

Impairment mitigation in dual-polarization single-span optical digital coherent systems using support vector classifiers

Ivan Aldaya, Lucio Borges, Camila Costa, Julian L. Pita, Rafael A. Penchel, José Augusto de Oliveira, and Grethell Pérez-Sánchez

Abstract— We report on using support vector classifiers (SVCs) to mitigate the residual and fiber-induced nonlinear distortions in a digital coherent optical communication system employing dual-polarization 16-ary quadrature amplitude modulation with a data rate of 100 Gbps. Simulation results reveal that SVC can partially tackle the effect of the intra-polarization and the inter-polarization nonlinear crosstalk. Simulations also show that processing the information of both polarizations leads to improved performance but needs to take special care to avoid overfitting and biasing effect, requiring the implementation of regularization. Regarding the training block size, processing each polarization individually and together require around 20,000 and 27,500 symbols, respectively.

Keywords— Kerr effect, machine learning, optical communications, support vector machines.

I. INTRODUCTION

The development of high-speed digital signal processors (DSPs) enabled the implementation of optical digital coherent communication systems [1], [2]. These systems represented a revolution in the field of optical communications because they allowed the adoption of advanced modulation formats exploiting amplitude, phase, and polarization diversity and paved the way to implement sophisticated impairment compensation techniques [3]. This way, linear impairments, such as phase noise, chromatic dispersion, and linear polarization crosstalk, can be efficiently compensated in the digital domain. However, noise and some practical issues lead to residual effects of linear impairments. Nevertheless, the system performance is mainly limited by the combination of the additive noise and the nonlinear distortion caused by the fiber's Kerr effect.

Due to the stochastic nature of the photodetector noise and its short coherence time, it is difficult to compensate. Nonlinear distortion, on the other hand, is prone to be mitigated as it is deterministic. Different DSP-based compensation techniques have been proposed in this scenario to overcome this impairment. The first attempts included digital back propagation (DBP) [4], [5], inverse Volterra series transfer function (IVSTF) [6], [7], [8], and Wiener-Hammerstein

(WH) [9], [10], which rely on model inversion. Various authors demonstrated the capacity to compensate for the nonlinear distortion of these methods, but their high computational complexity prevented them from being adopted in real-time applications. In order to find a trade-off between performance and computational complexity, different machine learning methods have been proposed, including clustering [11], [12], [13], [14], supervised regression [15], [16], [17], [18], and supervised classification [19], [20]. Supervised classification is particularly interesting because, on the one hand, it typically leads to better performance than clustering. On the other hand, compared to regression, classification is preferable since no further symbol detection is required. Support vector machine is a machine learning technique that has been applied to systems relying on multicarrier modulation formats [21] but remains unexplored in systems using single-carrier modulation formats.

In the present paper, we applied a support vector classifier (SVC) to a single-span digital coherent optical communication system employing dual-polarization (DP) 16-ary quadrature amplitude modulation (16QAM) and operating at 112 Gbps. Simulation results reveal that SVC can effectively mitigate the effects not only of the fiber-induced nonlinear distortion but also of the residual Mach-Zehnder modulator (MZM) impairments. We consider two different configurations: SVC operating on each polarization independently and on both polarizations simultaneously. Simulations reveal that the latter approach is preferable in terms of performance since it can tackle inter-polarization nonlinear crosstalk, but requires a more careful training configuration since it is prone to overfitting.

II. CLASSIFICATION USING SUPPORT VECTOR MACHINES

Supervised classification based on SVC has been applied to a broad variety of problems [22]. SVC was developed by Vapnik and applied to classify data into two distinct classes [23]. The main difference between SVC and other classification algorithms, such as logistic regression, is that SVC maximizes the gap between the data points belonging to different classes [24]. SVC was originally applied to linearly separable binary classification problems. However, SVC was soon extended to multi-class classification by implementing either a one-vs-one or one-vs-all approach [25]. In addition, SVC can be modified to problems with nonlinear boundaries by employing the denominated kernel trick [26]. This technique is similar to mapping the data into higher dimensional

Ivan Aldaya, Lucio Borges, Camila Costa, Rafael A. Penchel, and Jose A. de Oliveira are with the Center of Sustainable Technologies (CAST) at the School of Engineering of São João da Boa Vista (FESJ), Universidade Estadual Paulista (UNESP), São João da Boa Vista - SP, e-mail: ivan.aldaya@unesp.br. Julian Pita is with the Faculté des Sciences, Université du Québec à Montréal (UQÀM). Grethell Pérez Sánchez is with Unidad Azcapotzalco, Universidad Autónoma Metropolitana.

space but requires a much lower complexity. Different kernels can be adopted, the most popular being the polynomial and the radial basis functions (RBFs) [27]. Finally, as in any other machine learning algorithm, a regularization term can be included in SVC to avoid overfitting and improve prediction accuracy [28].

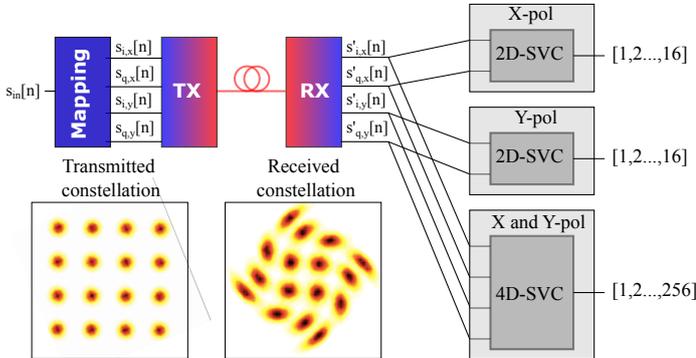


Fig. 1: Block diagram of the optical communication system alongside with the SVCs. We also include the constellations at the input and output of the fiber for a launched optical power of 10 dBm.

In order to mitigate the effects of nonlinearities, SVC is applied to the in-phase and quadrature components of the constellations obtained after the compensation of linear impairments. In this way, it is possible to reduce the effect of inter-symbol interference and consider that the distortion on the symbol, depends only on the actual symbol itself. We can adopt two approaches. On the one hand, we can classify the symbols of the constellations of each polarization independently. Thus, we use two parallel SVC where the inputs are the in-phase and quadrature component of of each constellation, Fig. 1. Alternatively, we can process the two polarizations simultaneously in a single SVC, which is fed by the in-phase and quadrature components of the two polarizations. Regarding the implementation of the SVC classifier, we used the *scikit-learn* Python library. We set the kernel function to RBF, we used the default value for γ , which is inversely proportional to the product of the number of features and and their variance, and the stop condition was configured by setting the tolerance to 10^{-3} . Regarding the regularization parameter C , its value was modified for particular configurations, which is discussed in Section IV.

III. SIMULATION SETUP

Simulations are carried out in VPI Transmission Maker and Python using the simulation setup shown in Fig. 2 and described in [11]. At the transmitter side, a pseudo-random bit sequence with a raw bit rate of 112 Gbps is divided into two subsequences that are mapped into 16-QAM constellations. The 16-QAM symbols are oversampled and filtered using a raised-cosine filter (RCF) with a 20% roll-off factor. The number of samples per symbol is further increased to 64 samples per symbol to emulate the digital-to-analog converter. These signals are used to drive two dual-parallel in-phase-quadrature (DP-IQ) MZMs the two orthogonal polarizations of

the output of a continuous wave (CW) laser with a linewidth of 100 kHz and an operating wavelength of 1550 nm. The two modulated signals are combined into a polarization beam combiner (PBC) and amplified using an erbium-doped fiber amplifier (EDFA) with a noise figure of 4 dB, which is used to sweep the launched optical power from 0 dBm to 10 dBm.

The fiber link is simulated using the vectorial split-step Fourier method (SSFM) configured with a dispersion parameter of $D = 16$ ps/(nm·km), a polarization mode dispersion parameter of $D_{PMD} = 0.1$ ps/ $\sqrt{\text{km}}$, an attenuation coefficient of $\alpha = 0.2$ dB/km, a nonlinear refractive index of $n_2 = 0.26$ $\mu\text{m}^2/\text{W}$ and an effective mode area of $A_{eff} = 80$ μm^2 . The fiber link length was set to 140 km.

The incoming signal is detected using polarization and phase diversity digital coherent receiver at the receiver side. The recovered in-phase and quadrature components of the two polarizations are then digitalized and processed in the DSP. The first block of the DSP is the orthogonalization stage responsible for correcting any unbalance in the 90° optical hybrid. The second stage is a frequency-domain static equalizer that compensates for the chromatic dispersion of the fiber. In the third stage, the polarization mode dispersion and the residual linear polarization crosstalk are mitigated using a dynamic equalizer employing the multi-modulus algorithm (MMA). The frequency drift and the phase noise are compensated using the blind phase search algorithm. Finally, the effects of nonlinear distortion are addressed using SVC operating in two modes: on the one hand, we process each polarization independently, and on the other hand, both polarizations are simultaneously considered.

The number of simulated bits is 2,097,152, corresponding to 262,144 symbols in each polarization. The bit error ratio (BER) is the figure-of-merit employed to assess the signal quality. Since for some launched optical power levels the constellation points do not present Gaussian distribution, BER cannot be estimated using EVM, and direct error counting was adopted.

IV. NUMERICAL RESULTS

In order to analyze the capacity of SVC to mitigate the effects of nonlinear distortion, we apply this classification method to received constellations under different launched optical power conditions. For dual-polarization systems, two approaches can be adopted: apply SVCs to the constellation of each polarization independently, or alternatively, both polarizations can be processed together. In Fig. 3(a), we show the results of applying SVC to each polarization, whereas in Fig. 3(b), the BER results when the two polarizations are processed simultaneously. In the two approaches, different values of the regularization parameter C are considered and the BER curves considering maximum-likelihood (ML) detection are included.

The results shown in Fig. 3(a) reveal that when we apply SVC to each polarization individually, it is possible to reduce the BER for all the considered launched optical power levels. Nevertheless, it can be observed that the BER improvement is more prominent at elevated power levels, indicating that SVC

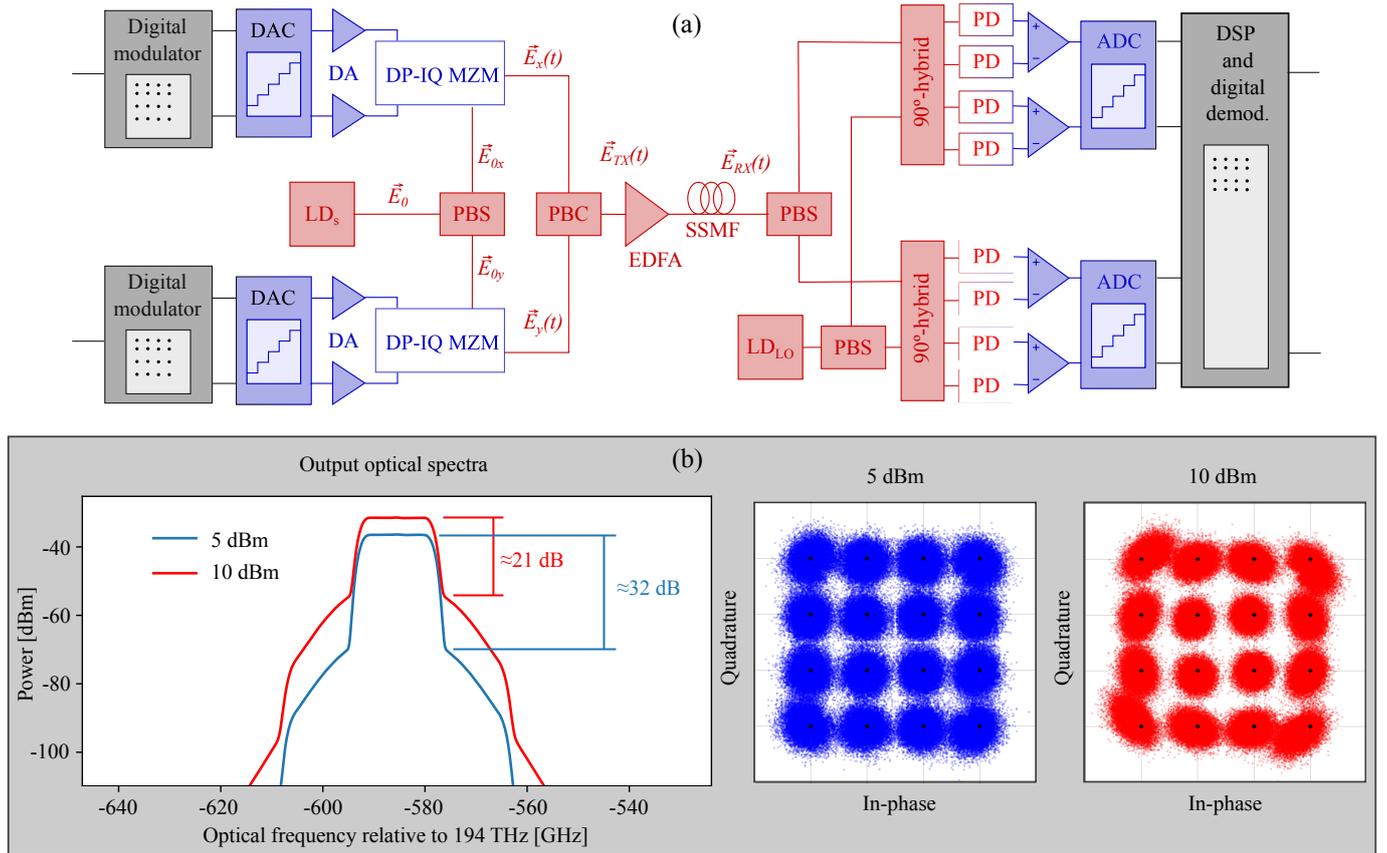


Fig. 2: (a) Block diagram of the employed simulation setup. DAC: digital-to-analog converter, DA: driving amplifier, DP-IQ-MZM: dual-parallel in-phase-quadrature Mach-Zehnder modulator, LD: laser diode, PBS: polarization beam splitter, PBC: polarization beam combiner, EDFA: erbium-doped fiber amplifier, SSMF: standard single-mode fiber, PD: photodiode, and ADC: analog-to-digital converter. (b) Output spectra and constellation diagram for launched optical power levels of 5 dBm and 10 dBm.

partially compensates for the effect of the nonlinear distortion. Regarding the effect of the regularization coefficient C , it can be observed that for the contemplated values $0 \leq C \leq 10$, the obtained BER curves are indistinguishable. In particular, it is important to note that even if no regularization is applied ($C = 0$), good performance is obtained.

In Fig. 3(b), we show the BER curves obtained when the information of both polarizations is processed simultaneously. In contrast to the independent processing of polarizations, the value of C has a dramatic effect on the performance of the SVC. Indeed, it is possible to observe that for $C = 0$, at low power levels, the performance of SVC is even worse than for ML. When we introduce a regularization with low-to-intermediate values, i.e., 0.1 and 1, the performance of SVC improves for all the range of launched optical power. However, when a value of $C = 10$ is employed, the performance degrades. This sensitivity to the value of C can be explained by noting that for $C = 0$, the classification based on SVC suffers from underfitting, whereas for an excessively high value of C , leads to biasing error. Furthermore, it is natural to observe overfitting and biasing effects only when the two polarizations are simultaneously processed because in this approach the dimensionality of the model is higher than when

each constellation is processed independently.

Finally, in Fig. 3(c), we represent the BER curves obtained using ML and the best configurations for the two considered approaches. The first important point to note is that, independently of the approach, SVC outperforms ML for the whole considered range of launched optical power levels. This indicates that SVC is capable not only to mitigate the effects of nonlinear distortion but also some residual linear effects. In particular, looking at the constellation of Fig. 2 for launched optical power of 5 dBm, it is possible to observe a larger separation between the central points. This is probably caused by the nonlinear transmission curve of the MZM. As the launched optical power increases, the fiber nonlinear distortion is more significant and the BER curves of ML and SVC separate. However, up to 7 dBm, the BER curves for the two SVC approaches remain similar. For even higher launched optical power levels, the curves for the two SVC approaches diverge, leading to better results for the combined processing of both approaches. This can be explained by noting that the cross-phase modulation between the orthogonal polarizations is neglected when each polarization is independently processed. When both polarizations are simultaneously processed, on the other hand, the effect of the nonlinear inter-

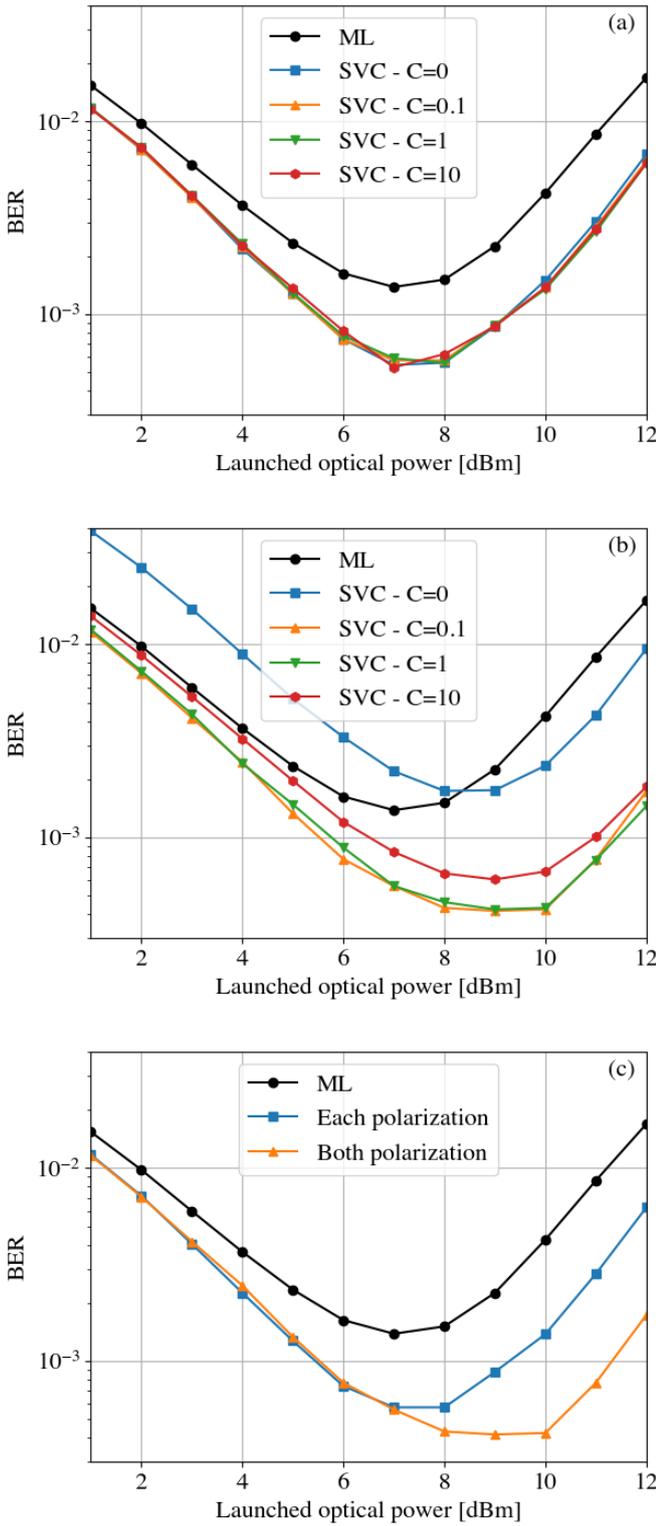


Fig. 3: BER in terms of the launched optical power when (a) each polarization is independently processed and (b) the two polarizations are simultaneously processed. In both cases, regularization parameter values of $C = 0, 0.1, 1$, and 10 were considered. (c) Comparison of the BER curves for $C=1$ of the two approaches. In the three cases, the BER obtained using ML detection is included.

polarization crosstalk caused by the cross-phase modulation can be partially mitigated. Numerically, for ML detection, the optimum BER is obtained for 7 dBm and yields a value of 1.38×10^{-3} . When SVC is applied to each polarization, the optimum BER level is reduced to 5.75×10^{-4} , which is achieved for 8 dBm launched optical power. In contrast, for dual polarization processing, the optimum BER is further reduced to 4.17×10^{-4} , for a launched optical power of 9 dBm.

The capability of SVCs to compensate for the intra-polarization self-phase modulation, inter-polarization cross-phase modulation and some of the transmitter impairments is important, but, to be feasible, the number of symbols required to train the SVCs should be as low as possible. In order to quantify the effect of the training block size on the classification performance, we sweep the number of training symbols from 1,000 up to 30,000 and train the SVC with 100 random samples from the training subset, whereas the test set is fixed. By contemplating different samples for each training block size, we avoid the effect of sensitivity to the possible outliers in the training subset, leading to a more robust metric. The obtained BER values for the different numbers of training symbols when we considered the polarizations individually and together are presented in Fig. 4. For the sake of interpretation, we included the evolution of the mean value of the 100 repetitions (solid line) and the value of the minimum average BER (discontinuous line) for both approaches. As expected, the larger the size of the training block is, the lower the average BER, as the training is more efficient. However, the performance of SVC stabilizes for a certain number of training symbols. Thus, optimum performance is achieved at around 20,000 symbols when processing each polarization. When processing the two polarizations, SVC requires a slightly larger number of symbols, around 27,500. This behavior was expected since higher-dimensionality problems are harder

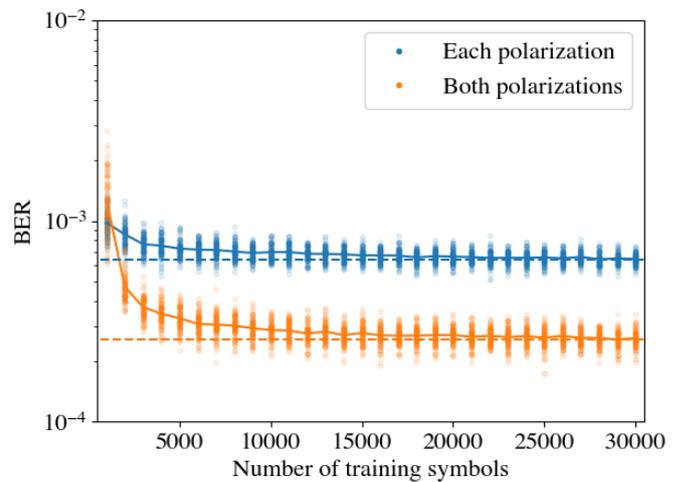


Fig. 4: Obtained BER for different numbers of training symbols. For each number of training symbols, 100 different subsets were considered. The straight line indicates the ensemble average for each number of training symbols, whereas the dotted line represents the BER for 50,000 training symbols.

to train, and consequently, tend to require a larger number of symbols.

V. CONCLUSIONS AND FUTURE WORK

In this paper, we have applied SVC to mitigate the effects of nonlinear distortion in single-span dual-polarization 16QAM digital coherent systems. Numerical results for a 112 Gbps optical interconnect reveal that SVC is capable not only to compensate for the fiber nonlinear distortion but also for some other residual effects. In addition, we show that processing the two polarizations simultaneously can improve classification performance. However, the higher dimensionality of the model requires a careful choice of the regularization constant to avoid overfitting and biasing effect. As described in Section III, the presented analysis is limited to a single-channel system. Future work will assess the impact of the interchannel nonlinear distortion, which is expected to degrade the performance of the classifier, since, assuming that no information from interfering channels is available, the interference would appear as a stochastic effect.

ACKNOWLEDGMENTS

This work was partially funded by the São Paulo State Research Foundation (FAPESP), grants 15/24517-8, 2020/11874-5 e 2020/09889-4, by the National Council for Scientific and Technological Development (CNPQ), grants 313378/2021-5 e 409146/2021-8, and by the Brazilian Innovation Agency (FINEP), project 0527/18.

REFERÊNCIAS

- [1] D. A. de Arruda Mello and F. A. Barbosa, *Digital Coherent Optical Systems: Architecture and Algorithms*. Springer Nature, 2021.
- [2] J. Zhao, Y. Liu, and T. Xu, "Advanced DSP for coherent optical fiber communication," *Applied Sciences*, vol. 9, no. 19, p. 4192, 2019.
- [3] K. Kikuchi, "Fundamentals of coherent optical fiber communications," *Journal of lightwave technology*, vol. 34, no. 1, pp. 157–179, 2015.
- [4] E. Ip and J. M. Kahn, "Compensation of dispersion and nonlinear impairments using digital backpropagation," *Journal of Lightwave Technology*, vol. 26, no. 20, pp. 3416–3425, 2008.
- [5] T. Sutili, S. M. Rossi, R. C. Figueiredo, and D. A. Mello, "Fast adaptive digital back-propagation algorithm for unrepeated optical systems," in *2020 Optical Fiber Communications Conference and Exhibition (OFC)*. IEEE, 2020, pp. 1–3.
- [6] H. Mrabet *et al.*, "A reduced complexity Volterra-based nonlinear equalizer for up to 100 Gb/s coherent optical communications," *Optoelectronics and Advanced Materials-Rapid Communications*, vol. 12, no. 3-4, pp. 186–192, 2018.
- [7] N. Castro and S. Sygletos, "A novel learned Volterra-based scheme for time-domain nonlinear equalization," in *CLEO: Science and Innovations*. Optica Publishing Group, 2022, pp. SF3M-1.
- [8] E. Giacomidis, I. Aldaya, M. A. Jarajreh, A. Tsokanos, S. Thai Le, F. Farjady, Y. Jaouën, A. D. Ellis, and N. J. Doran, "Volterra-based reconfigurable nonlinear equalizer for coherent OFDM," *IEEE Photonics Technology Letters*, vol. 26, no. 14, pp. 1383–1386, 2014.
- [9] J. Pan and C.-H. Cheng, "Wiener-Hammerstein model based electrical equalizer for optical communication systems," *Journal of lightwave technology*, vol. 29, no. 16, pp. 2454–2459, 2011.
- [10] T. Sasai, M. Nakamura, E. Yamazaki, A. Matsushita, S. Okamoto, K. Horikoshi, and Y. Kisaka, "Wiener-hammerstein model and its learning for nonlinear digital pre-distortion of optical transmitters," *Optics Express*, vol. 28, no. 21, pp. 30952–30963, 2020.
- [11] C. Costa *et al.*, "Self-phase modulation and inter-polarization cross-phase modulation mitigation in single-channel DP-16QAM coherent PON employing 4D clustering," *Optical Fiber Technology*, vol. 75, p. 103186, 2023.
- [12] X. Wang, Q. Zhang, X. Xin, R. Gao, Q. Tian, F. Tian, C. Wang, X. Pan, Y. Wang, and L. Yang, "Robust weighted k-means clustering algorithm for a probabilistic-shaped 64QAM coherent optical communication system," *Optics Express*, vol. 27, no. 26, pp. 37601–37613, 2019.
- [13] E. Giacomidis, I. Aldaya, J. Wei, C. Sanchez, H. Mrabet, and L. P. Barry, "Affinity propagation clustering for blind nonlinearity compensation in coherent optical OFDM," in *CLEO: Science and Innovations*. Optica Publishing Group, 2018, pp. STh1C-5.
- [14] I. Aldaya, E. Giacomidis, G. de Oliveira, J. Wei, J. L. Pita, J. D. Marconi, E. A. M. Fagotto, L. Barry, and M. L. F. Abbade, "Histogram based clustering for nonlinear compensation in long reach coherent passive optical networks," *Applied Sciences*, vol. 10, no. 1, p. 152, 2019.
- [15] I. Aldaya *et al.*, "Compensation of nonlinear distortion in coherent optical OFDM systems using a MIMO deep neural network-based equalizer," *Optics Letters*, vol. 45, no. 20, pp. 5820–5823, 2020.
- [16] D. Wang, M. Zhang, Z. Li, C. Song, M. Fu, J. Li, and X. Chen, "System impairment compensation in coherent optical communications by using a bio-inspired detector based on artificial neural network and genetic algorithm," *Optics Communications*, vol. 399, pp. 1–12, 2017.
- [17] A. Shahkarami, M. Yousefi, and Y. Jaouën, "Complexity reduction over bi-RNN-based nonlinearity mitigation in dual-pol fiber-optic communications via a CRNN-based approach," *Optical Fiber Technology*, vol. 74, p. 103072, 2022.
- [18] X. Liu, C. Li, Z. Jiang, and L. Han, "Low-complexity pruned convolutional neural network based nonlinear equalizer in coherent optical communication systems," *Electronics*, vol. 12, no. 14, p. 3120, 2023.
- [19] D. Wang *et al.*, "KNN-based detector for coherent optical systems in presence of nonlinear phase noise," in *2016 21st Optoelectronics and Communications Conference (OECC) Held Jointly with International Conference on Photonics in Switching (PS)*. IEEE, 2016, pp. 1–3.
- [20] R. de Paula *et al.*, "Mitigation of nonlinear phase noise in single-channel coherent 16-QAM systems employing logistic regression," *Optical and Quantum Electronics*, vol. 53, pp. 1–14, 2021.
- [21] E. Giacomidis, S. Mhatli, T. Nguyen, S. T. Le, I. Aldaya, M. McCarthy, and B. J. Eggleton, "Kerr-induced nonlinearity reduction in coherent optical OFDM by low complexity support vector machine regression-based equalization," in *Optical Fiber Communication Conference*. Optica Publishing Group, 2016, pp. Th2A-49.
- [22] J. Cervantes *et al.*, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," *Neurocomputing*, vol. 408, pp. 189–215, 2020.
- [23] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, pp. 273–297, 1995.
- [24] E. Alpaydin, *Introduction to machine learning*. MIT press, 2020.
- [25] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE transactions on Neural Networks*, vol. 13, no. 2, pp. 415–425, 2002.
- [26] V. Jakkula, "Tutorial on support vector machine (svm)," *School of EECS, Washington State University*, vol. 37, no. 2.5, p. 3, 2006.
- [27] B. Scholkopf, K.-K. Sung, C. J. Burges, F. Girosi, P. Niyogi, T. Poggio, and V. Vapnik, "Comparing support vector machines with Gaussian kernels to radial basis function classifiers," *IEEE transactions on Signal Processing*, vol. 45, no. 11, pp. 2758–2765, 1997.
- [28] B. Scholkopf and A. J. Smola, *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2018.