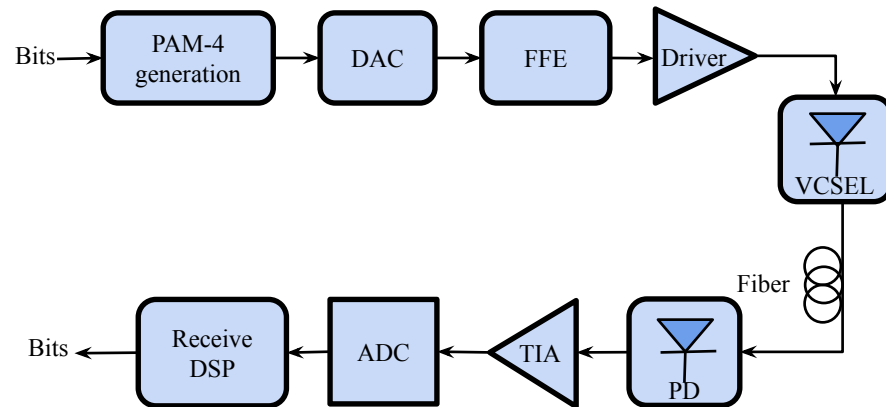

Figure 3. Model block diagram of a VCSEL-based OI system. DAC stands for digital-to-analog converter, ADC stands for analog-to-digital converter, PD stands for photodetector, and TIA stands for trans-impedance amplifier. The FFE weights are optimized in the paper.

However, FFEs have limitations, particularly in filtering out relaxation oscillations under varying biasing and data conditions. Relaxation oscillations or rapid fluctuations in laser output power, can degrade signal quality. Traditional FFE techniques may not fully compensate for these effects due to fixed or inadequately adaptive filter settings [29,30]. To overcome these limitations, we introduce an ML-based approach to dynamically optimize the FFE coefficients w_k. In this paper, we consider ideal driver/FFE electronics, as including transmitter and receiver non-idealities is beyond the scope of the current paper. The following section outlines the end-to-end pipeline for learning FFE weights within the OI system.

4. Pipeline for Learning FFE Weights

The end-to-end ML-based pipeline of the OI system and transmission chain, including the FFE, VCSEL, and the receiver, is shown in Figure 4. Detailed functionality from message encoding to output estimation is provided in the subsequent subsections.
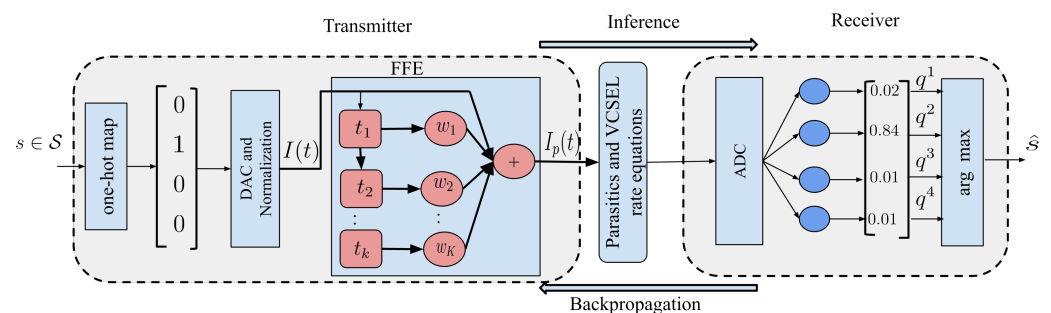

Figure 4. Model block diagram showing the placement of the FFE in the entire chain. Weights w_1 to w_K are learned in the paper.

4.1. Encoding and Input Transformation

The process starts by encoding a message $s \in \{1, \ldots, M\} \triangleq \mathcal{M}$, where $M = 4$. Each message s is encoded into a one-hot vector $\mathbf{x}$, where the s-th element is 1 and all other elements are 0. The output of the one-hot layer, ranging from $[0, 1]$, is scaled and shifted to the dynamic input current range of $[2, 12]$ mA. This ensures that the input remains above the VCSEL threshold current across all ambient temperatures, preventing the AE from arbitrarily increasing the bias current, which would lead to self-heating in a real system [28].

4.2. Signal Conversion and Transmission

After it is prepared, the input is sent to a DAC, converting the digital input into an analog signal for the FFE and VCSEL. The FFE adjusts the signal to compensate for potential distortions before it reaches the VCSEL. The fiber is modeled as a additive white Gaussian noise (AWGN) channel. The system can adapt to include additional features such as low-pass filtering and dispersion as well as intricate circuitry such as output driver circuits, which are beyond the current work's scope.

4.3. Output Processing and Estimation

At the receiving end, the photodiode output is processed through a fully-connected neural network layer with softmax activation. This converts the received signals into a probability vector $\mathbf{q} = [q_1, \ldots, q_M]^\top$, where the estimated message $\hat{s}$ is determined by selecting the highest probability from the softmax output, expressed as

$$\hat{s} = \arg\max_i q_i. \tag{10}$$

4.4. Optimization and Loss Minimization

The network optimization focuses on minimizing the categorical cross-entropy loss function, provided by

$$\mathcal{L} = \sum_{i=1}^{M} x_i \log(q_i), \tag{11}$$

where q_i for $i \in \{1, 2, \ldots, M\}$ is a predicted value and x_i is 1 for true classes and 0 otherwise. This measures the disparity between the predicted and true probability distributions, guiding the model towards accurate predictions by penalizing deviations from the true class probabilities. The loss $\mathcal{L}$ can be related to an achievable information rate using arguments from mismatched decoding [31].

4.5. Training FFE Weights with the Adjoint Method

The goal of the training is to find FFE weights w_k for $k = 1, \ldots, K$ that optimize performance in terms of the categorical loss function. A learning-based model is trained by adjusting its parameters to minimize the difference between its predictions and actual outcomes. Traditional backpropagation involves a backward pass through the network to update these parameters based on the gradient of the loss, leading to challenges with VCSEL components governed by differential equations. To address this, we propose using the adjoint method [26] for training the pipeline. Derived from the framework of neural ordinary differential equations (NODE) [25], this technique integrates ODEs as dynamically learnable components within the network. The adjoint method calculates the adjoint state during the backward pass, representing the gradient of the loss concerning the network state at any given time. By solving the reverse-time ODE for the adjoint state, it is possible to directly compute the gradients with respect to the differential equations governing the VCSEL.

To ensure robust learning, the training process utilized a dataset of 2.5×10^4 randomly selected message symbols processed in batches of 50 symbols over 7500 epochs. Training was conducted at an SNR of 18 dB and a temperature of 70 °C. Determining the number of taps is a critical factor, and is contingent on the desired equalization performance. A greater number of taps in the design enhances equalization performance, as it allows for fine-tuning the FFE frequency response; in turn, this fine-tuning enables more precise shaping of the system's limited bandwidth, leading to significantly improved overall bandwidth performance. The following section provides a detailed analysis of the numerical results and examines potential future extensions.

5. Numerical Results and Discussion

We performed different training for different FFE configurations ranging from two taps to five taps. The delay was chosen in multiples of T_s, where $T_s = 1/(15 * F_s)$ and $F_s = 56$ GBaud is the symbol rate, that is, $n_1 = 6T_s$, $n_2 = 9T_s$, $n_3 = 12T_s$, $n_4 = 15T_s$, and $n_5 = 18T_s$. During training, the FFE weights were optimized to minimize the error between the transmitted and received symbols. The optimized weights for various tap configurations are detailed in Table 1, reflecting the system's adaptation to diverse signal distortions encountered during the training phase.

The effectiveness of the optimized weights is demonstrated through eye diagrams in Figures 5 and 6a–d. Each diagram represents a different tap setting on the FFE, demonstrating how increasing the number of taps affects signal clarity. Figure 5 shows the eye diagram without the FFE. A clear trend is observed in Figure 6a–d, where the signal clarity improves as the number of taps increases. The average eye height increases from 0.6 mW for two taps to about 1.2 mW for five taps. Similarly, the average eye width is about 8.33 ps for two taps and 10.7 ps for five taps. The average jitter for two taps is 9.5 ps, while that for five taps is 7.14 ps. The added taps enhance the complexity of the weights, allowing for finer signal adjustments and reduced inter-symbol interference; however, as shown in Figure 6d, the improvement with five taps is minimal compared to the FFE with four taps.

Table 1. Optimized weights w_1 to w_5 for different tap settings.

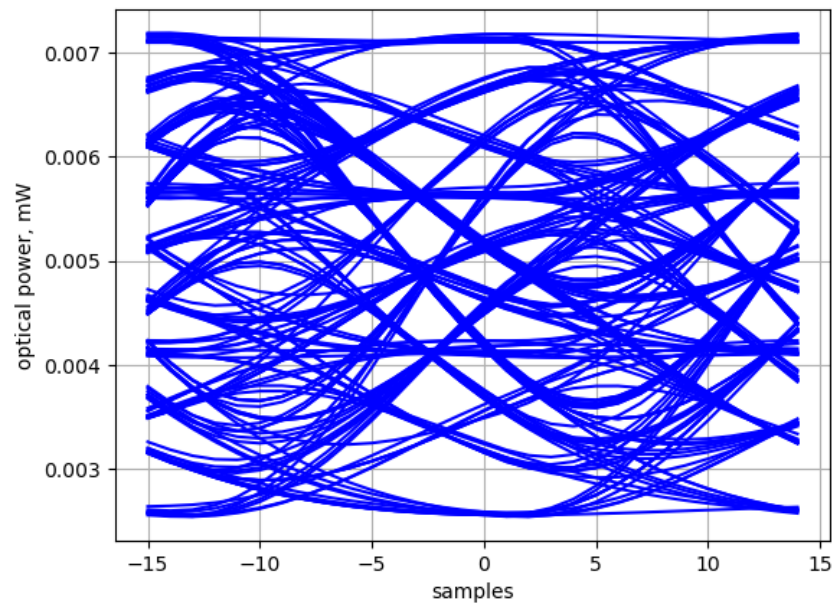| Taps	w_1	w_2	w_3	w_4	w_5
2	0	−0.3438	0	0.0285	0
3	0	−0.4329	−0.3564	0.4961	0
4	0	−0.3357	−0.1708	0.0421	0.1947
5	−0.0281	−0.2644	−0.1067	−0.0223	0.1516

Figure 5. Eye diagram illustrating the signal quality in the absence of FFE.

The improvement in clarity observed in the eye diagram is correlated with a reduction in the SER. This relationship is illustrated in Figure 7, where higher SNRs lead to lower SERs, highlighting the advantages of more advanced pre-equalization techniques. For instance, using five learned taps provides a sensitivity gain of approximately 1 dB over configurations with only two taps. This gain is notable at a low SER level of 10^{-4}, indicating a substantial enhancement in the system's ability to accurately interpret the received symbols. Increasing the number of taps enables more precise adjustments of the equalizer's response, translating to improved performance metrics, such as lower SER at higher SNRs.

The proposed method trained at an SNR of 18 dB generalizes well across high SNR conditions, but may require additional training at lower SNRs to combat increased noise. To address SNR variability more robustly, a similar approach to the distance training method discussed in [32] could be adopted; in this approach, during training the SNRs are drawn from a Gaussian distribution with a mean of 18 dB and a certain standard deviation. Similarly, VCSELs, being temperature-sensitive, demand adaptive models for reliable performance across a wide operating range from $-40\,°C$ to $+125\,°C$. Traditional methods often involve retraining FFEs for different temperatures or fine-tuning via transfer learning. Alternatively, a temperature-adaptive FFE that introduces temperature as an input to the neural network could be explored in future to enable dynamic adaptation without requiring retraining for each scenario. Similarly, to address nonlinearities such as VCSEL relaxation oscillations and temperature-induced variations, nonlinear equalizers based on other deep learning models could be trained to adapt to rapid shifts in operating conditions. This approach would enhance the system's robustness in extreme scenarios. In future work, the benefits of this approach can be explored for different symbol rates versus VCSEL bandwidths. Additionally, learning the delays along with the weights and including transmitter and receiver non-idealities could further enhance the adaptability of the system.
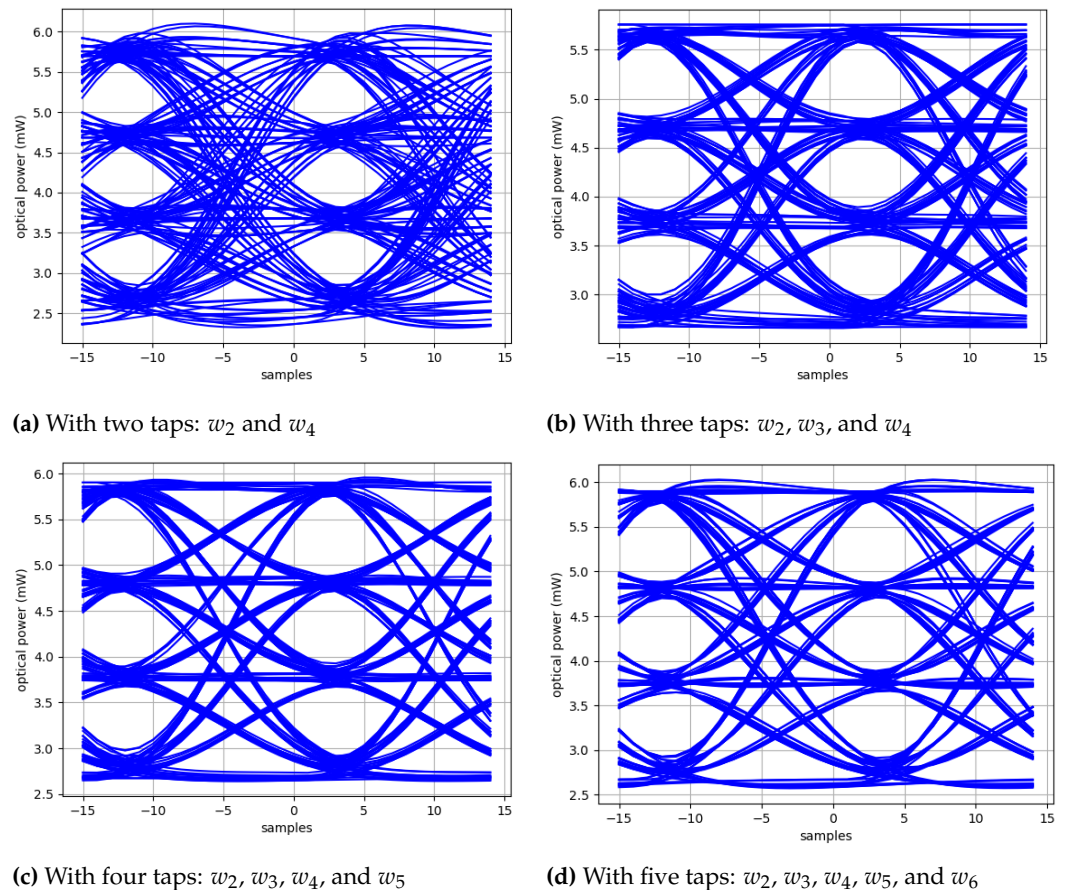

(a) With two taps: w_2 and w_4

(b) With three taps: w_2, w_3, and w_4

(c) With four taps: w_2, w_3, w_4, and w_5

(d) With five taps: w_2, w_3, w_4, w_5, and w_6

Figure 6. Eye diagrams for different optimized FFE tap configurations.
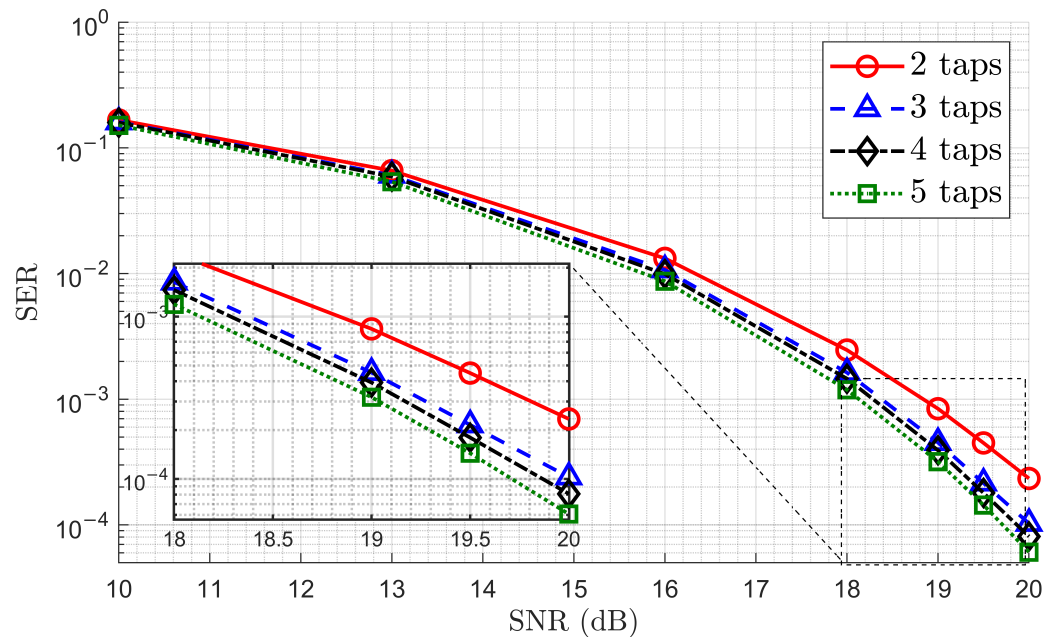
Figure 7. SER vs. SNR for different numbers of FFE taps.

6. Conclusions

In conclusion, we have developed an end-to-end pipeline for optimizing FFE weights within the OI system. By integrating the adjoint method with ODE solvers, we achieve gradient-based optimization of the FFE weights. The results demonstrate significant improvements in signal clarity and performance. Specifically, configurations with more taps enhance signal integrity, with the five-tap setup providing a 1 dB sensitivity gain over the two-tap setup and an SER of 10^{-4}. These findings validate the effectiveness of our approach and highlight the importance of the tap number in optimizing equalization strategies.

Author Contributions: Conceptualization, M.S., S.G. and H.W.; methodology, M.S.; software, M.S.; validation, M.S., A.P. and S.G.; formal analysis, M.S. and C.H.; investigation, M.S.; resources, S.G. and P.A.; data curation, A.P.; writing—original draft preparation, M.S. and A.P.; writing—review and editing, M.S., A.P., C.H., S.G. and H.W.; visualization, A.P.; supervision, H.W.; project administration, P.A., C.H. and H.W.; funding acquisition, P.A., C.H. and H.W. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest; the funders had no role in the design of the study, in the collection, analysis, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

References

1. Cheng, Q.; Bahadori, M.; Glick, M.; Rumley, S.; Bergman, K. Recent Advanced in Optical Technologies for Data Centers: A Review. *Optica* **2018**, *5*, 1354–1370. [CrossRef]
2. Rumley, S.; Bahadori, M.; Polster, R.; Hammond, S.D.; Calhoun, D.M.; Wen, K.; Rodrigues, A.; Bergman, K. Optical Interconnects for Extreme Scale Computing Systems. *Parallel Comput.* **2017**, *64*, 65–80. [CrossRef]
3. King, R. VCSEL Design for Automotive Datacom Experimental Results for 980 nm versus 850 nm. Technical Report, 2021. Available online: https://www.ieee802.org/3/cz/public/may_2021/king_3cz_01a_0521.pdf (accessed on 25 September 2024).

4. Aoki, T.; Kubota, R.; Hiiro, H.; Yoshimoto, S.; Yanagisawa, M. 50 Gb/s PAM-4 VCSELs operating up to 125 °C. In Proceedings of the Optoelectronics and Communications Conference (OECC), Hong Kong, China; 3–7 July 2021; Optica Publishing Group: Washington, DC, USA, 2021; p. T3D.4. [CrossRef]
5. Jasim, F.Z.; Omar, K.; Hassan, Z. Temperature effect on VCSEL output performance. *J. Optoelectron. Adv. Mater.* **2009**, *3*, 1136–1138.
6. Srinivasan, M.; Song, J.; Grabowski, A.; Szczerba, K.; Iversen, H.K.; Schmidt, M.N.; Zibar, D.; Schröder, J.; Larsson, A.; Häger, C.; et al. End-to-End Learning for VCSEL-based Optical Interconnects: State-of-the-Art, Challenges, and Opportunities. *J. Light. Technol.* **2023**, *41*, 3261–3277. [CrossRef]
7. Berenguer, P.W.; Nölle, M.; Molle, L.; Raman, T.; Napoli, A.; Schubert, C.; Fischer, J.K. Nonlinear Digital Pre-distortion of Transmitter Components. *IEEE J. Lightw. Technol.* **2015**, *34*, 1739–1745. [CrossRef]
8. Lavery, D.; Maher, R.; Liga, G.; Semrau, D.; Galdino, L.; Bayvel, P. On the Bandwidth Dependent Performance of Split Transmitter-receiver Optical Fiber Nonlinearity Compensation. *Opt. Express* **2017**, *25*, 4554–4563. [CrossRef] [PubMed]
9. Giannakopoulos, S.; He, Z.S.; Svensson, L.; Zirath, H. Scalable, Modular Feed-Forward Equalizer for Baseband Applications. In Proceedings of the 2022 17th European Microwave Integrated Circuits Conference (EuMIC), Milan, Italy, 26–27 September 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 280–283.
10. Maeda, K.; Yamamoto, S.; Kohmu, N.; Nishimura, K.; Fukasaku, I. An Active-Copper-Cable with Continuous-Time-Linear-Equalizer IC for 30-AWG 7-meters Reach Interconnect of 400-Gbit/s QSFP-DD. In Proceedings of the 2019 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS), Bangkok, Thailand, 11–14 November 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 217–220.
11. Belfiore, G.; Khafaji, M.; Henker, R.; Ellinger, F. A 50 Gb/s 190 mW asymmetric 3-tap FFE VCSEL driver. *IEEE J.-Solid-State Circuits* **2017**, *52*, 2422–2429. [CrossRef]
12. Zhong, K.; Zhou, X.; Huo, J.; Yu, C.; Lu, C.; Lau, A.P.T. Digital Signal Processing for Short-reach Optical Communications: A Review of Current Technologies and Future Trends. *IEEE J. Lightw. Technol.* **2018**, *36*, 377–400. [CrossRef]
13. Shen, T.S.R.; Lau, A.P.T. Fiber Nonlinearity Compensation using Extreme Learning Machine for DSP-based Coherent Communication Systems. In Proceedings of the Optoelectronics and Communications Conference (OECC), Kaohsiung, Taiwan, 4–8 July 2011; pp. 816–817.
14. Gaiarin, S.; Pang, X.; Ozolins, O.; Jones, R.T.; Da Silva, E.P.; Schatz, R.; Westergren, U.; Popov, S.; Jacobsen, G.; Zibar, D. High Speed PAM-8 Optical Interconnects with Digital Equalization based on Neural Network. In Proceedings of the 2016 Asia Communications and Photonics Conference (ACP), Wuhan, China, 2–5 November 2016; pp. 1–3.
15. Ge, L.; Zhang, W.; Liang, C.; He, Z. Compressed Neural Network Equalization based on Iterative Pruning Algorithm for 112-Gbps VCSEL-enabled Optical Interconnects. *IEEE J. Lightw. Technol.* **2020**, *38*, 1323–1329. [CrossRef]
16. Deligiannidis, S.; Bogris, A.; Mesaritakis, C.; Kopsinis, Y. Compensation of fiber nonlinearities in digital coherent systems leveraging long short-term memory neural networks. *IEEE J. Lightw. Technol.* **2020**, *38*, 5991–5999. [CrossRef]
17. Argyris, A.; Bueno, J.; Fischer, I. Photonic Machine Learning Implementation for Signal Recovery in Optical Communications. *Sci. Rep.* **2018**, *8*, 8487. [CrossRef] [PubMed]
18. Paryanti, G.; Faig, H.; Rokach, L.; Sadot, D. A Direct Learning Approach for Neural Network based Pre-distortion for Coherent Nonlinear Optical Transmitter. *IEEE J. Lightw. Technol.* **2020**, *38*, 3883–3896. [CrossRef]
19. Wu, Y.; Gustavsson, U.; Amat, A.G.I.; Wymeersch, H. Low Complexity Joint Impairment Mitigation of I/Q Modulator and PA Using Neural Networks. *IEEE J. Sel. Areas Commun* **2022**, *40*, 54–64. [CrossRef]
20. Minelli, L.; Forghieri, F.; Shao, T.; Shahpari, A.; Gaudino, R. TDECQ-Based Optimization of Nonlinear Digital Pre-Distorters for VCSEL-MMF Optical Links Using End-to-end Learning. *J. Light. Technol.* **2023**, *42*, 621–635. [CrossRef]
21. Schaedler, M.; Kuschnerov, M.; Calabrò, S.; Pittalà, F.; Bluemm, C.; Pachnicke, S. AI-based Digital Predistortion for IQ Mach-Zehnder Modulators. In Proceedings of the 2019 Asia Communications and Photonics Conference (ACP), Chengdu, China, 2–5 November 2019; pp. 1–3.
22. Srinivasan, M.; Song, J.; Häger, C.; Szczerba, K.; Wymeersch, H.; Schröder, J. Learning optimal PAM levels for VCSEL-based optical interconnects. In Proceedings of the 2022 European Conference on Optical Communication (ECOC), Basel, Switzerland, 18–22 September 2022; IEEE: Piscataway, NJ, USA, 2022, pp. 1–4.
23. Grabowski, A.; Gustavsson, J.; He, Z.S.; Larsson, A. Large-Signal Equivalent Circuit for Datacom VCSELs. *IEEE J. Lightw. Technol.* **2021**, *39*, 3225–3236. [CrossRef]
24. Chen, R.T.Q. torchdiffeq, 2018. Available online: https://github.com/rtqichen/torchdiffeq (accessed on 25 September 2024).
25. Chen, R.T.Q.; Rubanova, Y.; Bettencourt, J.; Duvenaud, D. Neural Ordinary Differential Equations. In Proceedings of the NIPS'18: Proceedings of the 32nd International Conference on Neural Information Processing Systems, Montréal, QC, Canada, 3–8 December 2018; pp. 6572–6583.
26. Kidger, P.; Chen, R.T.; Lyons, T.J. "Hey, that's not an ODE": Faster ODE Adjoints via Seminorms. In Proceedings of the ICML, Virtual, 18–24 July 2021; pp. 5443–5452.
27. Coldren, L.A.; Corzine, S.W.; Mashanovitch, M.L. *Diode Lasers and Photonic Integrated Circuits*; John Wiley & Sons: Hoboken, NJ, USA, 2012; Volume 218.
28. Szczerba, K.; Kocot, C. Behavioral Modeling of VCSELs for High-speed Optical Interconnects. In Proceedings of the Vertical-Cavity Surface-Emitting Lasers XXII. International Society for Optics and Photonics, San Francisco, CA, USA, 31 January–1 February 2018; Volume 10552, p. 1055204.

29. Shivashankar, V.; Kottke, C.; Jungnickel, V.; Freund, R. Investigation of linear and nonlinear pre-equalization of VC-SEL. In Proceedings of the Broadband Coverage in Germany; 11. ITG-Symposium, Berlin, Germany, 29–30 March 2017; VDE: Berlin, Germany, 2017; pp. 1–5.

30. Mowlavi, S.; Giannakopoulos, S.; Grabowski, A.; Svensson, L. A Review of IC Drivers for VCSELs in Datacom Applications. *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.* **2023**, *32*, 42–54. [CrossRef]

31. Li, S.; Häger, C.; Garcia, N.; Wymeersch, H. Achievable information rates for nonlinear fiber communication via end-to-end autoencoder learning. In Proceedings of the 2018 European Conference on Optical Communication (ECOC), Rome, Italy, 23–27 September 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1–3.

32. Karanov, B.; Chagnon, M.; Thouin, F.; Eriksson, T.A.; Bülow, H.; Lavery, D.; Bayvel, P.; Schmalen, L. End-to-end deep learning of optical fiber communications. *J. Light. Technol.* **2018**, *36*, 4843–4855. [CrossRef]