



## Reservoir-computing based equalization with optical pre-processing for short-reach optical transmission

Da Ros, Francesco; Magalhaes Ranzini, Stenio; Bülow, Henning

*Published in:*

I E E E Journal on Selected Topics in Quantum Electronics

*Publication date:*  
2020

*Document Version*  
Peer reviewed version

[Link back to DTU Orbit](#)

*Citation (APA):*

Da Ros, F., Magalhaes Ranzini, S., & Bülow, H. (Accepted/In press). Reservoir-computing based equalization with optical pre-processing for short-reach optical transmission. *I E E E Journal on Selected Topics in Quantum Electronics*.

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# Reservoir-computing based equalization with optical pre-processing for short-reach optical transmission

Francesco Da Ros, *Member, IEEE*, Stenio M. Ranzini, *Student member, IEEE*, Henning Bülow, and Darko Zibar

**Abstract**—Chromatic dispersion is one of the key limitations to increasing the transmission distance-rate product for short-reach communication systems relying on intensity modulation and direct detection. The available optical dispersion-compensation techniques have lost favor due to their high impact on the link loss budget. Alternative digital techniques are usually power-hungry and introduce latency. In this work, we compare different digital, optical and joint hybrid approaches to provide equalization and dispersion compensation for short-reach optical transmission links. Reservoir computing is reviewed as a promising technique to provide equalization with memory in an easily trainable fashion, and the properties of the reservoir network are directly linked to system performance. Furthermore, we propose a new hybrid method relying on reservoir computing combined with a simple signal pre-conditioning stage directly in the optical domain. The optical pre-processing uses an arrayed waveguide grating to split the received signal into narrow sub-bands. The performance of the proposed scheme is thoroughly characterized both in terms of reservoir properties and appropriate pre-processing. The benefits are numerically demonstrated for 32-Gb/s on-off keying signal transmission, and show an increase in reach from 10 km to 40 km, corresponding to 400%, compared with more complex digital-only techniques.

**Index Terms**—Short-reach transmission, direct-detection, reservoir computing, signal equalization, chromatic dispersion compensation.

## I. INTRODUCTION

NEW technologies massively relying on a high-interconnectivity of relatively simple devices have shifted the strain caused by an ever-increasing demand for data rates from long-haul systems to short-reach communications, i.e., inter- and intra-datacenter. Most of the information transfer happens on a machine-to-machine basis, with only a small fraction of the data reaching the human end-user. Unlike for long-haul systems where the energy and complexity can be spread over a large number of users, for short-reach communication, the transmission rate and reach need to be increased by focusing on low-complexity solutions. Medium and long haul systems are mainly limited by amplification noise and

nonlinear effects, as linear impairments can be compensated at the receiver side thanks to coherent detection providing access to the full optical field. However, even though coherent detection is already discussed for short-reach [1], it is not yet considered commercially viable, and only intensity modulation with direct detection (IM/DD) is deployed. Losing the information carried by the field's phase makes chromatic dispersion (CD) the major obstacle to extending the transmission reach. Several techniques are available for compensating for the inter-symbol interference (ISI) induced by CD, by acting in the optical domain [2], [3], in the digital/electric domain [4], [5], [6], [7], [8], [9], [10], [11], [12] or by considering a joint optoelectronic approach [13], [14], [15], [16], [17], [18], [20]. Optical dispersion compensation techniques mainly rely on negative dispersion media, such as dispersion-compensating fibers (DCFs) or fiber Bragg gratings (FBGs). These media are able to compensate fully for the accumulated dispersion, however, at the expense of latency (mainly DCFs) and, more importantly, of significant power loss. As power loss directly translates into lower received signal quality, in-line optical dispersion compensation is typically avoided in short-reach systems. Moving into the digital domain to perform the compensation requires to consider the CD to be a nonlinear effect due to its transformation through the square law of a photodetector (PD). Additionally, intensity detection introduces nonlinear mixing of signal and noise, requiring a computationally heavy estimation of a likelihood function and thus making a maximum a posteriori probability (MAP) detection strategy less practical. The proposed methods, therefore, rely on rather complex approaches such as maximum likelihood sequence estimation (MLSE) [5], Volterra equalization [4], fully-fledged feed-forward neural networks (FNNs) [9], [12] and recurrent neural networks (RNNs) [11]. Promising results have been shown by applying these methods. However, the improvements in reach are relatively small to justify the overall complexity of training FNNs and RNNs, as they require backpropagating through the full network. As a practical channel is not static and physical drifts (e.g., due to temperature fluctuations) periodic retraining will be required. It is therefore desirable to develop schemes where the training phase can be performed with low complexity. This may be achieved by moving to hybrid optoelectronic approaches, which leverage on sharing the complexity between optical and electrical domain, at the expense of only small added optical power loss, as we have shown in [20]. Additionally, a few recent investigations have also looked into applying a new machine learning paradigm known as reservoir computing (RC), which allows avoiding to train the hidden part (reservoir) of the neural network [13],

F. Da Ros and S.M. Ranzini have contributed equally to the presented work. This work is supported by the European Research Council through the ERC-CoG FRECOM project (grant agreement no. 771878), by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie (grant agreement No 766115), and by the Villum Fonden Young Investigator Programme (grant no. 29344). The authors would like to thank the reviewers for their constructive suggestions which improved the overall quality of the manuscript.

F. Da Ros, S.M. Ranzini, and D. Zibar are with the Department of Photonics Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark, emails: fdro, smara, dazi@fotonik.dtu.dk.

H. Bülow is with Nokia Bell Labs, Stuttgart, Germany, email: henning.buelow@nokia-bell-labs.com

Manuscript received April 19, 2005; revised August 26, 2015.

[14], [15], [17], [18], [19]. By providing an optoelectronic implementation of the system, significant gains in transmission reach have been demonstrated. Further analysis is, however, needed to understand the benefits of applying the RC approach to short-reach channel equalization, especially when considering optoelectronic equalization. The works reported so far have either mainly relied on linear reservoirs [13] or dealt with Kerr nonlinear effects other than CD [18].

In this work, we extend our hybrid optoelectronic approach of [20] by investigating the additional benefits of using RC rather than FNNs for 32-Gb/s on-off keying (OOK) transmission. A thorough analysis of the reservoir properties (memory/size/nonlinear dynamics) is reported and directly related to achievable system performance gains. The proposed scheme is benchmarked against digital techniques (MLSE and FNN), as well as hybrid approaches showing reach extensions over 400%. Overall different receiver structures are considered: (i) direct photodetection followed by electronic post-processing; (ii) an optical pre-processing of the signal before square-law photodetection; (iii) hybrid approaches considering both optical pre-processing and electronic post-processing.

The paper is organized as follows. In Section II, the concept of reservoir computing is reviewed and its application to signal equalization is discussed, focusing on the short-reach optical channel introduced in Section III. In Section IV, digital RC is compared to alternative digital equalization methods (MLSE and FNN) and the impact of reservoir parameters such as memory and reservoir size on the system performance are discussed. In Section V, optical pre-processing schemes are introduced and the improvement they enable is discussed. In Section VI, digital RC is combined with optical pre-processing in a hybrid optoelectronic approach. The optical pre-processing and RC post-processing stages are appropriately characterized and the full benchmark of the promised approach (in both a low and high complexity form) is also provided. The conclusions are summarized in Section VII.

## II. RESERVOIR COMPUTING

A recurrent neural network (RNN) is a neural network architecture with feedback connections, which allows for temporal information to be processed, making it very powerful to compensate for nonlinear effects with memory such as CD-induced ISI after direct-detection. The increased computational power compared to FNNs, however, comes at the price of a harder training process both in terms of time required and possible pitfalls. Using the gradient descent methods commonly used for FNN is more challenging due to vanishing-gradients problems (especially when considering first-order optimizations) [21] and discontinuities in the error surface [22]. More complex methods can then be applied for training, but in general, training RNNs for problems with long-range temporal dependencies is a rather challenging task [23], [24], [25].

The concepts of echo state network (ESN) [26] and liquid state machine (LSM) [27] were proposed independently as simpler alternatives to overcome these training challenges. The former is mostly defined with a discrete-time sigmoid-unit

network, while the latter focuses on a biological approach for neural modeling. Whereas coming from different paths, the core ideas are similar. The two methods have, therefore, being unified within the framework of reservoir computing (RC) [28]. The theoretical formulation of RC is reviewed in the following, after which a brief summary of the state-of-the-art on RC applications to signal equalization tasks is reported.

### A. Mathematical formulation

The RC network consists of three standard parts - an input layer, a hidden network, and an output layer - as shown in Fig. 1.

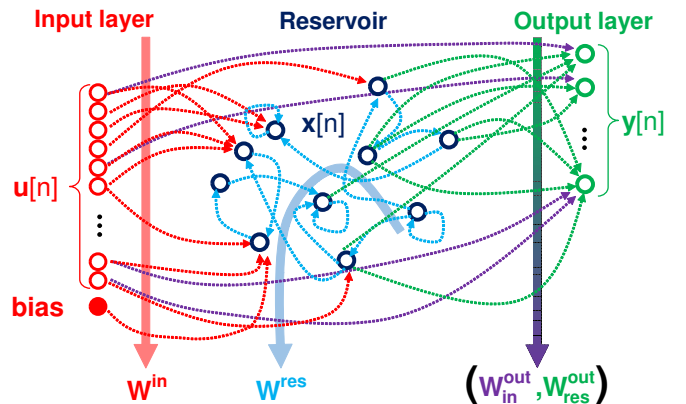


Fig. 1. General architecture of an RC network.

The input signal vector  $\mathbf{u}[n] = [u_1[n], \dots, u_N[n]]^T$  consists of a real  $N$ -dimensional vector at the discrete-time instant  $n$ , where  $N$  represents the number of inputs.  $\mathbf{u}[n]$  is injected into the reservoir through a linear transformation given by the  $M \times (N + 1)$ -dimensional  $\mathbf{W}^{in}$ , where  $N + 1$  takes into account a bias component and  $M$  represents the number of nodes within the reservoir. The weights in  $\mathbf{W}^{in}$  are randomly generated, commonly by drawing samples uniformly distributed between  $-1$  and  $1$ , or from a standard normal distribution [29]. The bias is added to make the signal operate in different regions of the nonlinear activation function of the hidden/computational part of the network [29].

This second part of the architecture, i.e., the hidden network, is structured as for a conventional RNN and it is initialized with random weights and sparse interconnections between hidden nodes. Unlike for RNNs, however, such weights and interconnections, as well as the input weights, are not trained but kept fixed to their random initial values. This hidden network is commonly referred to as a *reservoir* and it transforms the input signal from one state to another through nonlinear dynamics. Whereas the lack of training of the reservoir may seem to cause a strong limitation to the computing capacity of RC, the use of randomly generated connections is not uncommon in the machine-learning toolbox. Random projections are extensively used for hashing and dimensionality reduction [30], [31], and randomized weights are used for feed-forward neural networks within the extreme learning machine approach [32]. By removing the freedom to train the reservoir interconnection, a larger network compared

to fully-trained RNNs may be required, however, at the key advantage of a much simpler training procedure, as will be discussed in the following. Additionally, multiplications by constants can be practically implemented by using shifters and adders, resulting in a lower complexity implementation, e.g. as shown in [33]. The reservoir is described by the  $M \times M$  square matrix  $\mathbf{W}^{res}$ . Such a matrix defines the number of neurons and how they are connected. The weights of the interconnections between reservoir states, i.e., the elements  $\{w_{i,j}^{res}\}$  of  $\mathbf{W}^{res}$  are commonly drawn from uniform or normal distributions [29]. Through  $\mathbf{W}^{in}$ ,  $\mathbf{W}^{res}$  and the chosen activation function  $f_{act}$ , the evolution of the reservoir  $\mathbf{x}[n] = [x_1[n], \dots, x_M[n]]^T$  at the time instant  $n$  can be calculated as a function of the reservoir at the previous time instant ( $\mathbf{x}[n-1]$ ) and the input  $\mathbf{u}[n]$ , as in (1).

$$\mathbf{x}[n] = \alpha \cdot f_{act}(\mathbf{W}^{in} \cdot \mathbf{u}[n] + \mathbf{W}^{res} \cdot \mathbf{x}[n-1]) + (1 - \alpha) \cdot \mathbf{x}[n-1], \quad (1)$$

The leaking rate  $\alpha$  in (1) is used to emulate an exponential decay within the reservoir to both better satisfy the echo state property and mimic realistic physical implementations [26], [34].

At the reservoir's output, a linear transformation is applied to the states of the reservoir  $\mathbf{x}[n]$  and input signal  $\mathbf{u}[n]$  yielding the network's readout  $\mathbf{y}[n]$ . The output weights are the only components of the overall RC architecture which are trained and fulfill the task of transforming the reservoir states and input signal into the desired output. The RC output  $\mathbf{y}[n] = [y_1[n], \dots, y_K[n]]^T$  is calculated through the  $K \times M$ -matrix  $\mathbf{W}_{res}^{out}$  and the  $K \times (N+1)$  matrix  $\mathbf{W}_{in}^{out}$ .  $\mathbf{W}_{res}^{out}$  defines the weights and interconnection between the  $M$  reservoir states and the  $K$  network outputs, while  $\mathbf{W}_{in}^{out}$  describes a direct connection between input and output layers. The evolution of the output layer then follows (2).

$$\mathbf{y}[n] = \mathbf{W}_{res}^{out} \cdot \mathbf{x}[n] + \mathbf{W}_{in}^{out} \cdot \mathbf{u}[n], \quad (2)$$

In general, additional feedback from the output layer into the reservoir ( $\mathbf{W}_{out}^{res}$  of size  $K \times M$ ) can also be considered [26]. This feedback is here neglected as it is rather uncommon and strictly not necessary as a network including it can be usually re-mapped into one where such feedback is not explicitly defined [26].

The key advantage of the RC approach relies on the fact that, once a suitable  $\mathbf{W}^{res}$  is constructed, the evolution of the reservoir states depends only on the inputs (see (1)), and training is only applied to optimize the output weights ( $\mathbf{W}_{res}^{out}$ ,  $\mathbf{W}_{in}^{out}$ ). Such an optimization does not require the complex and computationally expensive back-propagation through the whole network, as is the case for RNNs (back-propagation-through-time algorithm) or even simpler FNNs (back-propagation algorithm). The training can be performed with a single linear regression operation, possibly with ridge regularization included, aiming at minimizing the squared error of the loss function. Given a target set matrix  $\mathbf{Y}_{target} =$

$\mathbf{Y}_1^{\bar{n}} := [\mathbf{y}[1], \dots, \mathbf{y}[\bar{n}]]$  defined for times  $n = [1, \dots, \bar{n}]$ , the output weights can be calculated through:

$$[\mathbf{W}_{res}^{out}, \mathbf{W}_{in}^{out}] = (\mathbf{A}^T \cdot \mathbf{A} + \lambda \mathbf{I})^{-1} \cdot \mathbf{A}^T \cdot \mathbf{Y}_{target}, \quad (3)$$

with  $\mathbf{A} = [\mathbf{X}_1^{\bar{n}}; \mathbf{U}_1^{\bar{n}}; \mathbf{B}]$ , where the last element  $\mathbf{B}$  represents the bias term;  $\lambda$  the ridge regularization factor; and  $\mathbf{I}$  an identity matrix of size  $M+N+1$ . The state and input matrices follow the same definition as for  $\mathbf{Y}_1^{\bar{n}}$ :  $\mathbf{X}_1^{\bar{n}} := [\mathbf{x}[1], \dots, \mathbf{x}[\bar{n}]]$  and  $\mathbf{U}_1^{\bar{n}} := [\mathbf{u}[1], \dots, \mathbf{u}[\bar{n}]]$ , respectively.

For the RC approach to provide effective computing power, a set of macroscopic reservoir hyper-parameters needs to be appropriately chosen, in particular, the activation function  $f_{act}$ , and the characteristics of the matrix  $\mathbf{W}^{res}$ , especially its sparsity and the resulting stability of the RC system. The latter is commonly defined in terms of the reservoir *spectral radius* ( $\rho$ ) defined as the maximum eigenvalue of  $\mathbf{W}^{res}$  [26].

Concerning  $f_{act}$ , the properties of ESN, including its universal approximation property, are strictly defined for sigmoid functions [26], [34], [35], [36]. However, in a number of demonstrations [13], [33], using alternative functions, even linear functions, to replace a sigmoid has been shown to be computationally effective. Linear activation functions have also been shown to provide a better memory capacity, then nonlinear functions [28], [37]. However, the performance of the network depends on the input and task at the end, and for general applications, a linear operator may not always be optimal [28], [37].

Regardless of the choice of the activation function, a fully randomly generated reservoir may not always provide an effective high-dimensional transformation allowing to compute the task at hand. The matrix  $\mathbf{W}^{res}$  needs to be properly constructed following a number of guidelines [34]. First of all, unlike what may be common for RNN, the reservoir interconnections are usually chosen with a high degree of sparsity. Whereas RNNs need full interconnectivity to ensure the universal approximation property [21], in RC on average 10 or fewer interconnections per node are recommended [26].

More importantly, it has been shown that the *echo state property*, i.e., a fading memory property for which the effect of input at time  $t_1$  vanishes for times  $t \gg t_1$ , is necessary to ensure computational capacity [26]. This property can be directly related to the spectral radius  $\rho$  of the matrix  $\mathbf{W}^{res}$ . Considering an activation function with a unitary slope around the origin - such as the  $\tanh(\cdot)$  considered in this work - a good rule of thumb is to scale  $\mathbf{W}^{res}$  such that the  $\rho \leq 1$ . This condition is only necessary and not sufficient to ensure the echo state property of a reservoir. Generally, the spectral radius requires tuning to the memory and nonlinear dynamics required by the specific task at hand. The spectral radius thus provides some insight on the expected behavior of the reservoir and it is one of the key parameters that will be investigated within this work, together with other macroscopic properties of a reservoir such as its size and memory. A general criterion to construct an effective reservoir to solve given tasks is rather challenging to define and a discussion in this direction focused on the relation between  $\rho$  and stability conditions and can be found in [34]. In this work, we focus on the specific task

signal equalization for short-reach IM/DD transmission and relate the properties of the reservoir used for nonlinear signal equalization to the received signal quality.

### B. RC for signal equalization: state-of-the-art

RC for channel equalization has been considered in a number of recent works, especially for a nonlinear wireless transmission [33], [38]. Its use for an optical fiber channel has instead been considered only more recently. In one of the first works targeting this task [18], a time-delayed reservoir implementation relying on a single nonlinear node and delayed feedback has been experimentally analyzed. A semiconductor laser provides the nonlinear node and the reservoir is able to effectively improve the quality of a 4-pulse amplitude modulated (PAM) signal affected by simulated optical impairments, focusing on the impact of Kerr nonlinearity. The performance of a time-delayed RC similar to [18] is then compared with linear digital signal processing techniques in [15] by applying it to experimentally measured 4-PAM waveforms after fiber transmission. Alternatively to time-delayed schemes, in [13], [14], a linear passive reservoir with a second-order nonlinearity only at the output layer is used to equalize an OOK signal after fiber propagation by focusing on the impact of memory effect due to CD. Preliminary work on a nonlinear all-fiber optics implementation has also been proposed in [16]. The simulation results showed very promising performance improvements for equalization of quadrature amplitude modulated (QAM) signals affected by Kerr nonlinearity by using an RC scheme allowing processing both electrical field quadratures. The results of [16] assume a static coherent transmission channel. As coherent transmission is inherently characterized by a fast time-varying channel, adaptive equalization is required to track effects such as polarization rotations. This work, therefore, focuses on the static IM/DD channel, where less frequent training may be sufficient to compensate for drifts in practical systems.

All the approaches mentioned have focused on the photonic implementation of RC. As the signals to equalize are already in the optical domain, optical or optoelectronic implementations of RC are of particular interest. However, the strong potential for integration and thus the implementation of a large reservoir, has driven a strong focus on photonics RC also for solving alternative tasks, e.g., spoken digit classification [39], [40], image classification [41], and nonlinear system prediction [42], [43], with impressive and promising demonstrations being reported. One of the very first approaches to optoelectronic RC was proposed in [39]. The authors showed simulation results for a semiconductor optical amplifier (SOA)-based neuron. Later, in [44], experimental implementation with passive nodes is provided and compared with numerical results with SOA-based neuron and sigmoid functions, showing similar performance for spoken digit classification. A significant effort has also been dedicated to mimic typical sigmoid nonlinear functions using photonics circuits, e.g., using SOAs [45] and quantum-dot lasers [46]. A comprehensive overview of the most significant contributions can be found in [19], [47], [48].

Regardless of the fast progress experienced by this new field, a systematic analysis of the impact of the reservoir prop-

erties on the performance for a specific task, e.g., equalization efficacy, is still missing. Additionally, fully optoelectronic implementations are still strongly affected by the system loss (e.g., approx. 5 dB in [14], the equivalent of 10 km of additional transmission), which significantly degrade the enhancement in the signal quality by decreasing the link power budget. Electronic implementations of RC for satellite communications have already been shown to provide complexity advantages over standard equalization techniques, e.g., Volterra equalization, with negligible loss of performance [33]. Along this direction, the following sections will compare digital implementations of RC with more standard digital equalization methods such as MLSE and FNN, as well as proposing to use low-complexity optical pre-processing to input a better-conditioned signal into the digital reservoir.

## III. SHORT-REACH TRANSMISSION SYSTEM

As already introduced, the scenario targeted by this investigation is an IM/DD transmission system, numerically modeled by the setup shown in Fig. 2. At the transmitter, a random stream of  $2^{18}$  bits is generated and pulse shaped with root-raised cosine (RRC, 0.1 roll-off) at 8 samples/symbol, yielding a 32-GBd OOK signal. The electrical signal is converted to the optical domain by an ideal Mach-Zehnder modulator (MZM) that encodes the signal onto the optical field of an ideal laser source. The nonlinearity of the MZM transformation is neglected as in [20] in order to focus only on the impact of the dispersive fiber channel. The optical signal is then injected into the transmission channel consisting of standard single-mode fiber (SSMF). As the focus is on compensating for CD-induced ISI after detection, the SSMF transmission is modeled only as a linear lossless dispersive element with a dispersion parameter  $D = 16.4$  ps/nm/km. An additive white Gaussian noise (AWGN) source is used to degrade the signal, e.g., simulating an optical pre-amplifier at the receiver. The signal-to-noise ratio (SNR) in the optical domain and the transmission fiber length (accumulated CD) are varied throughout the analysis to test and compare the performance of different receiver schemes. The bit error ratio (BER) is then computed as a function of SNR and transmission distance. The different schemes either involve (i) photodetection and digital post-processing (Section IV) or (ii) optical pre-processing and photodetection (Section V), or (iii) hybrid approaches considering optical pre-processing, photodetection and digital post-processing (Section VI). Digital and optical processing operate at 8 samples/symbol, with the only exception of the maximum likelihood sequence estimation (MLSE)-based detection operating at 1 sample/symbol. For all other receivers, the sampling instant is then optimized directly at the BER stage.

The performance metric considered to compare the receiver configurations under test is the received SNR required for a BER at the KP4 forward error correction (FEC) threshold ( $\text{BER} = 2.24 \cdot 10^{-4}$  [49]). The receivers will then be benchmarked considering the SNR penalty, i.e., the difference in required receiver SNR, with respect to a simple reference receiver. One of the simplest possible receivers for IM/DD

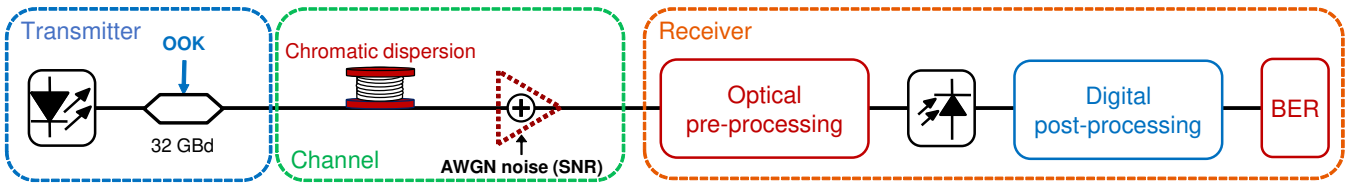


Fig. 2. Short-reach transmission setup under investigation, including optical pre-processing and digital post-processing.

systems consists of a 40-GHz optical band-pass filter, a photodetector (PD), an RRC filter, and a threshold detector, i.e., with no equalization performed on the electrical signal. This receiver scheme will be considered as a reference receiver throughout our analysis. Finally, all PDs are modeled as noiseless square-law elements followed by a low-pass Bessel filter and, for each BER analysis,  $5 \times 2^{18}$ -symbol sequences are transmitted and the BER values are averaged to increase the statistical relevance of the results.

#### IV. DIGITAL POST-PROCESSING

In this Section, RC will be benchmarked against more common nonlinear digital equalization/detection techniques and the impact of macroscopic reservoir properties on the system performance will be discussed. The three different digital processing schemes considered are sketched in Fig. 3 and consist of an MLSE detector, a fully-connected FNN equalizer, and an RC equalizer. Details on the specific implementations considered in this work are provided in the following subsections for each method.

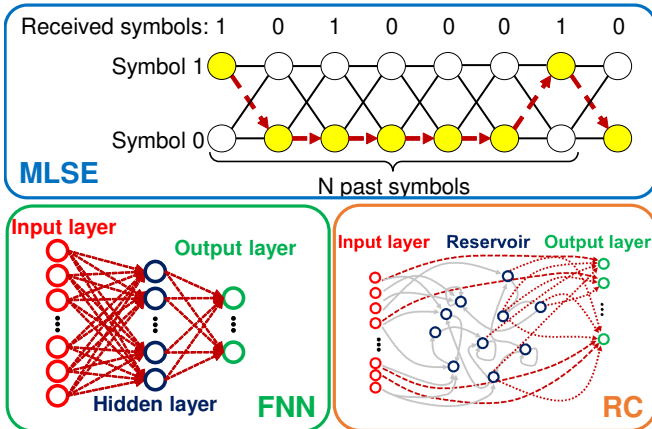


Fig. 3. Digital post-processing techniques benchmarked within this work: the dark red lines show the connections requiring training.

##### A. Maximum likelihood sequence estimation

An MLSE receiver is based on constructing a trellis of all the possible received sequences up to a target memory (in our case 7 symbols) and on estimating the sequence. This process yields the maximum likelihood of the transition probabilities of the channel, which depends on the probability density functions of the sequence. This approach is particularly effective in providing equalization for channels with memory. However, its complexity scales exponentially with the memory

of the channel and the size of the symbols' alphabet, making it quickly unpractical to implement [6].

##### B. Feed-forward neural networks

FNNs have started to be considered recently for the equalization of IM/DD systems [9], [12]. Whereas inherently without memory, time-dependent processing can be partially mimicked by inputting multiple symbols into the FNN in a sliding window approach. The FNN proposed in [20] receives a 5-symbol (40-sample) sliding window at the input layer and consists of a single hidden layer network with 16 hidden nodes and 8 output nodes (8 samples corresponding to 1 symbol). The FNN is trained using back-propagation, making the process relatively time-consuming even for such a simple network. In order to provide a fair comparison, the number of hidden layers and nodes of the FNN has been optimized based on cross-entropy training, as discussed in [20].

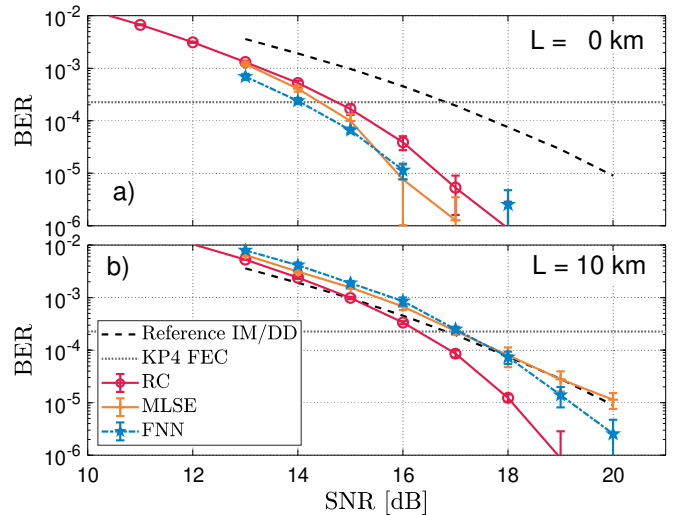


Fig. 4. Performance comparison between digital post-equalization techniques.

##### C. Reservoir computing

The implemented RC architecture follows the general considerations discussed in Section II. The specific implementation is described in detail in Algorithm 1. For all the reservoirs considered, input weights are drawn from a uniform distribution  $\mathcal{U}(-1, 1)$ , reservoir interconnections are drawn from a binary distribution with the probability of no interconnection ( $p(w_{ij}^{res} = 0)$ ) corresponding to the desired degree of sparsity of the reservoir matrix, and the reservoir weights from a standard normal distribution  $\mathcal{N}(0, 1)$ . The input signal



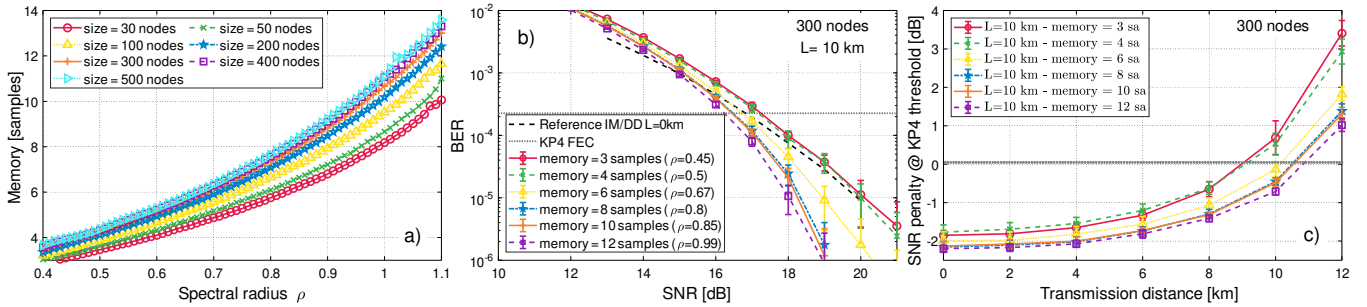


Fig. 5. Impact of reservoir memory on system performance: (a) memory as a function of the reservoir spectral radius ( $\rho$ ), (b) BER vs. SNR curves at 10-km transmission for 300-node reservoirs of different memory values, (c) SNR penalty as a function of the transmission distance for 300-node reservoirs of different memory values.

variance determines the operation regime of the reservoir [29]. Throughout this work, the variance has been scaled to 0.02, following the optimization discussed in Section VI-D. The reservoir size  $M$  and the spectral radius  $\rho$  have been varied throughout the following analysis. The sparsity has then been calculated for each reservoir size  $M$ , such that the number of average interconnected nodes within the reservoir has been kept fixed to 5 nodes. A brief analysis of the impact of such hyper-parameter has shown a minimal impact on system performance. As long as the matrix was sufficiently sparse ( $p(w_{ij}^{res} = 0) \geq 0.8$ ), no significant performance degradation could be seen, in line with [26], [29]. A hyperbolic tangent has been used as activation function  $f_{act}(\cdot) = \tanh(\cdot)$  for all reservoirs considered, and the output layer is trained through linear regression, as discussed in Section II, using 5% of the symbol sequence for training and 95% for testing. A regularization parameter  $\lambda = 10^{-8}$  has been used to avoid numerical instabilities when calculating the matrix pseudo-inversion, as in (3).

---

#### Algorithm 1 Reservoir computing

---

Given  $N$  inputs,  $M$  nodes, and  $K$  outputs:

##### Phase 1 – Initialization

Generate  $\mathbf{W}^{in}$  drawing weights from a uniform distribution  $\mathcal{U}(-1, +1)$

Generate elements of  $\mathbf{W}^{res}$  drawing them from a binary distribution with  $p(0) = \text{sparsity}$

Replace nonzero elements of  $\mathbf{W}^{res}$  drawing them from a normal distribution  $\mathcal{N}(0, 1)$

Re-scale  $\mathbf{W}^{res}$  :  $\rho(\mathbf{W}^{res}) = \bar{\rho}$

##### Phase 2 – Training

Use (1) to calculate  $\mathbf{X}_1^{\bar{n}}$  from  $\mathbf{U}_1^{\bar{n}}$

Use (3) to train  $[\mathbf{W}_{res}^{out}, \mathbf{W}_{in}^{out}]$  from  $\mathbf{X}_1^{\bar{n}}$ ,  $\mathbf{U}_1^{\bar{n}}$ , and  $\mathbf{Y}_{target}$

##### Phase 3 – Testing

Use (1) and (2) to calculate  $\mathbf{y}[n]$  given  $\mathbf{u}[n]$

---

Additionally, in order to increase the statistical relevance of the results, 3 different realizations of input and state matrices have been considered. The performance has therefore been averaged over 3 reservoir realizations, each simulated for  $5 \times 2^{18}$  transmitted sequences, i.e. 5 noise realizations. The

error bars in the following figures show the variance over these 15 simulation runs.

#### D. Performance comparison for IM/DD equalization

Fig. 4 shows the received BER as a function of the channel SNR for the three receivers under test and two different transmission distances: back-to-back in Fig. 4(a) and 10-km transmission in Fig. 4(b). As can be seen, already for back-to-back transmission, an increase in signal quality can be achieved for all the equalization schemes, when compared to the reference receiver. The improvement is expected since the reference receiver includes a low-pass filter but not a matched filter, which is not known for an IM/DD channel. Whereas MLSE and FNN achieve very similar performance, RC is slightly worse. This can be expected considering the significantly simpler, and thus less complete training scheme for RC, which leads to a minor penalty. The roles are, however, reversed considering a signal impaired by the ISI induced by a 10-km propagation. In this case, RC outperforms FNN and MLSE thanks to its inherent recurrence. Given the lower training complexity and improved performance, the potential for an RC-based receiver shown in Fig. 4(b) merits further investigation. Whereas these results show a first simple attempt at an RC-based equalizer, using a 300-node reservoir with  $\rho = 0.9$ , in the following, the impact of the spectral radius (i.e., memory) and reservoir size  $M$  are directly linked to system performance.

#### E. Impact of reservoir memory on RC-based equalizers

As the ability to equalize dispersed signals within a reservoir depends directly on its memory, in this subsection, the reservoir's memory is linked to the system performance by tuning its spectral radius ( $\rho$ ). In particular, Fig. 5(a) shows how the memory evolves with the spectral radius of  $\mathbf{W}^{res}$  for different reservoir sizes. The memory is defined here in terms of the settling time of the reservoir according to the control system theory [50]. A steady-state reservoir is triggered with a step-function at its input. The memory is calculated as the number of samples after which all the reservoir states have converged within less than 2% of their steady-state value (settling time). Up to 1000 different reservoirs have been randomly generated with the same macroscopic parameters,

and the memory values have been averaged over the reservoir realization to clarify the trends. As can be seen in Fig. 5(a), there is a direct relation between the memory of the reservoir and  $\rho$ , as expected from [26]. The memory grows as the spectral radius approaches unity and the growth rate increases as the radius increases beyond one. Additionally, the memory increases with the reservoir size but tends to saturate for large reservoirs beyond 300 nodes.

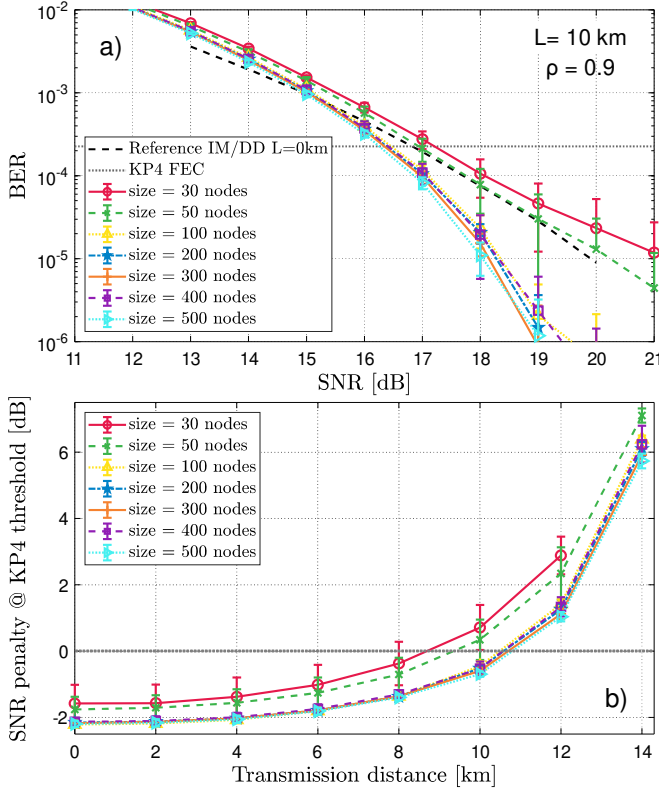


Fig. 6. Impact of reservoir size on system performance: (a) BER vs. SNR curves at 10-km transmission for reservoir of different size but with spectral radius  $\rho = 0.9$ , and (b) SNR penalty as a function of the transmission distance for the same parameters.

Considering a 300-nodes reservoir, the BER vs. SNR curves in Fig. 5(b) shows the performance enabled by an RC-based equalizer for a 10-km transmission and different values of reservoir memory (tuned by controlling  $\rho$ ). The performance improves with increasing the reservoir memory, as the reservoir increases its ability to invert the memory effect in the transmission channel. Such a trend can be generalized for different transmission distances, as shown in Fig. 5(c). The SNR penalty decreases with increasing memory and the gap between curves increases for longer distances.

An alternative metric to evaluate the memory of a reservoir is the linear memory capacity [51]. The relation between the two metrics is currently under investigation.

#### F. Impact of reservoir size on RC-based equalizers

Alternatively to changing the memory of the system by varying the spectral radius, the memory and nonlinear dynamics within the reservoir can be tuned by changing the reservoir size (see Fig. 5(a)). The impact of  $M$  on equalization

is shown in Fig. 6 for reservoirs with  $\rho = 0.9$ . By increasing the reservoir size, the signal will go through more neurons and experience increased nonlinear dynamics. In Fig. 6(a), the BER as a function of the SNR is shown for different reservoir sizes at a 10-km transmission. Larger reservoirs improve the overall performance with a saturation effect taking place as the size grows beyond 100 nodes. As for the memory analysis, the same trends can be extended for different transmission distances, as shown in Fig. 6(b).

## V. OPTICAL PRE-PROCESSING

As discussed in Section II, several attempts at photonic RC have already shown its strong potential for ISI equalization [13], [14], [15], [18]. However, as the loss of current implementations exceeds the loss of in-line optical compensation modules, the decrease in power budget (and thus in signal SNR) may dominate the improvement provided by the photonic RC equalizer.

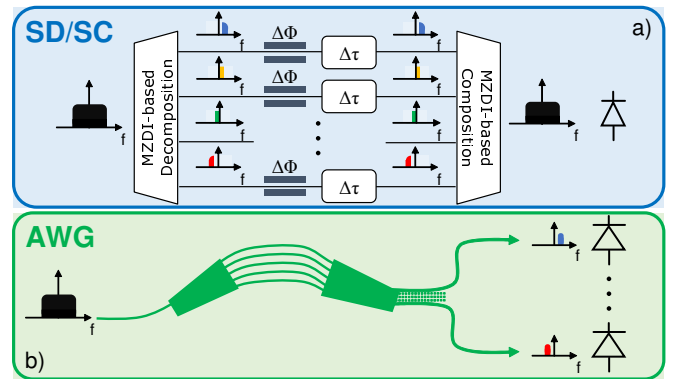


Fig. 7. Optical pre-processing techniques for spectral slicing: (a) MZDI-based spectral decomposition and spectral composition (SD/SC) with slice-wise phase ( $\Delta\Phi$ ) and delay ( $\Delta\tau$ ) shifting element for each slice, placed in between SD and SC; and (b) AWG-based spectral decomposition followed by a bank of PDs.

Alternatively, simpler and lower loss optical pre-processing schemes have already shown to provide performance gain in the numerical analysis of [20]. The proposed schemes have mainly relied on spectrally slicing of the received optical signal. The slicing can be performed either with a series of cascaded Mach-Zehnder delay interferometers (MZDIs) with decreasing free spectral range (FSR) [20] or with an arbitrary waveguide grating (AWG). Concerning the former, MZDIs have the advantage that their transfer function can be easily inverted and the spectrally decomposed slices can be perfectly recomposed into the original signal with a further series of cascaded MZDIs. In [20], the MZDIs are assumed ideal, i.e., with infinite extinction ratio, and we showed how spectral decomposition and composition (“SD/SC”) with MZDIs can be effectively used to compensate for signal dispersion as long as a phase shifter ( $\Delta\Phi$ ) and a delaying element ( $\Delta\tau$ ) are added in between the two spectral operations, as shown in Fig. 7. A phase rotation and delay is applied to each spectral slice independently, in an approach with corresponds to a slice-wise compensation of the phase shift and delay caused by the dispersion accumulated during propagation. The improvement



provided by this technique is shown in Fig. 8 and the higher tolerance to CD is clearly shown, especially for an increasing number of spectral slices. All phase shifters and delays have been jointly trained similarly to the FNN of Section IV, using cross-entropy.

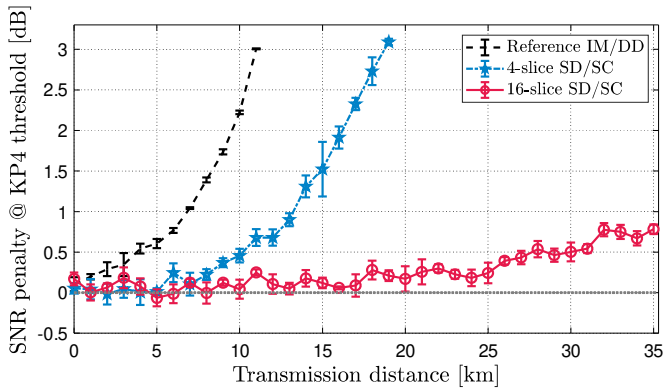


Fig. 8. Performance of spectral decomposition and composition for different number of slices, adapted from [20].

As narrower spectral slices are considered, the slice-wise compensation better approaches the inverse of the CD transfer function, yielding perfect compensation in the limit of an infinite number of slices. As shown in Fig. 8, however, a relatively low number of slices (16), already provides almost ideal compensation for distances approaching 20 km and less than 1 dB penalty for up to 35 km.

Nevertheless, whereas the optical pre-processing is particularly effective in bringing back the performance to the back-to-back level, no signal quality gain (negative SNR penalty) can be achieved as only zero-forcing equalization is performed by the optical circuit, not addressing the presence of noise. Additionally, MZDI may lead to higher insertion loss than AWGs and suffer from stronger polarization dependence, therefore a spectral slicing approach based on an AWG is also considered (Fig. 7(bottom)). An AWG can be modeled as a set of second-order Gaussian filters with a 3-dB bandwidth of each filter/output equal to the frequency separation between two neighboring output filters. Remark that in this case the spectral composition cannot be simply achieved by cascading an AWG-based demultiplexer with an AWG-based multiplexer. Therefore the spectral slices, in this case, are directly detected by a bank of parallel photodetectors and further digital processing is required.

## VI. HYBRID EQUALIZATION SCHEMES

In order to benefit from the higher-dispersion tolerance given by the optical pre-processing as well as the equalization provided by the digital post-processing, in [20] we proposed a hybrid scheme combining both approaches. An optical circuit performing the full spectral decomposition, phase/delay shifting, and spectral composition, however, leads inherently to non-negligible loss. Additionally, the use of spectral decomposition, phase/delay shifting, spectral composition and FNN (labeled “SD/SC+FNN”) requires to jointly train the full weights of the FNN together with the phases and delays in the

optical domain, thus back-propagating through the full digital and optical structure. A sketch of the setup, highlighting the trained quantities (phases, delays and FNN weights) is shown in Fig. 9(a).

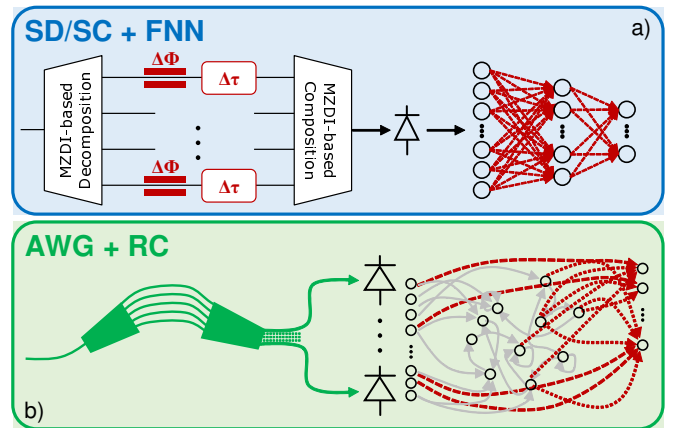


Fig. 9. Hybrid processing techniques: (a) SD/SC pre-processing followed by a photodetector and an FNN-based equalizer and (b) AWG-based spectral decomposition followed by a bank of PDs and an RC-based equalizer. The trainable quantities are marked in dark red, with arrows symbolizing the weights of the connections.

In order to decrease both the optical loss and the training complexity, here we propose to detect directly the spectral slices at the AWG outputs with a bank of PD and inject them into an RC-based equalizer with an increasing number of inputs (Fig. 9(b)). Even though this method requires a number of PDs that scales linearly with the number of spectral slices, the required bandwidth of each PD decreases to match the decrease in the spectral slice bandwidth, allowing to use simpler detectors. Additionally, note that the number of inputs, one per PD, is still much lower (at most 7 for all the slice bandwidth considered) compared to the 40 inputs required to introduce memory for the FNN case.

### A. Optimum number of spectral slices

In order to understand whether this approach yields any improvement compared to digital RC with one 40-GHz signal slice injected into the reservoir, the SNR penalty as a function of the AWG bandwidth is shown in Fig. 10(a). The AWG bandwidth is defined as the half-width half-maximum (HWHM) bandwidth of the AWG transfer function, and as such, the HWHM of each signal spectral slice in input to the reservoir. As the AWG bandwidth is decreased, i.e., the number of spectrally-sliced inputs to the reservoir is increased, the received signal quality is improved, especially for long transmission distances. Decreasing the bandwidth of each reservoir’s input is equivalent to relaxing the memory requirement to invert the effects of CD. The improvement for decreasing AWG bandwidths, however, reaches a minimum at approx. 8 GHz to 10 GHz, after which the signal quality gets significantly degraded. Other than being consistent over different transmission distances, this trend does not depend on the training size, on the input signal normalization, nor on the regularization (within ridge regression) used for calculating

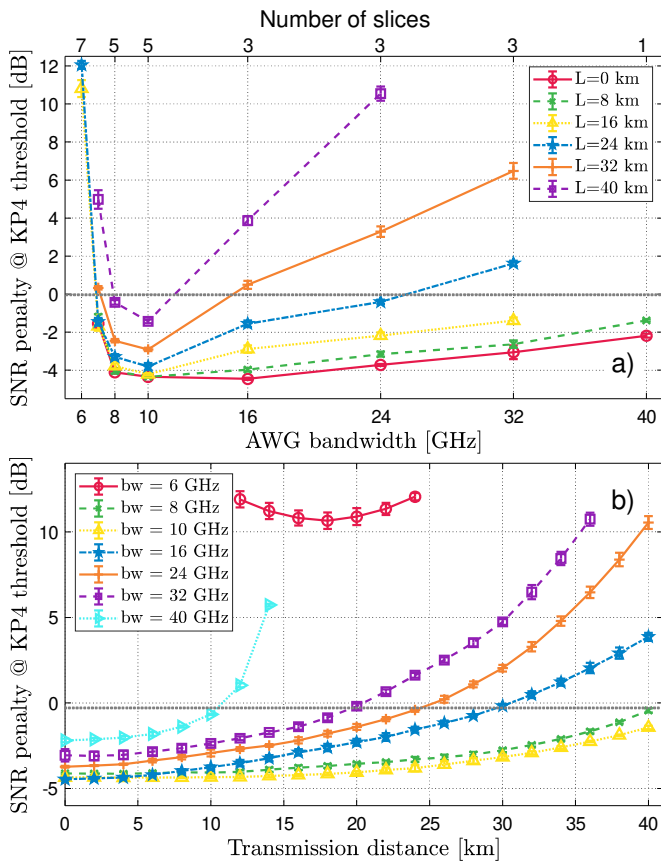


Fig. 10. Impact of AWG bandwidth on system performance for AWG+RC equalization: (a) SNR penalty as a function of the AWG bandwidth (and number of slices) for different transmission distances, and (b) SNR penalty as a function of the transmission distance for different AWG bandwidth. The reservoirs are constructed with 300 nodes and  $\rho = 0.9$ .

the matrix inverse during training. We believe the performance loss may be related to introducing narrow slices with low SNR, i.e., the high-frequency slices carrying less information into the reservoir. Additionally, the narrower low-pass filtering of the AWG may also result in information loss which impairs the ability of RC to provide equalization.

The SNR penalty over a broader range of transmission distances is shown in Fig. 10(b) for a selected choice of AWG bandwidths. As can be seen, the impact of the slice bandwidth increases with the transmission distance becoming more relevant as the channel memory increases.

### B. Impact of reservoir memory on AWG+RC-based equalization

From the previous characterization, optimal performance is achieved for a bandwidth of approx. 8 GHz. The optical pre-processing is then kept fixed to a 5-slice AWGN with 8-GHz of bandwidth/slice throughout the following analysis of the reservoir properties. This analysis follows the same line as in Sec. IV, looking at the impact of reservoir memory and size on the system performance. Fig. 11(a) focuses on the reservoir memory and shows an even stronger impact on the system performance compared to the case of a reservoir with a single 40-GHz input of Fig. 5. The trend, however, follows

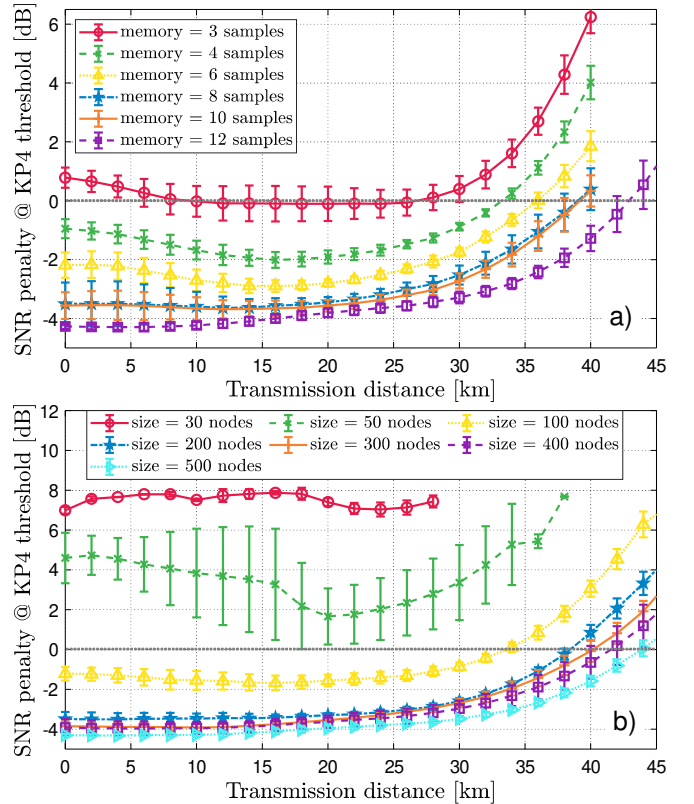


Fig. 11. Impact of reservoir properties on system performance: SNR penalty as a function of the transmission distance for (a) different reservoir memories, and (b) different reservoir sizes. The AWG bandwidth is kept fixed to 8 GHz.

the same direction, a significant improvement is achievable by increasing the reservoir memory but with lower signs of saturation for large memory values.

### C. Impact of reservoir size on AWG+RC-based equalization

As for the reservoir's memory, the size of the reservoir for multiple spectral slices at the RC input has a stronger impact on the performance compared to the case of a single 40-GHz wide reservoir input. This is shown in Fig. 11(b). If the reservoir size is not sufficiently large, the signal quality is actually worsened by the RC equalization. As the input size grows, a larger reservoir is needed to provide a sufficiently nonlinear transformation. Saturation of the improvement can, however, be seen once a reservoir of more than 200 nodes is considered.

### D. Impact of input scaling on AWG+RC-based equalization

The input scaling is another important hyper-parameter to consider, as it affects the operation of the reservoir [21], [29]. A small input scaling leads to a weaker nonlinear operation of the reservoir compared to a larger one. In this work, the variance of the input  $\mathbf{u}[n]$  has been varied and the SNR penalties after 10-km transmission are shown in Fig. 12 for different AWG bandwidths. Note that scaling  $\mathbf{u}[n]$  is equivalent to re-scaling the input-to-reservoir matrix  $\mathbf{W}^{in}$ , and we choose to scale the former for convenience in our

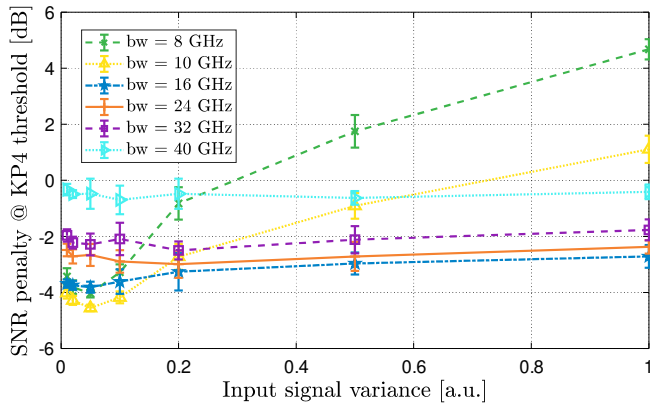


Fig. 12. Impact of input scaling on system performance: SNR penalty as a function of the AWG bandwidth for 10-km transmission.

implementation. No bounding of the input signal, as suggested in [29], was implemented.

The results of Fig. 12 show that the input scaling becomes relevant only as the number of slices increases (i.e. the slice bandwidth decreases). For bandwidths below 16 GHz, an optimum can be seen for a signal variance between 0.02 (values used throughout this work) and 0.05, showing that a weak nonlinear behaviour of the reservoir is sufficient to counteract the weak nonlinearity introduced by dispersion and square-law detection. It is however interesting to see that if the signal variance is significantly increased, i.e. beyond 0.5, the best (yet sub-optimal) performance is achieved for a lower number of wider spectral slices. The analysis reported in Fig. 12 focused on a 300-node reservoir with  $\rho = 0.9$ . However, we confirmed identical trends for spectral radius varying between 0.7 and 0.95 and distances up to 60 km.

#### E. System performance comparison

The overall performance comparison between the proposed hybrid technique with digital-only or alternative hybrid approaches is shown in Fig. 13. Considering a transmission reach definition at the 0-dB SNR penalty threshold compared to the reference receiver, MLSE and FNN achieve a reach of approx. 10 km. This number is slightly increased to 12 km by moving to a 500-nodes RC equalizer but with the key advantage of a much simpler and faster training procedure. A digital-only equalizer is, however, clearly outmatched by comparing it with the hybrid SD/SC+FNN approach of [20]. The joint training of pre- and post-equalization enables extending the reach to approx. 25 km for a 4-slices SD/SC pre-processing stage, i.e., a more than 100% increase over digital equalization. The price for the reach extension is, however, an even more complex training than for MLSE and FNN, as now also the optical layer (phase shifts and delays) needs to be jointly trained.

By moving to the newly proposed hybrid approach (AWG+RC), a reach of 40 km is achieved for a small 300-node reservoir and up to almost 45 km if a larger 500-node reservoir is considered. Other than the reach extension, the AWG+RC approach yields a significantly faster training step. Whereas a thorough complexity analysis is beyond the scope of this work, the training of the SD/SC+FNN requires

approx. 2 hours of processing on a standard GPU, whereas the proposed AWG+RC method can be trained in a matter of minutes on a laptop CPU.

The importance of the optical pre-processing can be clarified by analyzing the performance in the case of injecting 5 identical copies of the received signal, i.e., without spectral slicing, into the reservoir (“5-copies+RC”). In such a case, the reach is comparable to having a simple RC-based equalizer with one input but with a worsening of the SNR penalty for short distances. This result can be explained by considering that having 5 identical copies of the input does not enrich the dynamics of the reservoir but rather reduces them. This comparison clearly shows the potential of RC for equalization in short-reach transmission systems, especially when considering joint processing in both the optical and digital domains. The significant gain reported have been achieved with rather simple processing in the optical domain, more complex techniques may lead to even greater improvements, opening up a new exciting research direction.

## VII. CONCLUSION

In this work, we reviewed the concept of reservoir computing and discussed its application in the context of signal equalization for short-reach IM/DD transmission systems. RC-based equalization directly in the digital domain has been compared with alternative digital techniques, showing similar performance at potentially lower complexity and faster training. An overview of the impact of the main macroscopic properties of a reservoir on the system performance has been provided, focusing on memory and reservoir size. Finally, a new hybrid method combining optical pre-processing, by means of spectral slicing, and digital RC has been proposed. This new approach significantly outperforms digital-only techniques as well as previously proposed and more complex hybrid methods. A reach extension of 400% can be achieved for the proposed hybrid receiver compared to conventional digital-only techniques. These results clearly emphasize the strong potential of RC for equalization in short-reach communication systems and provide a better understanding of the link between RC network properties and system performance, especially in the context of signal pre-condition at the reservoir input.

## REFERENCES

- [1] J. K. Perin, A. Shastri, and J. M. Kahn, “Data center links beyond 100 Gbit/s per wavelength”, *Optical Fiber Technology*, **44**, pp. 69–85, 2018.
- [2] L. Grüner-Nielsen, S. N. Knudsen, B. Edvold, T. Veng, D. Magnussen, C. C. Larsen, and H. Damsgaard, Hans, “Dispersion compensating fibers”, *Optical Fiber Technology*, **6**(2), pp. 164–180, 2000.
- [3] N. M. Litchinitser, B. J. Eggleton, and D. B. Patterson, “Fiber Bragg gratings for dispersion compensation in transmission: theoretical model and design criteria for nearly ideal pulse recompression”, *Journal of Lightwave Technology*, **15**(8), pp. 1303–1313, 1997.
- [4] J. Jignesh, T. Eriksson, M. Chagnon, B. Corcoran, A. J. Lowery, F. Buchali, and H. Bülow, “Transmitter-side Volterra filtering for increased dispersion tolerance in 56 Gbaud PAM-4 systems”, in *Proc. Optical Fiber Communication Conference (OFC)*, San Diego, CA, USA, March 2018, paper M2C.6.
- [5] M. Rubsamen, J. M. Gene, P. J. Winzer, and R.-J. Essiambre, “ISI mitigation capability of MLSE direct-detection receivers”, *IEEE Photonics Technology Letters*, **20**(8) pp. 656–658, 2008.



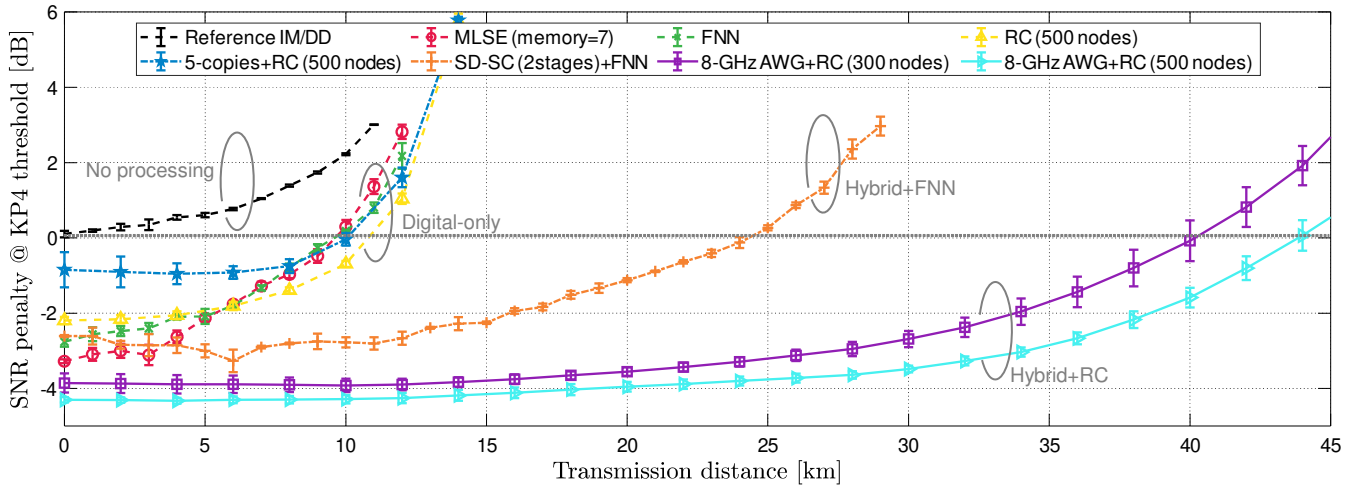


Fig. 13. Overall performance comparison between techniques: digital-only equalization schemes and hybrid schemes relying on FNNs and RC.

[6] O. Agazzi, M. R. Hueda, H. S. Carrer, and E. D. Crivelli, “Maximum-likelihood sequence estimation in dispersive optical channels”, *Journal of Lightwave Technology*, **23**(2), pp. 749–763, 2005.

[7] Q. Hu, M. Chagnon, K. Schuh, F. Buchali, and H. Bülow, “High Data Rate Tomlinson-Harashima precoding-Based Pam Transmission”, in *Proc. European Conference on Optical Communication (ECOC)*, Dublin, Ireland, September 2019.

[8] A. Mecozzi, C. Antonelli, M. Shtائف, “Kramers–Kronig coherent receiver”, *Optica*, **3**(11), pp. 1220–1227, 2006.

[9] S. Gaiarin, X. Pang, O. Ozolins, R. T. Jones, E. P. da Silva, R. Schatz, U. Westergren, S. Popov, G. Jacobsen, and D. Zibar “High speed PAM-8 optical interconnects with digital equalization based on neural network”. in *Proc. Asia Communications and Photonics Conference (ACP)*, Wuhan, China, November 2016, Paper AS1C-1.

[10] M. Chagnon, B. Karanov, and L. Schmalen, “Experimental demonstration of a dispersion tolerant end-to-end deep learning-based im-dd transmission system”, in *Proc. European Conference on Optical Communication (ECOC)*, Roma, Italy, September 2018. Paper Tu4F.6.

[11] B. Karanov, D. Lavery, P. Bayvel, and L. Schmalen, “End-to-End Optimized Transmission over Dispersive Intensity-Modulated Channels Using Bidirectional Recurrent Neural Networks”, *Optics Express*, **27**(14), pp. 19650–19663 (2019).

[12] J. Estaran, R. Rios-Müller, R. M. A. Mestre, F. Jorge, H. Mardoyan, A. Konczykowska, J.-Y. Dupuy, and S. Bigo, “Artificial neural networks for linear and non-linear impairment mitigation in high-baudrate IM/DD systems”, in *Proc. European Conference on Optical Communication (ECOC)*, Düsseldorf, Germany, September 2016.

[13] P. Bienstman, J. Dambre, A. Katumba, M. Freiberger, F. Laporte, A. Lugnan, S. Sackesyn, and C. Ma, “Neuromorphic information processing using silicon photonics”, in *Proc. SPIE 11081*, Active Photonic Platforms XI, 11081, September 2019.

[14] A. Katumba, X. Yin; J. Dambre, and P. Bienstman, “A Neuromorphic Silicon Photonics Nonlinear Equalizer For Optical Communications With Intensity Modulation and Direct Detection”, *Journal of Lightwave Technology*, **37**(10), pp. 2232–2239, May 2019.

[15] S. Li, S. Ohlendorf, and S. Pachnicke, “100 km 56 GBd PAM-4 Transmission using Photonic Reservoir Computing”, in *Proc. European Conference on Optical Communication (ECOC)*, September 2019, Paper Tu.2.B.

[16] M. Sorokina, S. Sergeev, and S. Turitsyn, “Fiber echo state network analogue for high-bandwidth dual-quadrature signal processing”, *Optics Express*, **27**(3) pp. 2387–2395, 2019.

[17] A. Argyris, J. Bueno, and I. Fischer, “PAM-4 Transmission at 1550 nm Using Photonic Reservoir Computing Post-Processing”, *IEEE Access*, **7** pp. 37017–37025, 2019.

[18] A. Argyris, J. Bueno, and I. Fischer, “Photonic machine learning implementation for signal recovery in optical communications”, *Scientific Reports*, **8**(1) pp. 8487, 2018.

[19] D. Brunner, M.C. Soriano, G. Van der Sande, “Photonic Reservoir Computing”, *Walter de Gruyter GmbH & Co KG*, 1st ed. , 2019

[20] S. Ranzini, F. Da Ros, H. Bülow, and D. Zibar, “Tunable Optoelectronic Chromatic Dispersion Compensation Based on Machine Learning for Short-Reach Transmission”, *Applied Sciences*, **9**(20) pp. 4332, 2019.

[21] S. Haykin, “Neural Networks: A Comprehensive Foundation”, 3rd ed. Prentice-Hall, 2007.

[22] D. Kenji, “Bifurcations in the learning of recurrent neural networks”. in *Proc. 1992 IEEE International Symposium on Circuits and Systems*, pp. 2777–2780, 1992.

[23] Y. Bengio, P. Simard, and P. Frasconi, “Learning long-term dependencies with gradient descent is difficult”, *IEEE transactions on neural networks*, **5**(2), pp. 157–166, 1994.

[24] I. Sutskever, “Training recurrent neural networks”, *Doctoral Thesis*, University of Toronto Toronto, Ontario, Canada, 2013.

[25] R. Pascanu, T. Mikolov, and Y. Bengio, “On the difficulty of training recurrent neural networks”, in *Proc. International conference on machine learning*, pp. 1310–1318, 2013.

[26] H. Jaeger, “The ”echo state” approach to analysing and training recurrent neural networks”. *Technical Report GMD Report 148*, German National Research Center for Information Technology, 2001.

[27] W. Maass, T. Natschläger, and H. Markram, “Real-time computing without stable states: A new framework for neural computation based on perturbations”. *Neural Computation*, **14**(11), pp. 2531–2560, 2002

[28] B. Schrauwen, D. Verstraeten, and J. Van Campenhout, “An overview of reservoir computing: theory, applications and implementations”. in *Proc. 15th European Symposium on Artificial Neural Networks*, pp. 471–482, 2007.

[29] M. Lukoševičius, “A practical guide to applying echo state networks”, *Neural networks: Tricks of the trade*, pp. 659–686, 2012.

[30] M. Datar, N. Immerlica, P. Indyk, and V. S. Mirrokni, “Locality-sensitive hashing scheme based on p-stable distributions”, in *Proc. of the twentieth annual symposium on Computational geometry*, pp. 253–262, 2004.

[31] W. B. Johnson, and J. Lindenstrauss, “Extensions of Lipschitz mappings into a Hilbert space”, *Contemporary mathematics*, **26**, pp. 189–206, 1984.

[32] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, “Extreme learning machine: a new learning scheme of feedforward neural networks”, *Neural networks*, **2**, pp. 985–990, 1984.

[33] M. Bauduin, S. Massar, and F. Horlin, “Non-linear satellite channel equalization based on a low complexity echo state network”, in *Proc. 2016 Annual Conference on Information Science and Systems (CISS)*, pp. 99–104, 2016.

[34] M. Lukoševičius, and H. Jaeger, “Reservoir computing approaches to recurrent neural network training”, *Computer Science Review*, **3**(3), pp. 127–149, 2009.

[35] L. Gonon, and J.-P. Ortega, “Reservoir computing universality with stochastic inputs”, *IEEE transactions on neural networks and learning systems*, Early-access 2019.

[36] L. Grigoryeva, and J.-P. Ortega, “Echo state networks are universal”, *Neural Networks*, **108**, pp. 495–508, 2018.

[37] M. Inubushi, and K. Yoshimura, “Reservoir computing beyond memory-nonlinearity trade-off”, *Scientific Reports*, **7**(1), pp. 10199, 2017.

- [38] H. Jaeger, and H. Haas, “Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication”, *Science*, **304**(5667), pp. 78–80, 2004.
- [39] K. Vandoorne, W. Dierckx, B. Schrauwen, D. Verstraeten, R. Baets, P. Bienstman, and J. Van Campenhout, “Toward optical signal processing using Photonic Reservoir Computing”, *Optics Express*, **16**(15), pp. 11182–11192, 2008.
- [40] D. Brunner, M.C. Soriano, C.R. Mirasso, and I. Fischer, “Parallel photonic information processing at gigabyte per second data rates using transient states”, *Nature Communications* **4**, pp. 1364, 2013.
- [41] B. Shi, N. Calabretta, and R. Stabile “Deep Neural Network Through an InP SOA-Based Photonic Integrated Cross-Connect,” *IEEE Journal of Selected Topics in Quantum Electronics*, **26**(1), pp. 7701111, 2020.
- [42] L. Appeltant, M.C. Soriano, G. Van der Sande, J. Danckaert, S. Massar, J. Dambre, B. Schrauwen, C.R. Mirasso, and I. Fischer, “Information processing using a single dynamical node as complex system,” *Nature Communications*, **2**, pp. 468, 2011.
- [43] A. Röhm, and K. Lüdge, “Journal of Physics Communications,” *Journal of Physics Communications*, **2**, pp. 085007, 2018.
- [44] K. Vandoorne, P. Mechet, T. Van Vaerenbergh, M. Fiers, G. Morthier, D. Verstraeten, B. Schrauwen, J. Dambre, and P. Bienstman, “Experimental demonstration of reservoir computing on a silicon photonics chip”, *Nature Communications*, **5**, pp. 3541, 2014.
- [45] G. Mourgias-Alexandris, A. Ysakyridis, N. Passalis, A. Tefas, K. Vysokinos, and N. Pleros, “An all-optical neuron with sigmoid activation function,” *Optics Express*, **27**(7), pp. 9620–9630, 2019.
- [46] C. Mesaritakis, A. Kapsalis, A. Bogris, and D. Syvridis, “Artificial Neuron Based on Integrated Semiconductor Quantum Dot Mode-Locked Lasers,” *Scientific Reports*, **6**, pp. 39317, 2016.
- [47] G. Tanaka, T. Yamane, J. B. Héroux, R. Nakane, N. Kanazawa, S. Takeda, H. Numata, D. Nakano, and A. Hirose, “Recent advances in physical reservoir computing: A review”, *Neural Networks*, 2019.
- [48] L. De Marinis, C. Marco, C. Piero, and A. Nicola, “Photonic Neural Networks: a Survey”, *IEEE Access*, accepted for publication, 2019.
- [49] E. Agrell, and M. Secondini, M. “Information-theoretic tools for optical communications engineers”, *In Proc. of the IEEE Photonics Conference*, pp. 99–103, 2018.
- [50] J. C. Doyle, B. A. Francis, and A. R. Tannenbaum, “Feedback control theory”, *Courier Corporation*, 2013.
- [51] J. Dambre, D. Verstraeten, B. Schrauwen, and S. Massar, “Information processing capacity of dynamical systems”, *Scientific reports*, **2**(1), pp.1–7, 2012.

**Francesco Da Ros** received his B.Sc. degree in information engineering from Università degli Studi di Padova, Italy, in 2008. He also received a dual M.Sc. degree in telecommunication engineering from Università degli Studi di Padova, Italy, and Technical University of Denmark, Denmark, in 2011 and the Ph.D. degree from the Technical University of Denmark in 2014. Between 2013 and 2014 he was a guest researcher at the Fraunhofer Institute for Telecommunications, Heinrich-Hertz-Institute, Germany. Between 2015 and 2018, he has been working as a post-doc researcher at the Department of Photonics Engineering at DTU within the Center for Silicon Photonics for Optical Communications (SPOC) focusing on all-optical signal processing and nonlinearity compensation techniques for optical transmission systems. Since 2019, he is a researcher in the Machine Learning in Photonic Systems group at DTU Fotonik, focusing on the inverse scattering transform and machine learning techniques applied to coherent communications. Dr. Da Ros has co-authored more than 100 journal and conference papers. He is an OSA Young Professional Volunteer member, an IEEE Young Professional member and has been serving as technical subcommittee member of the Conference on Lasers and Electro Optics (CLEO) since 2018.

**Stenio M. Ranzini** is pursuing a Ph.D. in optical communications at the Technical University of Denmark in the Machine Learning in Photonic Systems group. He received his bachelor’s degree at the University of Sao Paulo, Brazil, and his master’s degree at the State University of Campinas, Brazil. From 2011 to 2018, he was a researcher at CPqD Foundation, Campinas, Brazil. At CPqD, he developed and implemented state-of-art digital signal processing algorithms to be employed in a commercial physical layer ASIC for high speed (400G) optical communication transceivers. He also helped to found the Brazilian photonic society (SBFoton) in 2017, where he was the administrative director. Currently, he is in the fiscal council of the organization and a Marie-Curie Fellow.

**Henning Bülow** Bülow is with the Optical Networks Department at Bell Labs in Stuttgart, Germany. He received his Ph.D. in electrical engineering from the University of Berlin for work on integrated optical switching matrices. During his two decades tenure with Nokia Bell Labs he has worked on many aspects of optical transmission ranging from dynamic mitigation of optical transmission impairments by analog electronic, digital, or optical signal processing means, 4D modulation, and nonlinear Fourier transform based transmission. He was member of technical program committee of OFC and ECOC, respectively. From 2008 to 2011 he served as Guest Professor with the University of Erlangen-Nuremberg and is now leading the signal processing department of Nokia Bell Labs in Stuttgart, Germany.

**Darko Zibar** is Associate Professor at the Department of Photonics Engineering, Technical University of Denmark and the group leader of Machine Learning in Photonics Systems (M-LiPS) group. He received M.Sc. degree in telecommunication and the Ph.D. degree in optical communications from the Technical University of Denmark, in 2004 and 2007, respectively. He has been on several occasions (2006, 2008 and 2019) Visiting Researcher with the Optoelectronic Research Group led by Prof. John E. Bowers at the University of California, Santa Barbara, (UCSB). At UCSB, he has been working on topics ranging from analog and digital demodulation techniques for phase modulated analog optical links, as well as ultra-sensitive laser phase noise measurements techniques using machine learning. In 2009, he was a Visiting Researcher with Nokia-Siemens Networks, where he worked on clock recovery techniques for 112 Gb/s polarization multiplexed optical communication systems. In 2018, he was Visiting Professor with Optical Communication (OptCom) group Dipartimento di Elettronica e Telecomunicazioni, Politecnico di Torino working on the topic of machine learning based Raman amplifier design. His research efforts are currently focused on the application of machine learning techniques to classical and quantum optical communication and measurement systems. He is a recipient of Best Student paper award at Microwave Photonics Conference (2006), Villum Young Investigator Programme (2012), Young Researcher Award by University of Erlangen-Nurnberg (2016) and European Research Council (ERC) Consolidator Grant (2017). Finally, he was a part of the team that won the HORIZON 2020 prize for breaking the optical transmission barriers (2016).