Research Article Vol. 32, No. 13 / 17 Jun 2024 / Optics Express 23561
Optics EXPRESS

Opto-electronic machine learning network for Kramers-Kronig receiver linearization

SARAH MASAAD* AND PETER BIENSTMAN

Photonics Research Group, Department of Information Technology, Ghent University – imec, Belgium
*sarah.masaad@ugent.be

Abstract: We numerically demonstrate the use of an opto-electronic network comprising a photonic reservoir and an electronic feedforward equalizer (FFE) to linearize a Kramers-Kronig (KK) receiver. The KK receiver is operated under stringent conditions, with restricted sampling rates and low carrier powers, resulting in a nonlinear behavior. We propose two different network configurations, varying in the placement of the FFE component, and evaluate their ability to linearize the KK receiver. By training these networks on back-to-back systems, we arrive at a generic solution that significantly enhances the receiver performance, independent of specific link characteristics. The trained networks are tested in a plug-and-play manner across diverse short-reach links that employ standard digital signal processing blocks for the qualization. Our results show significant improvement in receiver linearity, resulting in a reduction in bit error rate of up to a factor of four.

1. Introduction

Optical data traffic has evolved from predominantly traversing long-reach networks that cover extensive distances to being increasingly transported within short-reach networks [1]. This stems from the surge in inter- and intra- data center traffic, as well as from new emergent applications where short but fast communication links are needed, including the widespread adoption of cloud services, the deployment of fiber-to-home networks, and the expanded use of optical front-hauling in mobile communication. These short-haul networks must be designed with power, footprint and cost considerations, which impact decisions regarding transceiver technologies and signal processing methods [2].

A pivotal challenge arises from the sheer volume of transceivers required within these networks. This demand has led to favoring intensity modulation and direct detection as opposed to coherent systems, since they are cheaper to deploy and operate. Although cost and complexity are lower, they come at the expense of poor spectral efficiency. To bridge this gap, self-coherent receivers have gained traction as potential solutions, since they offer high spectral efficiency at lower costs [3]. This is primarily achieved through co-propagating the local oscillator with the modulated message signal, which can reduce the number of hardware components at the receiver. In addition, fewer photodiodes can be used, followed by different processing techniques to retrieve the complex signal from the detector outputs. Notable examples include the Gerchberg-Saxton receiver [4], a silicon photonics based interference cancelling receiver [5], and the Kramers-Kronig receiver [6].

The Kramers-Kronig (KK) receiver, for one, employs a single photodiode and a phase retrieval algorithm derived from the KK relations to reconstruct the complex signal from the measured amplitude [7]. Despite its attractive hardware simplicity, ensuring linearity within this receiver necessitates two operational conditions. First, a minimum-phase signal is required, which is achieved by maintaining a high carrier-to-signal power ratio (CSPR) of around 9 dB. Second, due to the inherent nonlinear computations within the KK receiver's phase retrieval algorithm, at least 6 samples per symbol (sps) are needed for accurate reconstruction. These conditions underscore the tradeoffs between hardware complexity and processing requirements of self-coherent receivers, which affects their practicality in short-haul networks. Furthermore, essential digital signal

#522673 https://doi.org/10.1364/OE.522673
Journal © 2024 Received 8 Mar 2024; revised 8 May 2024; accepted 31 May 2024; published 12 Jun 2024

Optics EXPRESS

post-processing (DSP) operations, including equalization and forward-error correction (FEC), are operational prerequisites that contribute to the overall complexity and cost.

Indeed, DSP algorithms and circuitry are continuously adapted to provide gains in speed and reduce power consumption, tailoring them for optimal performance in short-distance networks. Equalizers integrating machine learning (ML) techniques are also studied under these constraints to find solutions that are feasible to deploy. While traditional ML focused on digitally implemented systems, photonic ML has recently gained momentum since photonics can offer favorable properties including high bandwidth, low power consumption, and a small footprint [8]. Additionally, the use of photonic networks for signal equalization was shown to be successful in several works. For example, passive photonic networks on an integrated chip were shown to perform dispersion compensation and nonlinear equalization of intensity modulated signals numerically [9,10] and experimentally [11,12]. In [13], numerical evaluation of a photonic network comprising ring resonators showed successful equalization of short-reach PAM4 signals. Nonlinear equalization in coherent systems was also experimentally demonstrated in [14].

In this paper, we explore using opto-electronic machine learning networks to ease the CSPR and the sps constraints of KK receivers, such that they are more applicable to short-haul systems. The optical portion of the network is trained following a machine learning paradigm known as reservoir computing [8]. The reservoir is cascaded by a trainable electronic feedforward equalizer. This trainable opto-electronic system sandwiches the KK receiver and is used to compensate for nonlinear errors arising from a KK receiver operated under severely limited receiver resources. Namely, the receiver is operated using 3 sps instead of the required 6 and at CSPRs below 9 dB. Numeric results show significant improvement in receiver linearity, which allows for better integration in a digital signal post-processing pipeline for chromatic dispersion compensation (CDC) and blind phase search (BPS). Importantly, the optical reservoir replaces a bandpass filter before the KK receiver, which is typically required in self-coherent systems. Furthermore, the training is done on a back-to-back system, without requiring knowledge of the link parameters, allowing the network to be inserted in a "plug-and-play" manner in varying links. We target short-reach systems of up to 250 km deploying 64 quadrature-amplitude modulation, and show up to 4 times reduction in bit error rate (BER).

Efforts in utilizing machine learning networks to aid self-coherent receivers have also been reported in other works, albeit utilizing fully digital neural networks. For example, authors in [15] report the use of a convolutional neural network that can emulate the behavior of the KK algorithm while lowering its CSPR requirements. Deep neural networks have also been shown to lower the complexity and the required CSPR of the Gerchberg-Saxton receiver [16,17].

Compared to our previous work [18], this paper utilizes an *opto-electronic* network for equalizing *receiver-based* nonlinearity, as opposed to a *fully optical* network for equalizing *channel-based* nonlinearity. Furthermore, the opto-electronic network is trained on a back-to-back transceiver system such that it can be inserted in various links and seamlessly cascaded with a DSP pipeline for channel impairments. This arrangement allows the network to be utilized for transceiver impairments while channel impairments like dispersion are deferred to the DSP.

In the next section, the operation of the Kramers-Kronig receiver is reviewed. Section 3 then introduces photonic reservoir computing and details the architecture used in this paper. Next, Section 4 describes the full network, the training methods, and the system under study. Section 5 then discusses the achieved results, and the paper is finally concluded in Section 6.

2. Self-coherent Kramers-Kronig receiver

The Kramers-Kronig receiver is a phase-retrieval based self-coherent receiver that uses the known KK relations to reconstruct a complex-valued signal from its magnitude. While the KK relations relate real and imaginary parts of a signal, the KK receiver uses a logarithmic trick to relate the amplitude and phase [19]. Due to the use of the logarithm, the frequency spectrum of the signal

Optics EXPRESS

is broadened and requires at least 3-fold upsampling for correct operation. The reconstruction accuracy of the KK receiver also depends on having satisfied the minimum-phase condition. This can be achieved by maintaining a high power ratio between the co-transmitted carrier and the message signal, which is termed the Carrier-to-Signal Power Ratio (CSPR).

The pipeline of the KK receiver [19] is shown in Fig. 1, with insets of the frequency components of the signal at different stages. Initially, the optical single-sideband signal is photo-detected and then sampled using an analog-to-digital converter (ADC). The signal is then up-sampled, the logarithm is applied, and then the KK relations are applied in the frequency domain. The complex signal produced after the KK relations is exponentiated to reverse the effect of the logarithm and is subsequently down-sampled. The reconstructed signal will include the co-propagated carrier, similar to the optical signal impinging on the photodiode. Further processing is required to retrieve the message signal, which involves removing the carrier and frequency shifting the signal.

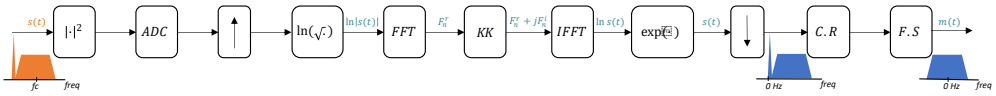

Fig. 1. KK receiver pipeline, showing receiver blocks, several signals along the pipeline, and their frequency spectrum. Receiver blocks: $|\cdot|^2$: Photodiode – **ADC**: Analog-to-Digital Converter – **↑:** upsampling**ln**: natural logarithm – **FFT**: Fast Fourier Transform – **KK**: KK relations – **IFFT**: Inverse Fast Fourier Transform **exp**: exponential – ↓: down-sampling – **C.R**: carrier removal (through filtering) – **F.S**: frequency shifting. Signals: ***s(t)***: single sideband signal (includes unmodulated carrier and message signal) – $\boldsymbol{F_n^r}$: frequency components of real signal – $\boldsymbol{F_n^r + jF_n^i}$: frequency components of complex signal – ***m(t)***: complex message signal

Figure 2 illustrates the combined impact of altering the CSPR and samples per symbol (sps) on the receiver's performance. It showcases BER vs CSPR in a back-to-back KK transceiver configuration utilizing 64 QAM signals at an Optical Signal-to-Noise Ratio (OSNR) of 27 dB. The trendlines depict the influence of employing signals with 6, 4, and 3 sps in the logarithmic and exponential functions. The figure shows that the bit error rate drops as the CSPR increases, irrespective of the sampling rate. Since the BER plateaus at around 9 dB and 6 sps, this can be considered near ideal reconstruction for the system parameters under consideration.

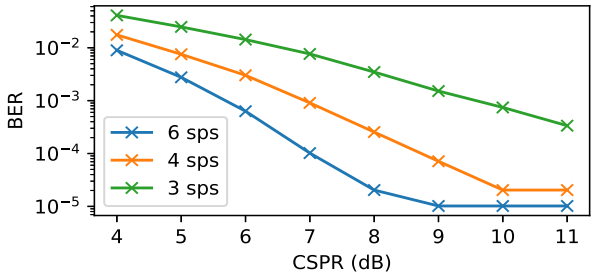

Fig. 2. BER vs CSPR of a back-to-back 64 QAM KK transceiver system operated at 6,4 and 3 samples per symbol

The errors incurred in the back-to-back scenario stem exclusively from the receiver's non-linearity. Some of these error values appear relatively low, but since they are the result of nonlinear transformations, they are expected to significantly impact subsequent digital signal post-processing stages, which typically require linear receiver operation. For instance, attempts to compensate for chromatic dispersion are unlikely to achieve full mitigation, leading to a

Optics EXPRESS

significantly higher BER penalty than that seen in a back-to-back system. This will be shown in Section 5, where we numerically study the impact of the receiver's distortion when short links are introduced in the system. We will also show how our proposed solution can positively influence the receiver's behavior to result in lower BER at seemingly unfavorable operational conditions.

3. Photonic reservoir computing

Neural networks are machine learning models that are quickly finding applications in several areas, including telecommunications. Used as universal approximators, they can be applied to model complex problems such as those that originate from optical fibers and other system components, which allows using them for equalization [20]. These neural networks vary in their architecture to make them better suited for certain tasks, but they are generally composed of function-performing nodes that are connected either recurrently to create memory or in a feed-forward manner for memoryless tasks. Training the network requires using large datasets to find the best weights for the different node connections.

In reservoir computing, a recurrent network's inner connections are left unchanged during training to simplify an otherwise computationally expensive process [21]. Instead, the output weights, known as the readout weights, are trained. Figure 3 shows an example of a 16-node reservoir, where a network of function-performing nodes has fixed weighted interconnects with trainable output weights.

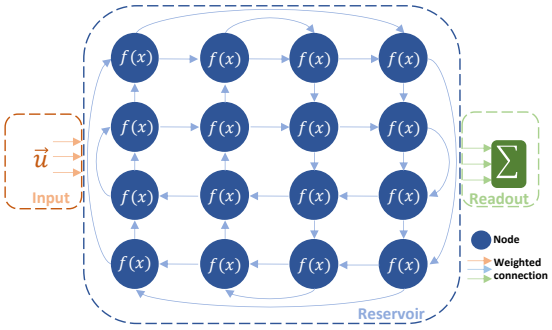

Fig. 3. 16-node reservoir showing function performing-nodes and weighted connections. Input weights (orange) and output weights (green) may be trained, while inner connections (blue) remain fixed.

Besides simplifying the training of digitally realized networks, reservoir computing also enables using physically realized networks, where it is either expensive or not possible altogether to have trainable inner network weights. In such physical systems, dynamics that are naturally occurring or easily realizable in that physical substrate are used for computing [22]. However, weighting elements that allow controlled manipulation of the signals may not be so easy to realize, especially if they are needed for internal connections in the dynamic system. In such cases, reservoir computing has been shown as an effective framework for training and thus utilizing physical systems to compute [23].

To this end, we rely on the reservoir computing framework to enable computing in an integrated photonic system. This allows us to leverage the advantages of photonics for processing signals, especially those that are transported optically. Our implementation of a photonic reservoir consists of 3×3 multimode interferometers (MMIs) connected by spiraled waveguides [24]. The MMIs are the nodes of the network and perform linear mixing of the inputs in the complex domain. Every node has two inputs that come from within the network and two that feed back to the network. The third input allows injecting an external signal at each node, and the third output allows connection to the readout. The fixed internal weights of the network arise from

Optics EXPRESS

the waveguide interconnects, which alter the phase and amplitude of the signals. Due to the sidewall roughness of the waveguides and random variance of their length around the designed value, every reservoir will have a different but fixed set of internal phase weights. Trainable complex-valued optical weights are implemented at the readout.

This network is composed solely of passive components, since both the MMIs and the waveguides do not require driving. While this offers a power advantage, it also makes the reservoir fully linear in the complex domain, thus hindering its use for nonlinear tasks. However, we have shown that by following the reservoir with a nonlinear receiver, the reservoir behaves as a pre-distorting filter for a nonlinear kernel (i.e. the receiver), allowing the entire system to solve nonlinear tasks. This was numerically and experimentally shown to perform well on a variety of tasks, mainly focusing on channel impairments [25]. In the next section, we detail how this photonic reservoir can be utilized as part of an opto-electronic network to equalize KK receiver errors.

4. System details

Our network architecture is composed of a 24-node photonic reservoir co-trained with a 16-tap electronic feed-forward equalizer (FFE). The photonic reservoir replaces the optical bandpass filter standardly required before a KK receiver and hence is not an additional component in the system. Both the reservoir and FFE are composed of generic delay and weighting elements implemented either in the optical or electrical domain, and are therefore simple to implement.

Naturally, the photonic part of the network is placed before the detector, while the FFE part is after the detector. The algorithmic portion of the KK receiver can then be performed before the FFE block or after it. As such, two possible architectures can be modelled and are shown in Fig. 4. Note that the *algorithmic portion* of the KK receiver denotes all the blocks in Fig. 1 that follow the detector, including the carrier removal and the frequency shifting.

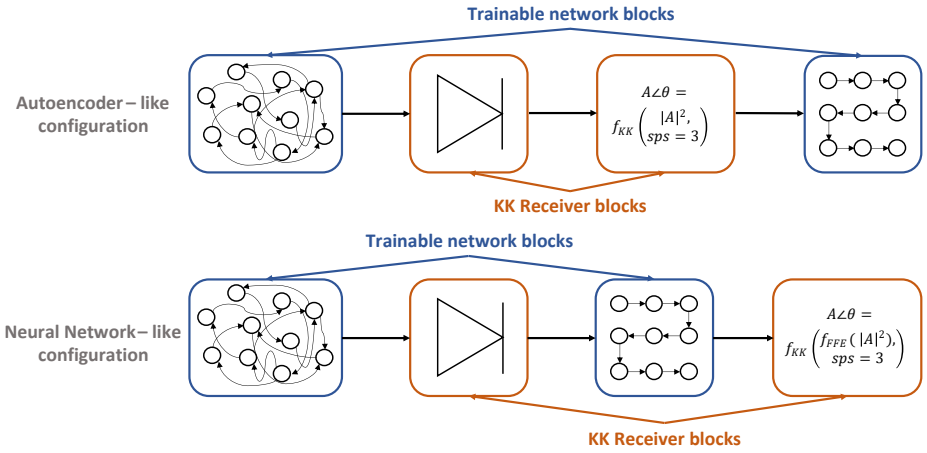

Fig. 4. Two configurations for the opto-electronic network with the KK receiver. The network is composed of two trainable blocks (blue), namely the photonic reservoir (always preceding the photodiode) and the electronic FFE. The KK receiver consists of two blocks (orange), namely the photodiode and the algorithmic steps shown in Fig. 1. The complex-valued output of the receiver in the top configuration (autoencoder-like) is a direct function of the raw measured amplitude and the sps. In the lower configuration (neural network-like), the measured amplitude undergoes an FFE transformation before it continues to the algorithmic portion of the receiver.

Optics EXPRESS

In the top configuration, the FFE portion of the trainable network comes at the end of the full KK receiver. The idea is to create a network where a photonic pre-distorting filter imposes meaningful distortions on the signal making it more robust to the KK receiver's nonlinearity. Then, an FFE post-filter is used to negate these distortion effects and reconstruct the message signal. To better enable a symmetric behavior between the reservoir and FFE, we use a single input node in the reservoir. However, the input node is chosen in the middle of the reservoir to allow for richer reservoir dynamics. Note that the reservoir can use the KK receiver's behavior as a nonlinear kernel. Both the FFE and reservoir have interconnections that introduce half symbol-period delays. This architecture bears similarity to equalization autoencoders used in end-to-end learning of fiber optic systems [26].

In the lower configuration, the KK receiver pipeline is separated by the FFE block, where the FFE is performed after the detector, but before the KK receiver algorithm. This configuration separates the KK receiver nonlinearities such that every linear trainable part of the network is followed by a nonlinear KK receiver transformation. The reservoir's delay lines are designed to introduce half symbol-period delays, while the FFE effectively delays the signal by 40% of the symbol period in every tap. This is because the ADC samples the signal at 2.5 samples/symbol, and the FFE granularity is set to delay 1 sample per tap. Because we aim to enhance the reservoir's dynamics, three input nodes are used: one from the top left corner, one in the center, and one from the bottom right nodes. This configuration bears resemblance to a two-layer neural network where a linear layer is followed by a nonlinear activation function.

For training the weights in both configurations, we set up a back-to-back system to ensure that only the KK receiver errors are compensated for. This facilitates arriving at a generic solution that will linearize the KK receiver without being trained on a specific channel length. Then, the resulting architecture can be inserted in a plug-and-play manner into any channel length, provided that the transceiver operational conditions remain fixed and the channel impairments are compensated for using DSP modules. In addition to reducing the number of times the system needs to be trained since it is unrelated to the system length, this method also simplifies the training process, because the DSP pipeline is not traversed in every training step.

The system schematic is shown in Fig. 5, where the training pipeline is seen in (a) and the testing pipeline in (b). VPI transmission suite [27] is used to simulate the behavior of the transmitter and fiber, while the receiver and trainable network are simulated using PyTorch and the photonic library based on it, Photontorch [28]. The figure shows that three different configurations are deployed for both the training and testing pipelines. Therese are the two network configurations shown in Fig. 4 in addition to a benchmark configuration. For the benchmark, an ideal bandpass filter before the KK receiver is used to filter out the noise pre-detection instead of the trainable photonic reservoir. The benchmark pipeline also includes a trainable 16-tap FFE equalizer. This is done to showcase that the added benefits of the proposed solutions do not come from linear compensations like those achievable by a linear equalizer, despite our trainable network elements being linear. Rather, the gains from our solution are achieved through leveraging the entire pipeline to arrive to the autoencoder-like or neural network-like behaviors described in Fig. 4.

The simulation parameters are detailed in Table 1. The setup employs a 64-QAM modulator transmitting at a rate of 64 Gbaud. This data stream modulates an optical carrier with a linewidth of 200 KHz. The signal's average power remains fixed at 3 dBm, and an OSNR of 27 dB is achieved through gaussian noise loading. Subsequently, an unmodulated continuous wave (CW) laser, positioned approximately 2 GHz away from the message signal's left edge, is optically generated and combined with the signal. The CW laser's power level is varied to achieve CSPRs within the range of 5 dB to 8 dB. The KK receiver algorithm is operated at 3 samples per symbol, which is achieved by upsampling from the ADC's 2.5 samples.

In the testing pipeline, standard single-mode fiber spanning distances from 20 km to 250 km is used. Standard DSP blocks are implemented to execute dispersion compensation and blind phase

Optics EXPRESS

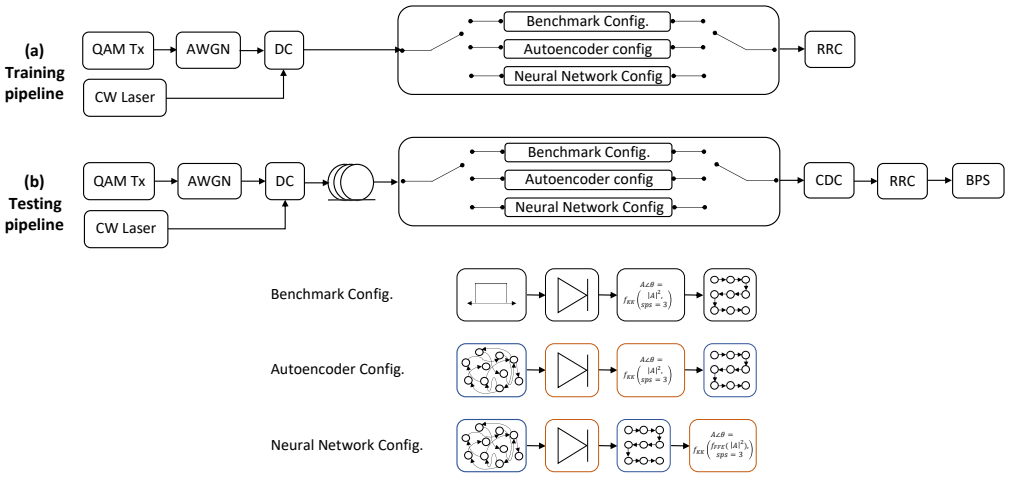

Fig. 5. Schematic of (a) training system and (b) testing system. Systems are simulated three times: with the autoencoder-like configuration, the neural network-like configuration, and a benchmark configuration. The autoencoder-like and the neural network-like configurations follow the description detailed in Fig. 4; the benchmark configuration utilizes an optical bandpass filter before the photodiode and an FFE block after the full KK receiver. **Tx**: Transmitter – **CW** laser: Continuous wave laser – **AWGN**: Additive white gaussian noise **DC**: Directional coupler **CDC**: Chromatic dispersion compensation – **RRC**: Root-raised cosine filter –**BPS**: Blind phase search

Table 1. Transmission system parameters

Transmitter		*Fiber*	
Symbol rate	64 Gbaud	*Length*	20 km – 250 km
OSNR	27 dB	*Attenuation*	0.02 dB/km
Signal power without coupled carrier	3 dBm	*Nonlinear index*	2.6e-20 m^2/W
Pulse shape	root-raised, 0.01 roll off	*Dispersion parameter*	16 ps/(nm km)
Carrier linewidth	200 Khz	*Polarization dispersion*	0 $s/m^{1/2}$
CSPR	5 dB – 8dB		
Guard band	2 GHz		
Carrier linewidth	200 KHz		

search (BPS) processes. Note that the training pipeline excludes a BPS block due to the laser linewidth being set to 0 Hz during training. This deliberate setting expedites the training process by eliminating the necessity for blind phase searches in every training epoch.

The training data set is composed of over 16,000 symbols that are generated using a Wichmann-Hill generator [29]. In the first training stage, the network parameters are initialized in preparation for the second stage which is backpropagation. Initialization of the network parameters is typically an important aspect when backpropagation is used, since the presence of multiple local minima can deter the optimizer from finding the optimal solution and instead get stuck at suboptimal ones. For example, in our networks, a random initialization followed by backpropagation continuously underperformed the benchmark. As such, Linear Regression (LR) is used to find an initial set of weights for the reservoir weights. A closed form solution, given by Eq. (1), is found through fitting the reservoir states to a target signal that is generated similarly to the training pipeline but

Optics EXPRESS

does not include noise.

$$W = (X^H X + \alpha I)^{-1} X^H y, \tag{1}$$

where W is the readout weight matrix, X is the readout state matrix, α is a regularization parameter to curb overfitting, I is the identity matrix, and y is the target signal. The superscript H refers to the conjugate-transpose of the matrix, which is needed for complex-valued data. It turns out that this first LR step conditions the reservoir to behave as a bandpass filter, which is necessary to avoid signal-noise beating in the detector. Indeed, it has a similar performance as the benchmark configuration, where such a bandpass filter is explicitly implemented. Initializing the reservoir for this behavior allows starting at the benchmark performance where noise is filtered from the signal before attempting to optimize this solution further.

Next, we address the initialization of the FFE weights. The straightforward initialization for the FFE weights is setting all the weights to 0 except the center tap which is set to 1. This initialization is used for the autoencoder-like configuration. For the neural network-like configuration, we were concerned that the first network layer (i.e., the reservoir) would have a more advanced initialization than the second layer. Since the FFE in this case is also followed by a nonlinear receiver transformation, it might not be able to take advantage of this if the initialization is near a suboptimal local minima. As such, we used a genetic algorithm known as CMA-ES (Covariance matrix adaptation evolution strategy) [30]to find a new initialization point. To allow flexibility, the algorithm was also able to vary the LR initialization of the reservoir weights. However, we did not exhaustively investigate the influence of this approach.

After initialization, backpropagation is used to alter the optical reservoir weights as well as the electrical FFE weights. The same procedure is followed for both configurations. A gradient descent optimizer with momentum [31] is used to minimize the mean-square error (MSE) loss between the signal and the target symbols

$$MSE = \Delta I^2 + \Delta Q^2, \tag{2}$$

where I and Q are in the in-phase and quadrature components of the QAM signal. However, since the reservoir behaves as a trainable filter, a solution that backpropagation can arrive at is one where the signal power is attenuated with respect to the carrier. While this would lead to a higher CSPR and thus to MSE reduction, this is an undesired effect since it achieves performance improvements without leveraging the network's behavior and at the expense of signal attenuation. To prevent this from happening, we introduce in the error function a second term that heavily penalizes growth of CSPR while also being neutral to the reduction of the CSPR. Such behavior is found in a softplus [32] function, which is also differentiable and continuous. The training error is then a sum of the CSPR and MSE errors as described by Eq. (3).

$$error_{training} = error_{CSPR} + MSE, \tag{3}$$

$$error_{CSPR} = \frac{1}{\beta} \ln(1 + \mathbf{exp}(\beta \times \Delta CSPR)), \tag{4}$$

$$\Delta CSPR = CSPR_{reservoir} - (CSPR_{transmitted} + \alpha), \tag{5}$$

where $CSPR_{transmitted}$ is the desired CSPR to maintain (which corresponds to the CSPR the signal was transmitted at) and $CSPR_{reservoir}$ is the CSPR measured at the output of the readout. At the end of the training, $CSPR_{reservoir}$ must be less than or equal to $CSPR_{transmitted}$. The hyperparameter β and α allow optimizing the behavior of the softplus function to appropriately penalize $\Delta CSPR$ with respect to MSE. This ensures that the MSE contribution is still significant and optimized for during training.

To escape potential local minima, the learning rate (LR) of the optimizer is chosen such that it is slightly high, causing the training error to temporarily and sporadically increase during

Optics EXPRESS

training. This is because there are two error terms, and allowing the optimization algorithm to explore options where the CSPR is temporarily increased allows arriving at better overall solutions. Naturally, the LR should not be too high, as the error can then spiral reaching large values and getting stuck at a local minima. To showcase the training dynamics, a segment of the training progress is shown in the Fig. 6. The plot shows $error_{training}$ increasing momentarily during the training before decreasing again. Note that overfitting is not commonly a problem in these networks since the network's nonlinearity is limited. As such, we rely on the training error and do not use a validation set.

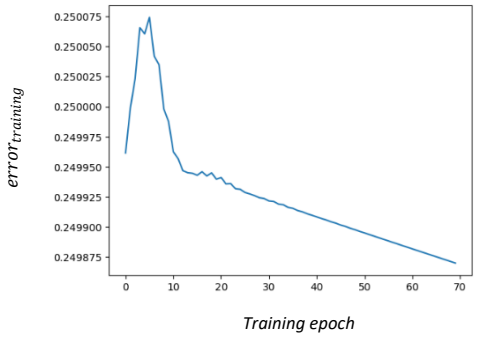

Fig. 6. Plot showing a segment of the training error vs epoch. A momentary increase in error is seen before it starts decreasing again

For the benchmark pipeline, the only trainable block is the FFE since the optical bandpass filter has fixed parameters. The FFE weights are first initialized with 0s except for the center tap which is initialized with 1. Then, backpropagation is used to optimize the weights using the MSE error function of Eq. (2). Note that monitoring the CSPR is irrelevant in this configuration since the optical bandpass filter is not trainable and will not change the CSPR. However, backpropagation training contributed to no improvements compared to the performance of the initialized weights. On the contrary, it worsened the overall BER since optimizing for MSE does not necessarily lead to better BER. Note that backpropagation cannot use BER as an error function since it is a non-continuous function. To rule out the possibility of being stuck in a local minimum due to our initialization, we moved to finding FFE weights through CMA-ES and used BER as a metric instead of MSE. However, the training resulted in less than 3% improvement, if any, compared to the initialization. This shows that there are little to no gains achievable through standard linear equalization. The benchmark results reported in the next section are based on the CMA-ES solution for the FFE block.

5. Results and discussion

The training BER achieved in back-to-back systems is shown in Fig. 7 for CSPR values between 5 and 8 dB. For each CSPR, a different reservoir with random input and internal phase variations is simulated and a different dataset is also used. Both proposed network configurations had substantially lower BERs compared to the benchmark, with the autoencoder-like configuration outperforming the neural network-like configuration.

For the testing dataset, over 32000 symbols were generated using a Whichmann-Hill generator, seeded at random and different seeds to the training ones. A different test set was also used for every CSPR value. Figure 8 shows the achieved test BER against fiber lengths spanning between 20 and 120 km for three CSPR values. Results for the autoencoder-like configuration are shown on the left subplot and results for the neural network-like configuration are shown on the right subplot. Results from the proposed networks are shown in solid lines, contrasted against results

Optics EXPRESS

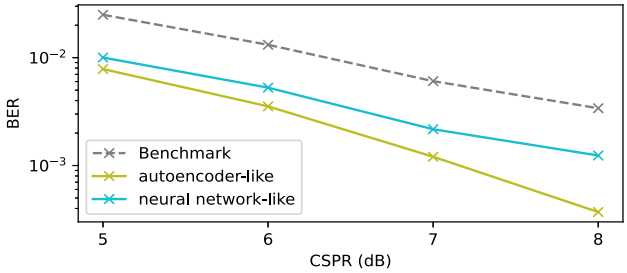

Fig. 7. Back-to-back BER measured during the training phase for CSPRs ranging between 5 and 8 dB and at 3 sps. The two proposed network configurations are shown against the benchmark pipeline.

from the benchmark configuration which are shown as dashed trendlines. Each color corresponds to a different CSPR. When compared to a back-to-back system, the fiber channel in the testing pipeline adds linear errors that are nonlinearly transformed due to the receiver and therefore are incompletely compensated using standard CDC. As such, the errors of the testing pipeline are substantially higher than those observed in the training configurations. Note that the nonlinear Kerr effect in the fiber is negligible at the power levels investigated in this paper.

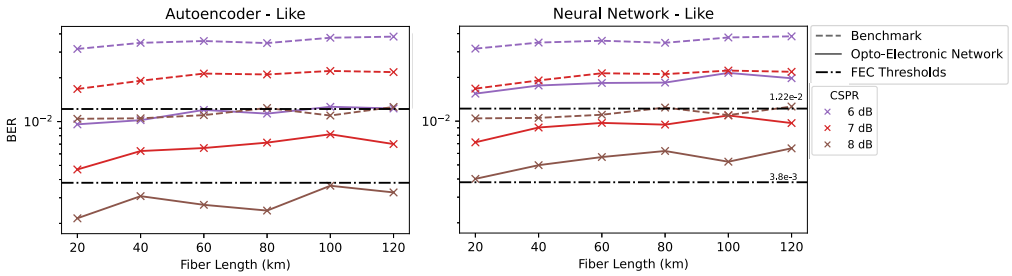

Fig. 8. results for the testing set showing BER vs fiber length for the autoencoder-like configuration (left) and the neural network-like configuration (right). Results from the opto-electronic networks are shown as solid lines, while results from the benchmark pipeline are shown as dashed lines. The colors refer to the different CSPRs which range between 6 dB and 8 dB.

As the opto-electronic networks enhance the receiver's linearity, lower BERs are achieved, with around 3-4 times improvement seen from the autoencoder-like configurations and around 2-3 times improvement from the neural network-like configuration. Furthermore, Fig. 8 indicates that similar BER results were achieved by the autoencoder-like configuration at 2 dB lower CSPRs compared to the benchmark, while a 1 dB improvement was contributed by the neural network-like configuration. These improvements are consistent over the range of lengths investigated in the paper.

Compensation is also evident from the constellation diagrams in Fig. 9, which show signals with a CSPR of 7 dB after traversing 120 km for the benchmark pipeline (left), the autoencoder-like configuration (center), and the neural network -like configuration (right).

Figure 10 shows the impact of evaluating the system at a fiber length of 250 km, i.e. roughly twice as long as the maximum of Fig. 8, so that dispersion accumulation becomes more pronounced. A notable increase in errors is observed at 5 dB and 6 dB, attributed to the failure of the BPS in accurately compensating for the phase noise. While in theory the severity of the phase noise should remain unaffected by the length of the fiber, in practice however, when the

Optics EXPRESS

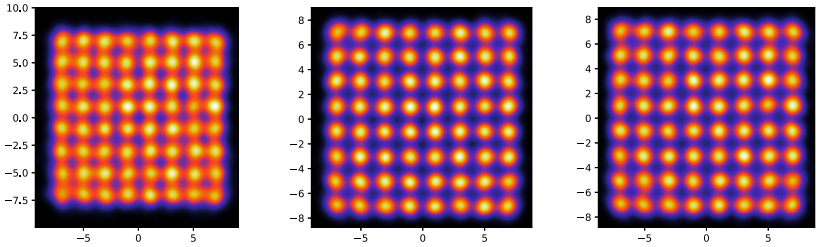

Fig. 9. Constellation diagrams for CSPR 7 dB after 120 km when detected by the benchmark pipeline (left), the auto-encoder like configuration (center), and the neural network -like configuration (right).

dispersion residual errors are significant, the BPS algorithm will be suboptimal and achieve a poor phase compensation. To potentially address this issue, a larger averaging window might improve performance, but this enhancement would demand increased computational resources and memory allocation.

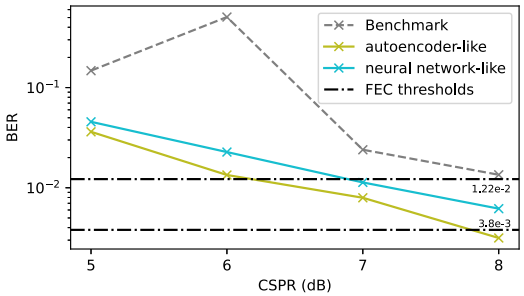

Fig. 10. BER vs CSPR after 250 km transmission for the opto-electronic networks against the benchmark pipeline. A spike in the BER is seen in the benchmark pipeline due to residual phase error.

We then explore the connection between our networks' performances and the receiver's nonlinearity, aiming to discern the extent to which system performance relies on this nonlinearity. The receiver's nonlinear behavior is exhibited at the detector when CSPR levels are low, as well as at the algorithmic portion when the samples per symbol are low. While the CSPR value is an aspect of the signal itself, the sps is easily adapted on the receiver's end. As such, we study the effect of changing the sampling rate of the KK receiver *after* the networks have been trained for operation with 3 sps, and do this for two CSPR levels.

Figure 11 shows the BER vs sps of the testing set at 250 km for both configurations. The sps is varied between more nonlinear (i.e., 2.5 sps) and more linear (i.e., 4 and 6 sps) receiver operation. Note that the lower sps system can exhibit highly nonlinear behavior which, as seen in Fig. 10, may lead to residual phase noise and thus a high BER. As such, to ensure the results of Fig. 11 reflect the effects of changing the nonlinearity of the receiver, we set the linewidth of the laser to zero. The left subplot of Fig. 11 shows the performance of the systems when the CSPR is 6 dB, and thus the overall receiver is more nonlinear compared to the right subplot with a CSPR of 8 dB.

Both network configurations at both CSPRs continue to have good performance at 2.5 sps, resulting in errors 2 to 5 times lower than the benchmark. However, as the sps increases above 3, the neural-network like configuration becomes distorting to the signal and thus performs worse than the benchmark. This is more nuanced when the CSPR is higher, and thus the overall receiver

Optics EXPRESS

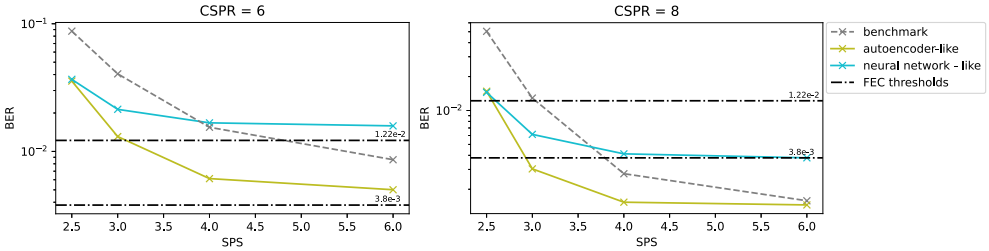

Fig. 11. BER vs SPS for testing pipelines at 250 km. Networks were trained to operate at 3 sps but were tested at sps values between 2.5 (more nonlinear) and 6 (more linear). Both 6 dB CSPR (left) and 8 dB CSPR (right) are shown.

behavior is more linear. In contrast, the autoencoder-like configuration continues to perform better than the benchmark as the sps increases. However, similar to the neural network -like configuration it too deteriorates in performance as the CSPR increasing. These results indicate that the neural network-like configuration is very dependent on the nonlinear behavior from the receiver's sampling rate. The autoencoder-like configuration on the other hand continues to achieve better performance than the benchmark as long as a portion of the KK receiver (i.e., the CSPR) remains nonlinear.

When constructing the two different configurations, several aspects were varied, including the placement of the FFE block, the input nodes of the reservoir, the delay line lengths of the FFE block, and the initialization before backpropagation. These factors make it hard discern which aspect or combination of aspects led to the difference in achievable BERs. Nonetheless, the results of both networks indicate that significant BER gains are achievable through leveraging the autoencoder-like and neural network-like behaviors.

Moreover, the opto-electronic networks reduce the complexity of the KK receiver, especially at higher baud rates, by halving the system's sampling rate, thereby streamlining computationally intensive operations. This reduction in complexity is notable even with the incorporation of the additional FFE component, as this component functions effectively at low sampling rates and can be realized in either analog or digital form.

The reservoir and FFE node count were chosen to offer increased degrees of freedom and higher programmability, although further investigation could focus on minimizing the network costs. These nodes, alongside their programmability, introduce memory that could potentially compensate—at least partially—for dispersion. Although this paper does not explore this aspect, training the reservoir to perform several functions has been explored in other works [33].

Finally, we address two limitations of the opto-electronic network. Since the photonic reservoir replaces an optical bandpass filter, the network training optimizes the reservoir's behavior to achieve suitable cutoff and roll-off characteristics. Consequently, changes in the system's noise profile may necessitate new training. Moreover, the performance remains limited when compared to a fully coherent receiver. Since the opto-electronic neural networks have linear architectures, they are not comparable to a fully capable machine learning network. Rather, the approach in this paper offers a simple alternative designed to boost the performance of a less capable system by harnessing the nonlinear functions inherent in the receiver.

6. Conclusions

We numerically demonstrated the use of an opto-electronic network composed of an integrated photonic reservoir and an electronic feed forward equalizer for the linearization of a KK receiver operating at 3 sps and CSPRs below 9 dB. Up to a 4x reduction in BER was achieved on links of 250 km and shorter, without explicitly training the reservoir on these channels and with seamless

Optics EXPRESS

DSP integration. The opto-electronic network components are fully linear which makes them easily realizable, and they rely on the receiver's nonlinearity to achieve the desired behavior.

Funding. Horizon 2020 Framework Programme (GA101070195, GA860360, GA871330, GA871658); Fonds Wetenschappelijk Onderzoek (G006020N).

Acknowledgements. Parts of this work were performed in the context of the European projects Postdigital (GA860360), Nebula (GA871658), Neoteric (GA871330), Prometheus (GA101070195), and the Belgian FWO project (G006020N)

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. K. Zhong, X. Zhou, J. Huo, *et al.*, "Digital Signal Processing for Short-Reach Optical Communications: A Review of Current Technologies and Future Trends," J. Lightwave Technol. **36**(2), 377–400 (2018).
2. M. Chagnon, "Optical communications for short reach," J. Lightwave Technol. **37**(8), 1779–1797 (2019).
3. I. Alimi, R. Patel, N. Silva, *et al.*, "A review of self-coherent optical transceivers: Fundamental issues, recent advances, and research directions," Appl. Sci. **11**(16), 7554 (2021).
4. H. Chen, N. K. Fontaine, J. M. Gene, *et al.*, "Dual Polarization Full-Field Signal Waveform Reconstruction Using Intensity only Measurements for Coherent Communications," J. Lightwave Technol. **38**(9), 2587–2597 (2020).
5. M. Lyu, W. Shi, and L. Rusch, "SIP-based SSBI cancellation for OFDM," IEEE Photonics J. **11**(5), 1–13 (2019).
6. A. Mecozzi, C. Antonelli, and M. Shtaif, "Kramers–Kronig receivers," Adv. Opt. Photonics **11**(3), 480–517 (2019).
7. A. Mecozzi, C. Antonelli, and M. Shtaif, "Kramers–Kronig coherent receiver," Optica **3**(11), 1220–1227 (2016).
8. A. Lugnan, A. Katumba, F. Laporte, *et al.*, "Photonic neuromorphic information processing and reservoir computing," APL Photonics **5**(2), 020901 (2020).
9. A. Katumba, X. Yin, J. Dambre, *et al.*, "A Neuromorphic Silicon Photonics Nonlinear Equalizer for Optical Communications with Intensity Modulation and Direct Detection," J. Lightwave Technol. **37**(10), 2232–2239 (2019).
10. E. Gooskens, F. Laporte, C. Ma, *et al.*, "Wavelength dimension in waveguide-based photonic reservoir computing," Opt. Express **30**(9), 15634–15647 (2022).
11. S. Sackesyn, C. Ma, J. Dambre, *et al.*, "Experimental realization of integrated photonic reservoir computing for nonlinear fiber distortion compensation," Opt. Express **29**(20), 30991–30997 (2021).
12. E. Gooskens, S. Sackesyn, J. Dambre, *et al.*, "Experimental results on nonlinear distortion compensation using photonic reservoir computing with a single set of weights for different wavelengths," Sci. Rep. **13**(1), 21399 (2023).
13. C. Mesaritakis, K. Sozos, D. Dermanis, *et al.*, "Spatial Photonic Reservoir Computing based on Non-Linear Phase-to-Amplitude Conversion in Micro-Ring Resonators," in *Optical Fiber Communication Conference (OFC)* (2021), pp. 1–3.
14. C. Huang, S. Fujisawa, T. F. de Lima, *et al.*, "Demonstration of photonic neural network for fiber nonlinearity compensation in long-haul transmission systems," *Optical Fiber Communication Conference (OFC)* (2020) pp. 1–3.
15. D. Orsuti, C. Antonelli, A. Chiuso, *et al.*, "Deep Learning-Based Phase Retrieval Scheme for Minimum-Phase Signal Recovery," J. Lightwave Technol. **41**(2), 578–592 (2023).
16. D. Orsuti, M. Santagiustina, A. Galtarossa, *et al.*, "Edge-Carrier-Assisted Phase Retrieval At Low CSPR and Low Dispersion Diversity With Deep Learning," J. Lightwave Technol. **42**(7), 2285–2295 (2024).
17. X. Li, S. A. N. Haohua, H. J. I. Onglin, *et al.*, "Deep-learning-enabled high-performance full-field direct detection with dispersion diversity," Opt. Express **30**(7), 11767–11788 (2022).
18. S. Masaad, E. Gooskens, S. Sackesyn, *et al.*, "Photonic reservoir computing for nonlinear equalization of 64-QAM signals with a Kramers-Kronig receiver," Nanophotonics **12**(5), 925–935 (2023).
19. C. Fullner, M. M. H. Adib, S. Wolf, *et al.*, "Complexity Analysis of the Kramers–Kronig Receiver," J. Lightwave Technol. **37**(17), 4295–4307 (2019).
20. P. J. Freire, A. Napoli, B. Spinnler, *et al.*, "Neural Networks-Based Equalizers for Coherent Optical Transmission: Caveats and Pitfalls," IEEE J. Sel. Top. Quantum Electron. **28**(4), 1–23 (2022).
21. H. Jaeger, "Foreword," in *Reservoir Computing Theory, Physical Implementations, and Applications* (Springer, 2016).
22. G. Tanaka, T. Yamane, J. B. Héroux, *et al.*, "Recent advances in physical reservoir computing: A review," Neural Networks **115**, 100–123 (2019).
23. G. Van der Sande, D. Brunner, and M. C. Soriano, "Advances in photonic reservoir computing," Nanophotonics **6**(3), 561–576 (2017).
24. S. Sackesyn, C. Ma, A. Katumba, *et al.*, "A Power-Efficient Architecture for On-Chip Reservoir Computing," *Artificial Neural Networks and Machine Learning (ICANN)* (2019).
25. K. Vandoorne, P. Mechet, T. Van Vaerenbergh, *et al.*, "Experimental demonstration of reservoir computing on a silicon photonics chip," Nat. Commun. **5**(1), 3541 (2014).
26. S. Li, C. Häger, N. Garcia, *et al.*, "Achievable Information Rates for Nonlinear Fiber Communication via End-to-end Autoencoder Learning," *European Conference on Optical Communication (ECOC)* (2018) pp. 1–3.

Optics EXPRESS

27. "VPItransmissionMaker," Availabe at https://www.vpiphotonics.com/Tools/OpticalSystems/
28. F. Laporte, J. Dambre, and P. Bienstman, "Highly parallel simulation and optimization of photonic circuits in time and frequency domain based on the deep-learning framework PyTorch," Sci. Rep. **9**(1), 5918 (2019).
29. B. A. Wichmann and I. D. Hill, "Algorithm AS 183: An Efficient and Portable Pseudo-Random Number Generator," Appl. Stat. **31**(2), 188–190 (1982).
30. N. Hansen, "The CMA evolution strategy: A comparing review," in *Towards a New Evolutionary Computation* (Springer, 2006) Vol. 192.
31. I. Sutskever, J. Martens, G. Dahl, *et al.*, "On the importance of initialization and momentum in deep learning," *Proceedings of the 30th International Conference on Machine Learning* (2013) pp. 2176–2184.
32. "Softplus," Available at https://pytorch.org/docs/stable/generated/torch.nn.Softplus [accessed: April 20, 2024].
33. A. Zelaci, "Thesis: Photonic reservoir computing for equalizing phase-modulated signals," Ghent University (2023).