

Semi-Supervised and Supervised Nonlinear Equalizers in Fiber-FSO Converged System

Rui Zhang^{ID}, Xizi Tang, Chin-Wei Hsu, You-Wei Chen^{ID}, and Gee-Kung Chang^{ID}, *Fellow, IEEE, Fellow, OSA*

Abstract—We leverage the supervised and semi-supervised Volterra nonlinear equalizers (VNLE) to mitigate the system nonlinearity. Two methods are employed to estimate the coefficients: ordinary least square (OLS) estimator and the least absolute shrinkage and selection operator (Lasso). Due to the additional coupling loss and higher propagation loss in bad weather conditions, FSO-fiber link requires a more stringent power budget. Higher modulation depth and transmitter output power can improve the link budget but need to make nonlinearity correction. Thus, we comprehensively perform a proof-of-concept demonstration in a fiber-FSO converged link with pulse amplitude modulation (PAM). Compared with conventional supervised VNLE using OLS, the coefficients estimated from Lasso require a smaller training symbol overhead. In both the 50-Gbaud PAM4 (at the 1.22×10^{-2} threshold) and 35-Gbaud PAM8 (at the 2×10^{-2} threshold) cases, when the labeled data proportion is 5%, supervised VNLE using Lasso exhibits a received optical power (ROP) improvement up to 3 dB, compared to supervised VNLE using OLS. Moreover, the semi-supervised method can utilize the unlabeled data and further improve the performance without adding signal overhead to the system. In our 50-Gbaud PAM4 experiment, with 60% unlabeled data, the semi-supervised VNLE based on the soft decision (SD) and Lasso demonstrates up to 3-dB sensitivity gain at the BER threshold of 4.5×10^{-3} compared with the supervised VNLE using Lasso. The semi-supervised VNLE using SD and Lasso also demonstrates a line rate improvement $>100\%$ at the 4.5×10^{-3} Pre-FEC BER threshold over the conventional supervised VNLE using OLS.

Index Terms—Fiber optical communication, fiber-wireless integration, free space optics, lasso nonlinearity correction.

I. INTRODUCTION

WITH the growth of the data traffic from emerging services, fiber-wireless integration is slated to become an enabler of realizing diverse user-specific applications while maximizing the capacity of the physical layer infrastructure [1]. With abundant un-licensed spectrum resources and simplicity of installation, free-space optics (FSO)-fiber convergence becomes a promising candidate for providing a flexible, ultra-high

speed wireless delivery when fibers are unavailable or difficult to deploy, which solves the last-mile bottleneck and spectrum congestion issues [1], [2]. Recently, FSO-fiber convergence has been investigated with coherent detection and intensity modulation and direct detection (IM-DD) [2], [3]. IM-DD scheme is more cost-effective compared to coherent detection, but it needs a higher power budget. Compared with fiber link, the hybrid fiber-FSO suffers from additional coupling loss and link loss that depends on the weather conditions. Therefore, high receiver sensitivity is crucial to Fiber-FSO systems. Typically, deeper modulation depth, larger signal amplitude, as well as higher amplifier gain could improve the receiver sensitivity performance at the expense of introducing nonlinearities from modulators and amplifiers [4], [5]. Moreover, high-order modulation format further improves the spectral efficiency but is less tolerant to nonlinearities. Thus, nonlinearity correction techniques are adopted to tackle with these nonlinear impairments.

Volterra nonlinear equalizer (VNLE) is one of the common digital signal processing (DSP) techniques used in optical communication systems to mitigate modulation and amplifier nonlinearities [6]. It employs a polynomial regression and restores the signal by fitting and applying the inverse nonlinear transfer function incurred by both the linear and the nonlinear inter-symbol interferences (ISI) [7]. A simplified VNLE is introduced to reduce the computational complexity by removing the interaction terms and only keeping the memory polynomial terms [5], [8]. Furthermore, there are many research works that theoretically or experimentally investigate sparse Volterra [9], as well as supervised VNLE using Lasso and coefficient pruning [10]–[12]. The added l_0 or l_1 regularization term in VNLE using Lasso enforces the insignificant tap coefficients to be zero. It has demonstrated significant reductions in complexity using regularization in both the passive optical network and optical interconnect systems. On the other hand, supervised and semi-supervised neural network (NN) based nonlinear equalizers have been widely investigated for self-interference cancellation and system nonlinear compensation [13], [14]. However, NN typically exhibits slower convergence rates and higher complexity compared to VNLE, as the later can be solved through convex optimization. The nonlinearity in a fiber-FSO access link mainly comes from modulation and amplifiers, which can be well approximated by polynomial regression. Thus, VNLE is a sufficiently enough. However, the Volterra nonlinear equalizer requires a large amount of training data, especially when the memory length is long. Large training overhead will decrease the transmission efficiency in the case of burst frame, multi-user

Manuscript received March 16, 2021; revised May 20, 2021 and July 5, 2021; accepted July 16, 2021. Date of publication July 21, 2021; date of current version October 4, 2021. This work was supported in part by the NSF Grant Award 1821819 to Industry/University Cooperative Research Center for FiberWireless Integration and Networking at Georgia Institute of Technology. (Corresponding author: Rui Zhang.)

The authors are with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30308 USA (e-mail: ruizhang@ece.gatech.edu; txizi3@gatech.edu; chsu95@gatech.edu; ywchen77115@gmail.com; gkchang@ece.gatech.edu).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/JLT.2021.3098337>.

Digital Object Identifier 10.1109/JLT.2021.3098337

links, or dynamic channel conditions [15], [16], where the taps need to be updated within several μ s or ms intervals.

To further improve the VNLE performance and reduce the requirement on training data size, in this paper, we investigate both the semi-supervised and supervised VNLEs and perform an experimental validation in a fiber-FSO converged transport system. We leverage Lasso and ordinary least square (OLS) in the VNLEs and propose to use either the hard decision (HD) or soft decision (SD) in the semi-supervised VNLEs. Note that one benefit of Lasso is to reduce tap coefficients, which has been comprehensively investigated in supervised VNLEs and can help to reduce the implementation complexity [10], [11]. Another benefit of Lasso is that it can reduce the training symbol size requirement, especially when the ambient dimension of data vectors is much larger than the number of observations, which relaxes the latency and reduces training complexity. The supervised VNLE utilizes only the pilot symbols (labeled data). The semi-supervised VNLE utilizes both the labeled data and part of the unlabeled data to do polynomial regression based on pseudo-label [17], [18]. Experimental results have demonstrated that the semi-supervised VNLEs and the supervised VNLEs using Lasso achieve superior performances than the conventional supervised VNLE using OLS. In our previous work, the performance of supervised VNLE using Lasso and the semi-supervised VNLE using SD and Lasso are compared with the conventional supervised VNLE using OLS [19]. This paper extends the previous work by investigating additional types of VNLEs and conducting a more in-depth analysis of algorithm setting tuning and performance benchmarks.

The paper is organized as follows. Section II recaps the conventional supervised VNLE using OLS and introduces the principle of Lasso and semi-supervised VNLEs. Section III depicts the experimental setup. Section IV shows the experimental results with different settings and labeled and unlabeled data ratios and compares different algorithms. Section V gives a concluding remark.

II. PRINCIPLE OF OPERATION

To mitigate the nonlinearities in the system, the full VNLE fits the nonlinear transfer curve can be represented by the following class of polynomials [7]:

$$z(n) = f_{full}^{(n)}(h, s) = \sum_{k=-(K-1)/2}^{(K-1)/2} h_k \cdot s(n-k) + \sum_{p=2}^{N_P} \sum_{i_1=-(L-1)/2}^{(L-1)/2} \dots \sum_{i_p=-(L-1)/2}^{(L-1)/2} h_{i_1, \dots, i_p} \cdot \prod_{i_p} s(n-i_p), \quad (1)$$

where s is the input data, z is the output data, and h are the tap coefficients. N_P determines the polynomial orders in the VNLE. High N_P can fit the nonlinear model better but requires higher computational complexity. For a given sequence s , the notation $f(h, s)$ stands for the output sequence of VNLE parametrized with h applied to s . We use $f^{(i)}(h, s)$ to denote the i -th element of such sequence. Also, a simplified VNLE (sVNLE) is widely deployed to reduce the computational complexity. It only keeps

the power terms and is presented as [5], [8]:

$$z(n) = f^{(n)}(h, s) = \sum_{k=-(K-1)/2}^{(K-1)/2} h_k \cdot s(n-k) + \sum_{p=2}^{N_P} \sum_{i=-(L-1)/2}^{(L-1)/2} h_{i,p} \cdot s^p(n-i). \quad (2)$$

With known input samples and known labels, the estimator chooses the tap coefficients that minimize the least square error between the estimated output samples and the labels. This method is the conventional VNLE using OLS, which utilizes the labeled data as the training data (i.e., pilot symbols). In the following analysis, our methods generalize it by using the unlabeled data and sparsity-inducing regularization.

Let X denote the labels at the transmitter side and let $\hat{X}$ denote the received labeled data. The objective function for the supervised mode with Lasso consists of two terms, the mean-squared error on the training symbols $\|f(h, \hat{X}) - X\|_2^2$, and an ℓ_1 regularization term. The supervised method objective is thus formulated as a convex optimization program:

$$h = \arg \min \left(\|f(h, \hat{X}) - X\|_2^2 + \lambda \cdot \|h\|_1 \right). \quad (3)$$

When $\lambda = 0$, it degenerates into the supervised VNLE using OLS estimator. The added regularization term can exploit the sparsity structure in the optimal coefficients. Even if the ambient dimension of data vectors is much larger than the number of observations, the Lasso estimator can still recover the optimal coefficients, provided that it is (approximately) sparse [20], [21]. Therefore, it can reduce the requirement on training data size. In addition, the lasso estimator automatically performs model selection: when superfluous high-order coefficients are included, its weight will typically become zero. The regularization parameter λ can be chosen through cross validation or the theoretical formula [20]. In this paper, we choose λ using 10-fold cross validation. Precisely, the algorithm partitions the training data into 10 folds, and use 10% of the training data as a validation dataset to choose λ . This methodology is known to be able to select the optimal value of λ in Lasso [22]. For each value of λ , we perform cross-validation separately, yielding the optimal λ in each case. The Lasso program can be solved via Alternating direction method of multipliers (ADMM). The OLS problems are solved via method of steepest descent in our paper. In each iteration of ADMM for Lasso, the complexity is a matrix-vector product, which is the same as the per-iteration cost of steepest descent for OLS. The ADMM algorithm used in this paper belongs to the class of gradient-based methods. When applied to Lasso with high-dimensional problems, the convergence rates of such methods typically depend on the Restricted Strong Convexity condition of the problem [23]. This is also comparable to steepest descent, for which the convergence rates of gradient-based methods depend on the condition number of the data matrix [24]. In each iteration of both Lasso and OLS, the number of multiplication operations needed is $2n \cdot (K + L \cdot (N_P - 1))$, where n is number of data points, and $K + L \cdot (N_P - 1)$ is the tap number. Moreover, a total number

Algorithm 1: Semi-supervised VNLE using Lasso or OLS

Input: PAM order M , Labeled data X , received labeled data $\hat{X}$, received unlabeled data $\hat{Y} \in \mathbb{R}^N$, initial tap coefficients h^0

Output: Estimated tap coefficients h

for $t = 1, \dots, T$ **do**

 Recover the received unlabeled data using the current tap coefficients: $Z^{(t)} = f(h^{t-1}, \hat{Y})$.

if *soft decision* **then**

 Perform soft decision: $P^{(t)} = \text{SD}(\hat{Y})$, where $P^{(t)} \in \mathbb{R}^{M \times N}$.

 Estimate tap coefficients using Eq(4) or Eq(6).

end

if *hard decision* **then**

 Perform hard decision: $Z^d = \text{HD}(Z^{(t)})$, where $Z^d \in \mathbb{R}^N$.

 Estimate tap coefficients using Eq(5) or Eq(7).

end

end

Fig. 1. Semi-supervised VNLE with Lasso using SD and HD.

of $n \cdot (N_p - 1)$ multiplication operations are needed prior to the iterates, to compute the polynomial terms.

Moreover, we propose to use semi-supervised methods using self-training (also known as pseudo-label) in polynomial regression [17], [18]. It exploits the unlabeled data and further reduces the amount of labeled data needed. This type of methods may get stuck in some cases as the decision error made by the algorithm can reinforce itself, instead of being corrected. Other approaches, such as MixMatch [25] and temporal ensembling [26] may be used to improve the semi-supervised algorithms, albeit costing higher computational complexity. In this work we will focus on the simple self-training method. We use an iterative algorithm, which alternates between estimating the labels and estimating the coefficients. The algorithm takes the labeled data X , received labeled data $\hat{X}$ and received unlabeled data $\hat{Y}$ as inputs. We will consider and investigate both the HD and SD in the decision step. HD is performing maximum likelihood detection on each symbol independently (i.e., based on minimum Euclidean distance). The SD probabilities are computed through the posterior distribution of the input labels given the observed value, where the noise is assumed to be Gaussian. For each iteration, the following computations are performed sequentially as shown in Fig. 1. First, the algorithm recovers the received data using the current Volterra series. Second, perform SD or HD. The SD output is a probability matrix that denotes the probability of the symbols belong to each alphabet. The HD output is the symbol sequence after decision. Let N be the number of the unlabeled symbols and M be the alphabet size (e.g., $M = 8$ for pulse amplitude modulation (PAM) 8), the VNLE tap coefficients at the t -th iteration using SD and HD are estimated as

$$h^t = \arg \min_h \left(\sum_{i=1}^N \sum_{\beta=1}^M P^{(t)}(\beta, i) \cdot \left(f^{(i)}(h, \hat{Y}) - \beta \right)^2 + \alpha \cdot \left\| f(h, \hat{X}) - X \right\|_2^2 + \lambda \|h\|_1 \right), \quad (4)$$

and

$$h^t = \arg \min_h \left(\left\| f(h, \hat{Y}) - Z^d \right\|_2^2 + \alpha \cdot \left\| f(h, \hat{X}) - X \right\|_2^2 + \lambda \|h\|_1 \right), \quad (5)$$

respectively. In Eq (4), $P^{(t)}(\beta, i)$ is the soft-decision probability of the i -th unlabeled symbol being symbol β in the alphabet at the t -th iteration. In Eq (5), $Z^d \in \mathbb{R}^N$ is the symbol sequence generated by HD. Compared to Eq. (3), the objective function Eq. (4) and (5) contain an additional term, which is the weighted mean-squared error on the unlabeled data. This additional term aims at minimizing the difference between the signals after and before the decision. We add a coefficient α to tune the weight between the labeled and unlabeled data. When $\alpha = 0$, the estimator becomes un-supervised mode, which only utilizes the unlabeled data and does not need training symbol or introduce overhead. When $\lambda = 0$, they degenerate into semi-supervised VNLEs using OLS, which are presented as:

$$h^t = \arg \min_h \left(\sum_{i=1}^N \sum_{\beta=1}^M P^{(t)}(\beta, i) \cdot \left(f^{(i)}(h, \hat{Y}) - \beta \right)^2 + \alpha \cdot \left\| f(h, \hat{X}) - X \right\|_2^2 \right), \quad (6)$$

and

$$h^t = \arg \min_h \left(\left\| f(h, \hat{Y}) - Z^d \right\|_2^2 + \alpha \cdot \left\| f(h, \hat{X}) - X \right\|_2^2 \right). \quad (7)$$

The semi-supervised sVNE involves additional unlabeled data in multiplications and needs several times of decision-direct iterations, which shows higher complexity than supervised sVNE. In each iteration, SD needs $M + K + L \cdot (N_p - 1)$ multiplications, M exponential operations, and a logarithmic operation to compute the probability matrix, while HD only needs $K + L \cdot (N_p - 1)$ multiplication. Moreover, to compute tap coefficients in SD, additional $M \cdot n \cdot (K + L \cdot (N_p - 1))$ multiplication operations are needed to compute the product with SD probabilities as shown in Eq (4) and (6).

Proper initial tap coefficients reduce the iteration times of the semi-supervised VNLE using OLS and help to avoid bad local minima. We tried two tap initializations. One approach is a simple initialization that has been used for linear equalizer: the initial tap coefficients in Eq (2) are set to $h_0 = 1$, while the rest of the coefficients are set to zero. The other is called supervised initialization, where the output of supervised VNLE using Lasso serves as the initial tap. The initial BER of supervise initialization is the BER of supervised VNLE using Lasso. The algorithm terminates under certain stopping criteria. In the following experimental evaluation, the algorithm terminates when the Euclidean distances between tap coefficients obtained by 3 consecutive iterates are less than a certain threshold ε (i.e., 0.001) or the iteration times exceed 50. We use normalized values of symbol energy in our algorithms.

III. EXPERIMENTAL SETUP

Fig. 2 shows the experimental setup of the fiber-FSO link. At the transmitter side, the PRBS sequence is mapped to PAM signal and then pulse shaped using a root-raised-cosine filter with roll off factor equals to 0.01 in offline DSP. The symbol number is 10^5 , which includes both the labeled data

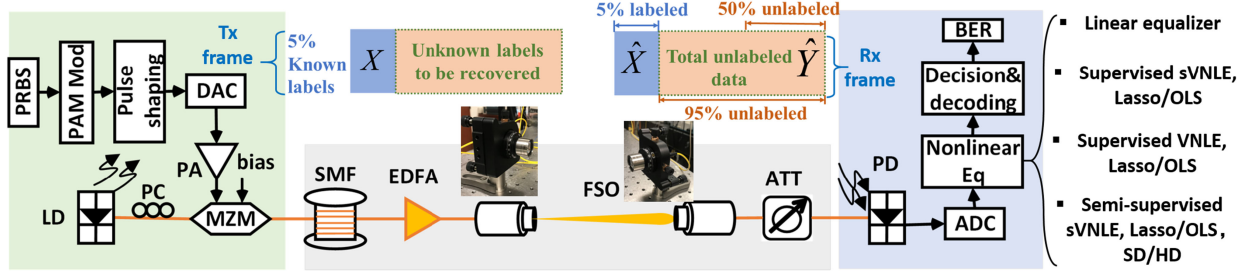


Fig. 2. Experiment setup of the fiber-FSO converged link. PAM mod: PAM modulation. Insets: an instance of the data frame with 5% labeled data.

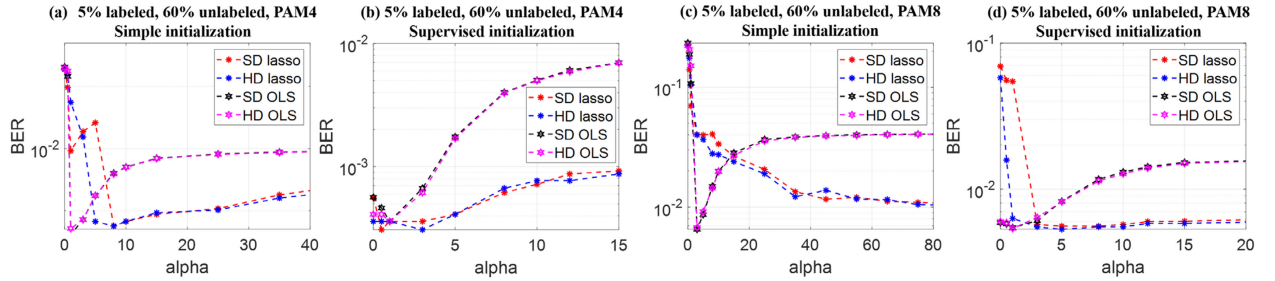


Fig. 3. BER versus weight α with different initialization settings for 50-Gbaud PAM4 and 35-Gbaud PAM8 (5% labeled data and 60% unlabeled data).

and the unlabeled data. The offline generated signal is loaded into digital-to-analogue converter (DAC) and then amplified by a 25-dB power amplifier (PA). The DAC resolution is 8 bits, and its bandwidth is 25 GHz. The transmitter laser diode (LD) operates at the wavelength of 1550 nm and its power is set to be 12 dBm. The electrical signal is then modulated by a 40-GHz Mach-Zehnder Modulator (MZM) with the V_{π} of 5.2 V. The optical signal propagates through a 5-km standard single-mode fiber (SMF) with 1-dB loss and then amplified by an Erbium-doped fiber amplifier (EDFA). The EDFA output power is 18.32 dBm. A pair of collimators transmit the signal through a 2-m free-space optical link with 5.32-dB loss. At the receiver side, an attenuator (ATT) adjusts the received signal power for conducting the sensitivity evaluation. A 50-GHz PIN photodiode (PD) detects the optical signal, which is followed by an 80-GSa/s analog-to-digital converter (ADC) and offline DSP. The VNLEs recover the received signal and then signal decision and decoding are performed to count the bit error rate (BER) for performance evaluation. Several equalizer schemes are investigated in the next section. Linear equalizer only keeps the linear (first order) ISI. Thus, it does not compensate for the nonlinear impairments that result in higher order polynomial terms. For the nonlinear equalizer, the sVNLE in Eq. (2) will be investigated with supervised and semi-supervised methods using Lasso or OLS. The insets of Fig. 2 illustrate an instance of the transmitter data frame and receiver data frame with 5% labeled data. We will sweep labeled data ratio and unlabeled data ratio in the next section. The labeled data ratio is pilot symbol ratio. As for unlabeled data ratio, it means the ratio of the received data that need to be recovered in the semi-supervised methods. For instance, in our experiment, there are 10^5 symbols in each frame. The 5% labeled data means $0.05 \times 10^5 = 5 \times 10^3$ symbols. The 20% unlabeled data means $0.2 \times 10^5 = 2 \times 10^4$ symbols. For

BER evaluation, we use 9×10^4 unlabeled symbols considering the highest portion of labeled data is 10% in the investigation.

IV. EXPERIMENTAL RESULTS

First, we sweep and optimize the semi-supervised VNLE settings including the tap initialization, unlabeled data ratio, memory lengths, as well as the weight (α in Eq. (4) to (7)) between the labeled data and unlabeled data. The peak-to-peak voltage (V_{pp}) is set as 4.2 V for 50-Gbaud PAM4 and 35-Gbaud PAM8. The ROP is 6.6 dBm. The memory lengths are set as $K = 128$, $L = 15$, $N_p = 5$. Fig. 3 presents the BER versus α under two different initialization methods for semi-supervised VNLEs described in Section II. The optimal BER and α can be obtained in the “dip” of each curve. Larger α gives more weight to the labeled data. Using simple initialization, the number of iterations required by the algorithm using OLS (from 15 to 35 typically) is much higher than that using Lasso (< 10). On the other hand, when combined with Lasso, it is harder for the semi-supervised algorithm to converge to a good local minimum. The Lasso program needs enough amount of weight on true labels, so that the support of the sparse vector can be approximately recovered in the initial iterations. This situation is aggravated with PAM8, where the structure of non-linearity is more complicated. As shown in Fig. 3(c), the optimal BER of OLS is better than Lasso in PAM8 signal under simple initialization. Supervised initialization utilizes the output of supervised VNLE using Lasso and help the algorithms to get rid of the bad local optimal. As shown in Fig. 3(b) and (d), Lasso shows better or comparable performance than OLS, and the optimal BER of OLS is also improved compared with the simple initialization. The optimal α of supervised initialization (< 5) becomes smaller than that under simple initialization, which means the algorithm puts more weight on the unlabeled data due to better initial decision

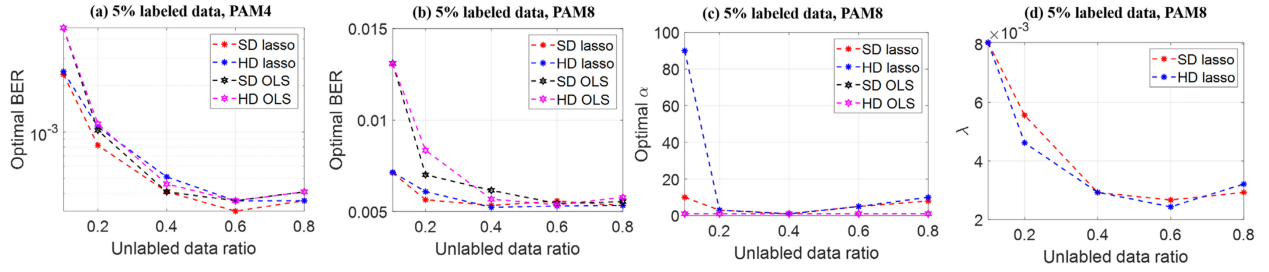


Fig. 4. (a) and (b): Optimal BER versus unlabeled data ratio with 5% labeled data for 50-Gbaud PAM4 and 35-Gbaud PAM8. (c) and (d): optimal α and λ versus unlabeled data ratio with 5% labeled data for PAM8.

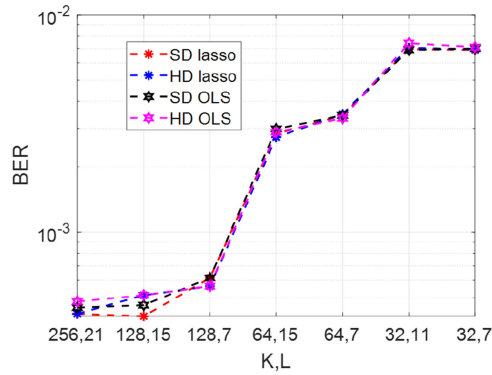


Fig. 5. BER of 50-Gbaud PAM4 versus memory lengths in sVNLE using 3% labeled data and 60% unlabeled data.

errors. Moreover, both semi-supervised methods require smaller iterations (3~7) to converge, under the supervised initialization. In the following analysis, we will use the supervised initialization to get better results.

Then by sweeping the weight α , we obtain the optimal α and BER for different unlabeled data ratios as shown in Fig. 4. Fig. 4(a) and (b) show the optimal BER obtained from the optimal α with different unlabeled data ratios and 5% labeled data for PAM4 and PAM8, respectively. Overall, the performances improve with the increasing of the unlabeled data ratio. When the unlabeled data ratio is small (i.e., $\leq 20\%$), Lasso estimator shows better BER performance than OLS. When the unlabeled data ratio is large (i.e., $>20\%$), the performances of OLS are similar to that of Lasso. Fig. 4(c) present the optimal α with respect to different unlabeled data ratios using 5% labeled data for 35-Gbaud PAM8. When the unlabeled data ratio is small (i.e., 10%), Lasso uses large value of α (i.e., 90) to put more weight on labeled data due to insufficient unlabeled data. When the unlabeled data ratio is larger than 10%, the optimal α becomes smaller (<10). The optimal weights of OLS remain unchanged and the optimal α is 1. As mentioned in Section II, we use 10-fold cross validation to select the optimal value of λ in lasso. Fig. 4(d) shows the optimal λ with different unlabeled data ratios. The value of λ decreases as the unlabeled data size increases. It is due to that the increasing of the observation number can relax the regularization term.

After sweeping the weight and unlabeled data ratios, we sweep different memory lengths as presented in Fig. 5. The unlabeled data ratio is 60% and the labeled data ratio is 3%. With

the decreasing of the linear and power term memory lengths, the BER becomes worse. Note that using even larger memory lengths (i.e., $K = 256, L = 21$) cannot further improve the system performance but brings more complexity. Thus the best BER value is attained by setting $K = 128$ and $L = 15$. As shown in Fig. 3 to Fig. 5, the four semi-supervised algorithms show comparable or similar BER performances in most cases. In some cases, the semi-supervised sVNLE using SD and Lasso shows slightly better performances over other semi-supervised algorithms. Intuitively, this is because SD makes use of the log-likelihood information, instead of only the binary decision. Also, Lasso can reduce the number of tap coefficients and implementation complexity in semi-supervised methods. Thus, in the following analysis, for simplification, we will use SD and Lasso for semi-supervised sVNLE. In practice, HD could be considered due to its lower computational complexity and similar BER performances to SD in most cases. Also, we set $K = 128, L = 15$ to have the best BER performance according to the previous results. The unlabeled data ratio is set as 60%.

Fig. 6 compares the performances of supervised VNLE using Lasso and semi-supervised VNLE using Lasso and SD with different peak-to-peak voltages (V_{pp}) into the modulator. The insets show the eye diagrams after nonlinear compensation under different V_{pp} . Fig. 6(a) and (b) present the results of 50-Gbaud PAM4 signal and 35-Gbaud PAM8, respectively. With 60% unlabeled data, the semi-supervised VNLE using Lasso and SD can improve the BER value by one order of magnitude. As shown in the eye diagrams, lower V_{pp} yields a lower signal to noise ratio (SNR) and lower modulation extinction ratio. Thus, increasing the V_{pp} can improve the signal SNR and BER. However, when V_{pp} exceeds the linear operation region, nonlinearity degrades the signal quality. When V_{pp} is approaching $V\pi$ and the nonlinearity become significantly high, the signal performance would still degrade even if the VNLE is used. One can observe that in insets (vi), the eye diagram still gets compressed after nonlinear correction. Based on these results, in the following analysis, we choose the V_{pp} of the PAM4 and PAM8 as 4.2 V and 3.75 V, respectively.

Fig. 7(a) and (b) show the BER versus labeled data ratios using different VNLEs for both the 50-Gbaud PAM4 ($V_{pp} = 4.2$ V) and 35-Gbaud PAM8 ($V_{pp} = 3.75$ V). The ROP is 6.6 dBm. Three forward error correction codes (FEC) are considered and the corresponding pre-FEC BER thresholds are plotted as reference lines (dashed, horizontal). The thresholds are 4.5×10^{-3} , 1.22×10^{-2} and 2×10^{-2} , which are staircase FEC with 6.69%

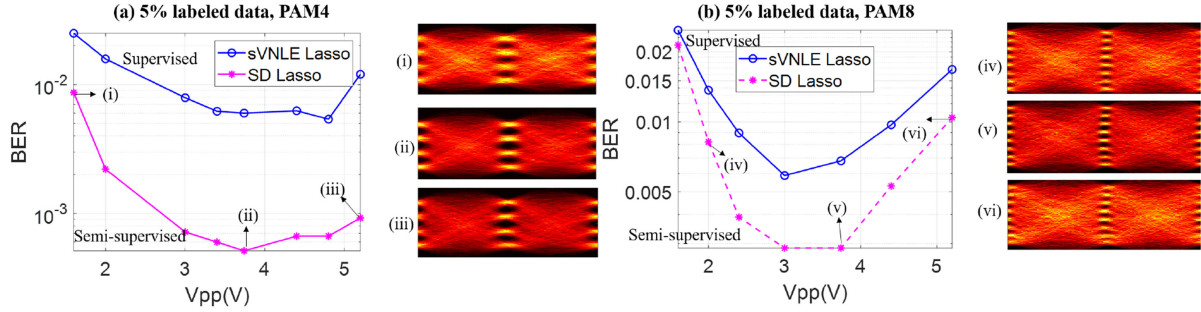


Fig. 6. BER versus Vpp of (a) 50-Gbaud PAM4 using 5% labeled data, (b) 35-Gbaud PAM8 using 5% labeled data. (i) to (vi): signal eye diagrams.

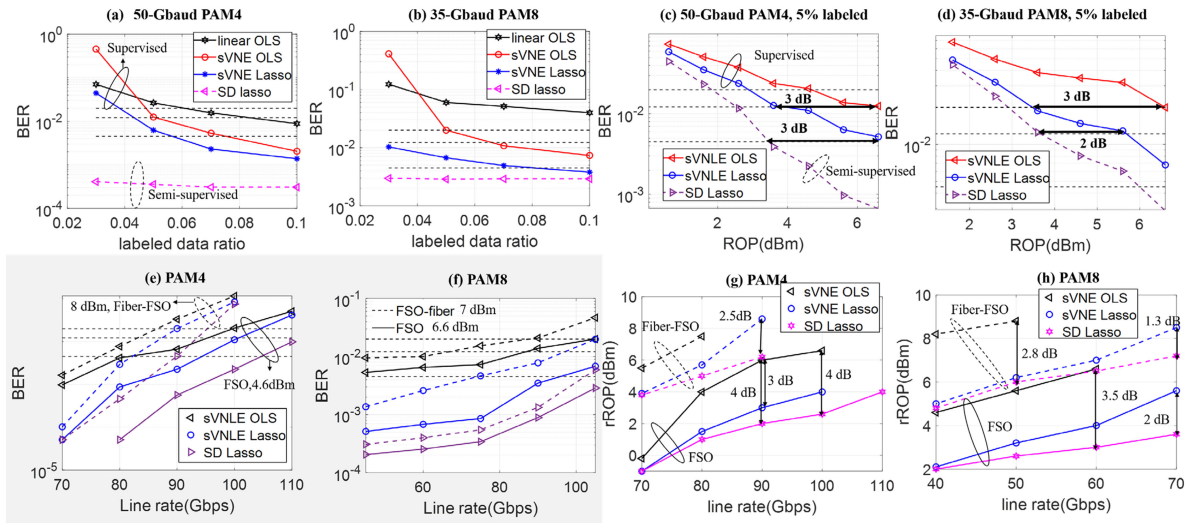


Fig. 7. BER versus labeled data ratio of (a) 50-Gbaud PAM4 and (b) 35-Gbaud PAM8 using 60% unlabeled data. BER versus ROP of (c) 50-Gbaud PAM4 and (d) 35-Gbaud PAM8 using 5% labeled data in FSO transmission. Required ROP versus line rate using 5% labeled data at the 1.22×10^{-2} threshold of (e) PAM4 and (f) PAM8. BER versus line rate of (g) PAM4 and (h) PAM8 using 5% labeled data. (e) to (h): Dashed lines are fiber-FSO transmission, solid curves are pure FSO.

overhead (OH), concatenated FEC (cFEC) with 14.8% OH, openFEC (oFEC) with 15.3% OH, respectively. In both figures, the supervised sVNLE outperforms supervised sVNLE using OLS for the unlabeled data ratio ranging from 3% to 10%. The semi-supervised VNLE using SD and lasso can always achieve a lower BER value than the 4.5×10^{-3} threshold and show a relative flat curve when the labeled data ratio varies. Thanks to the unlabeled data, the semi-supervised VNLE show little degradation when the labeled data ratio is decreased. Moreover, for 35-Gbaud PAM8, linear equalizer fails to achieve the 2×10^{-2} FEC threshold while supervised sVNLE using OLS fails to attain 4.5×10^{-3} threshold. Moreover, at the 2×10^{-2} threshold, the supervised sVNLE using lasso needs 3% labeled data while OLS needs 7% labeled data. Fig. 7(c) and (d) show the sensitivity performance of 50-Gbaud PAM4 and 35-Gbaud PAM8 using 5% labeled data in FSO transmission. At the threshold of 4.5×10^{-3} in PAM4 transmission, the semi-supervised sVNLE using SD and Lasso demonstrates 3-dB sensitivity improvement in comparing with supervised sVNLE using Lasso. The supervised sVNLE using OLS fails to achieve the 4.5×10^{-3} threshold. At the 1.22×10^{-2} threshold in PAM4 transmission, the semi-supervised and supervised VNLE using lasso show 4-dB and 3-dB ROP improvement when compared

with the supervised sVNLE using OLS, respectively. As for PAM8, the semi-supervised algorithm shows 2-dB gain over the supervised sVNLE using lasso at the 1.22×10^{-2} threshold. The supervised sVNLE using lasso achieves 3-dB sensitivity gain when compared with supervised sVNLE using OLS at the 2×10^{-2} FEC threshold. Moreover, the tap number reduction percentage was analyzed in both cases to show the benefit of tap coefficients reduction. In 50-Gbaud PAM4 transmission, the supervised sVNLE using Lasso achieves 15% tap coefficients reduction while the semi-supervised sVNLE achieves 9% reduction. As for 35-Gbaud PAM8 transmission, the supervised sVNLE using Lasso and the semi-supervised sVNLE achieves 47% and 36% tap number reduction, respectively.

Fig. 7(e) and (f) show the BER with different line rates using PAM4 and PAM8 with FSO and FSO plus fiber, respectively. Fig. 7(e) shows the results of PAM4 using 5% labeled data, where the ROP of FSO is 4.6 dBm and ROP of FSO-fiber is 8 dBm. In FSO transmission, the semi-supervised method can always support BER lower than the 2×10^{-2} or 1.22×10^{-2} threshold while other methods cannot support 110 Gbps at all three thresholds. In FSO-fiber transmission, semi-supervised sVNLEs can achieve 4.5×10^{-3} threshold at 90 Gbps while the supervised sVNLE using Lasso and OLS are at 80 Gbps

and 75 Gbps. Fig. 7(f) shares the same legend as Fig. 7(e), which shows the PAM8 results. At the 4.5×10^{-3} threshold, the supervised sVNLE using OLS can only support 45 Gbps in FSO transmission while the supervised sVNLE using Lasso and semi-supervised sVNLEs can support 95 Gbps and >110 Gbps, respectively. Thus, the proposed methods show a line rate improvement higher than 100% in this case. In FSO-fiber transmission (ROP = 7 dBm), sVNLE using OLS can support 65 Gbps at the 2×10^{-2} threshold but fails to achieve 1.22×10^{-2} . The sVNLE using Lasso and semi-supervised sVNLEs can attain the 1.22×10^{-2} threshold at 75 Gbps and 100 Gbps.

Fig. 7(g) and (h) present the required ROP (rROP) at 1.22×10^{-2} thresholds of PAM4 and PAM8 using 5% labeled data, respectively. Higher gains could be achieved using Lasso or semi-supervised method when the line rate is higher. This is due to that the system has fixed swing so the SNR will be lower when the line rate is higher. Lower SNR requires larger amount of data, rendering the data size in our experiments insufficient. In such case, the regularization effect by Lasso and semi-supervised methods can alleviate this issue and achieve better performance. In Fig. 7(g), semi-supervised sVNLE show up to 4-dB and 3-dB sensitivity gain over the supervised sVNLE using OLS and the supervised sVNLE using Lasso, respectively. In Fig. 7(h), the supervised sVNLE using Lasso is better than the conventional method by up to 2.8 dB. Also, the semi-supervised sVNLE shows up to 3.5-dB and 2-dB ROP gain over the supervised sVNLE using OLS and using Lasso.

V. CONCLUSION

We leveraged and investigated semi-supervised methods and Lasso to carry out VNLE, which can mitigate the nonlinearity in optical communication systems. Experimental results in a fiber-FSO link validate that Lasso can reduce the required pilot symbol number by exploiting the sparsity of the tap coefficients. Among the supervised VNLEs, Lasso yields better performance. Experimental results showed that the supervised sVNLE using Lasso outperforms the supervised sVNLE using OLS by up to 3-dB sensitivity gain in 50-Gbaud PAM4 transmission (at the 1.22×10^{-2} threshold) and 35-Gbaud PAM8 transmission (at the 2×10^{-2} threshold). Moreover, the semi-supervised VNLE further improved the BER performance, while maintaining the minimum frame overhead. Among the semi-supervised VNLEs, VNLE using SD and Lasso achieves the best performance. With sufficient unlabeled data, our experimental results have shown that, in both the 50-Gbaud PAM4 transmission (at the 1.22×10^{-2} threshold) and 35-Gbaud PAM8 transmission (at the 2×10^{-2} threshold), the semi-supervised sVNLE using SD and Lasso demonstrates ROP gains up to 4-dB and 3-dB over the supervised sVNLE using OLS and the supervised sVNLE using Lasso, respectively. Furthermore, the proposed methods show significant line rate improvements over existing techniques at certain BER thresholds.

REFERENCES

- [1] Y.-W. Chen, R. Zhang, C.-W. Hsu, and G.-K. Chang, "Key enabling technologies for the Post-5G era: Fully adaptive, all-spectra coordinated radio access network with function decoupling," *IEEE Commun. Mag.*, vol. 58, no. 9, pp. 60–66, Sep. 2020.
- [2] A. Trichili, M. A. Cox, B. S. Ooi, and M.-S. Alouini, "Roadmap to free space optics," *J. Opt. Soc. Amer. B*, vol. 37, pp. A184–A201, 2020.
- [3] Y. Alfadhli *et al.*, "Real-time FPGA demonstration of hybrid bi-directional MMW and FSO fronthaul architecture," in *Proc. Opt. Fiber Commun. Conf. Exhib.*, 2019, pp. 1–3.
- [4] S. P. Yadav and S. C. Bera, "Nonlinearity effect of high power amplifiers in communication systems," in *Proc. Int. Conf. Adv. Commun. Comput. Technol.*, 2014, pp. 1–6.
- [5] G. Khanna *et al.*, "A memory polynomial based digital pre-distorter for high power transmitter components," in *Opt. Fiber Commun. Conf.*, 2017, Paper M2C.4.
- [6] F. Pittalà *et al.*, "800ZR+ DWDM demonstration over 600 km G.654D fiber enabled by adaptive nonlinear TripleX equalization," in *Proc. Opt. Fiber Commun. Conf.*, San Diego, CA, 2020, Paper M4K.5.
- [7] E. Changsoo and E. J. Powers, "A new volterra predistorter based on the indirect learning architecture," *IEEE Trans. Signal Process.*, vol. 45, no. 1, pp. 223–227, Jan. 1997.
- [8] S. Zhalehpour, J. Lin, and L. Rusch, "SiP IQ modulator linearization by memory polynomial pre-distortion model," in *Proc. IEEE Photon. Conf.*, 2017, pp. 317–318.
- [9] V. Kekatos and G. B. Giannakis, "Sparse volterra and polynomial regression models: Recoverability and estimation," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 5907–5920, Dec. 2011.
- [10] W.-J. Huang *et al.*, "93% Complexity reduction of volterra nonlinear equalizer by ℓ_1 -Regularization for 112-Gbps PAM-4 850-nm VCSEL optical interconnect," in *Proc. Opt. Fiber Commun. Conf.*, San Diego, CA, 2018, Paper M2D.7.
- [11] S. Lu, C. Wei, C. Chuang, Y. Chen, and J. Chen, "81.7% Complexity reduction of volterra nonlinear equalizer by adopting L1 regularization penalty in an OFDM long-reach PON," in *Proc. Eur. Conf. Opt. Commun.*, 2017, pp. 1–3.
- [12] C. Chuang *et al.*, "Sparse volterra nonlinear equalizer by employing pruning algorithm for high-speed PAM-4 850-nm VCSEL optical interconnect," in *Proc. Opt. Fiber Commun. Conf. Exhib.*, 2019, pp. 1–3.
- [13] Q. Zhou, S. Shen, Y. W. Chen, R. Zhang, J. Finkelstein, and G. K. Chang, "Simultaneous nonlinear self-interference cancellation and signal of interest recovery using dual input deep neural network in new radio access networks," *J. Lightw. Technol.*, vol. 39, pp. 2046–2051, 2021.
- [14] Q. Zhou, F. Zhang, and C. Yang, "AdaNN: Adaptive neural network-based equalizer via online semi-supervised learning," *J. Lightw. Technol.*, vol. 38, pp. 4315–4324, 2020.
- [15] J. Zhang, Z. Jia, M. Xu, H. Zhang, L. A. Campos, and C. Knittle, "High-Performance preamble design and upstream burst-mode detection in 100-Gb/s/λ TDM coherent-PON," in *Proc. Opt. Fiber Commun. Conf.*, San Diego, CA, 2020, Paper W1E.1.
- [16] M. T. Dabiri, S. M. S. Sadough, and M. A. Khalighi, "FSO communication for high speed trains: Blind data detection and channel estimation," in *Proc. 11th Int. Symp. Commun. Syst., Netw. Digit. Signal Process.*, 2018, pp. 1–4.
- [17] D.-H. Lee, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Proc. Workshop Challenges Represent. Learn.*, 2013.
- [18] X. Zhu and A. B. Goldberg, "Introduction to semi-supervised learning," *Synth. Lectures Artif. Intell. Mach. Learn.*, vol. 3, pp. 1–130, 2009.
- [19] R. Zhang *et al.*, "115-Gbps/λ Fiber/FSO transmission with supervised and semi-supervised nonlinearity correction using lasso," in *Proc. Opt. Fiber Commun. Conf.*, 2021, Paper Th5E.4.
- [20] P. J. Bickel, Y. Ritov, and A. B. Tsybakov, "Simultaneous analysis of Lasso and Dantzig selector," *Ann. Statist.*, vol. 37, pp. 1705–1732, 2009.
- [21] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Stat. Soc.: Ser. B (Methodological)*, vol. 58, pp. 267–288, 1996.
- [22] D. Chetverikov, Z. Liao, and V. Chernozhukov, "On cross-validated Lasso in high dimensions," *Ann. Statist. (Forthcoming)*, vol. 40, 2020.
- [23] A. Agarwal, S. Negahban, and M. J. Wainwright, "Fast global convergence of gradient methods for high-dimensional statistical recovery," *Ann. Statist.*, vol. 40, no. 5, pp. 2452–2482, Oct. 2012.
- [24] M. H. Gruber, *Statistical Digital Signal Processing and Modeling*, ed: Oxfordshire U. K.: Taylor & Francis Group, 1997.
- [25] D. Berthelot, N. Carlini, I. Goodfellow, N. Papernot, A. Oliver, and C. Raffel, "Mixmatch: A holistic approach to semi-supervised learning," *Adv. Neural Info. Process. Syst.*, vol. 32, 2019. [Online]. Available: <https://proceedings.neurips.cc/paper/2019/file/1cd138d0499a68f4bb72bee04bbec2d7-Paper.pdf>
- [26] S. Laine and T. Aila, "Temporal ensembling for semi-supervised learning," in *Proc. 5th Int. Conf. Learn. Represent.*, Toulon, France, Apr. 2017.