Machine-learning-based equalization for short-reach transmission: neural networks and reservoir computing

F. Da Ros¹, S. M. Ranzini¹,², R. Dischler², A. Cem¹, V. Aref², H. Bülow², and D. Zibar¹

¹DTU Fotonik, Technical University of Denmark, DK-2800, Kgs. Lyngby, Denmark
²Nokia Bell Labs, Lorenzstr. 10, 70435, Stuttgart, Germany

ABSTRACT

The substantial increase in communication throughput driven by the ever-growing machine-to-machine communication within a data center and between data centers is straining the short-reach communication links. To satisfy such demand - while still complying with the strict requirements in terms of energy consumption and latency - several directions are being investigated with a strong focus on equalization techniques for intensity-modulation/direct-detection (IM/DD) transmission. In particular, the key challenge equalizers need to address is the inter-symbol interference introduced by the fiber dispersion when making use of the low-loss transmission window at 1550 nm. Standard digital equalizers such as feed-forward equalizers (FFEs) and decision-feedback equalizers (DFEs) can provide only limited compensation. Therefore more complex approaches either relying on maximum likelihood sequence estimation (MLSE) or using machine-learning tools, such as neural network (NN) based equalizers, are being investigated. Among the different NN architectures, the most promising approaches are based on NNs with memory such as time-delay feedforward NN (TD-FNN), recurrent NN (RNN), and reservoir computing (RC). In this work, we review our recent numerical results on comparing TD-FNN and RC equalizers, and benchmark their performance for 32-GBd on-off keying (OOK) transmission. A special focus will be dedicated to analyzing the memory properties of the reservoir and its impact on the full system performance. Experimental validation of the numerical findings is also provided together with reviewing our recent proposal for a new receiver architecture relying on hybrid optoelectronic processing. By spectrally slicing the received signal, independently detecting the slices and jointly processing them with an NN-based equalizer (wither TD-FNN or RC), significant extension reach is shown both numerically and experimentally.

Keywords: Short-reach optical communication, machine learning, neural networks, reservoir computing

1. INTRODUCTION

The steep increase in machine-to-machine communication traffic sets a strain on short-reach links up to a 100 km.¹ For such distances, latency and low-power consumption are key. Therefore, the main technology employed relies on low-complexity intensity-modulation/direct-detection (IM/DD) transmission.² This choice leads to a trade-off between complexity and transmission reach. By avoiding coherent detection, the impact of linear transmission impairments such as fiber group velocity dispersion (GVD) significantly limits the achievable transmission reach. The GVD induces pulse broadening and consequently intersymbol interference (ISI). In coherent systems, the GVD can easily be compensated for by recovering the phase of the received optical field. However, for an IM/DD receiver that only detects the received signal intensity, ISI cannot be compensated through simple electronic dispersion compensation (ECD).

All-optical dispersion compensation can be applied to remove the impact of ISI before opto-electronic conversion, but that usually leads to a significant decrease in the available received power which, in turn, has a negative impact on the transmission reach.

Alternatively, digital techniques such as linear or nonlinear equalizers can be applied for equalization of IM/DD systems. Linear feed-forward equalizers (FFEs) and decision-feedback equalizers (DFEs) are commonly employed and provide reasonable performance gain by approximating ECD.³ More complex methods such as

Further author information: (Send correspondence to F.D.R.)
F.D.R.: E-mail: fdro@fotonik.dtu.dk
maximum likelihood sequence estimation (MLSE)\textsuperscript{4} or nonlinear equalizers based on neural networks (NN),\textsuperscript{5–8} provide a further performance boost but at the expense of increased received complexity, which generally grows steeply with the system memory. In this work, we review our systematic comparison between linear and nonlinear equalizers focusing on a simple OOK transmission system.\textsuperscript{9–11} In particular, we analyze the performance improvement provided by two recent machine-learning tools which still allow for relatively limited training complexity, i.e. time-delay feed-forward NN (TD-FNN) and reservoir computing (RC).

In the following section (Section 2), the two NN schemes are briefly reviewed with a special focus on the memory properties of RC. The application of TD-FNN and RC for equalization and the performance improvement they provide for OOK transmission is shown in Section 3, first through a numerical analysis (Section 3.1) followed by the experimental validation (Section 3.2). The key results are summarized in Section 4

2. NEURAL NETWORKS

Neural networks are an effective tool which allows to approximate the black-box response of a specific system without necessarily requiring knowledge beyond a dataset of inputs and expected outputs to train the model.

Among the several NN architectures proposed, TD-FNNs and RNNs allow to predict time-dependent processes. This is particularly useful for communication systems with memory (e.g. ISI) as in IM/DD transmission system. While demonstrations of full RNNs for IM/DD system equalization have been reported,\textsuperscript{7} RNNs suffer from training challenges due to vanishing gradients which may prevent the use of standard training methods such as back-propagation. Simplified RNN models have therefore been considered, focusing on TD-FNNs,\textsuperscript{5,10,11} and RC.\textsuperscript{10,11,14–16} Alternative lower-complexity RNNs architectures such as gated recurrent units (GRUs),\textsuperscript{12} and long short-term memory (LSTM)\textsuperscript{13} have been applied mainly to coherent transmission so far.

In the following we briefly review the architecture and training of TD-FNNs and RC focusing on the memory they can provide.

2.1 Time-delay feed-forward neural networks

The architecture of a single-hidden layer TD-FNN is shown in Fig. 1(a) consisting of input (size \(M \times (K + 1)\)), hidden (size \(N_{HN}\)) and output (size \(N_{out}\)) layers. As FNNs do not have intrinsic memory, TD-FNNs are fed with multiple time-instants as \(K + 1\) parallel inputs \(\{u_i[t_n-K/2], \ldots, u_i[t_n+K/2]\}_{i=1}^M\). Such a window of time-instants shifts in time, ensuring that the TD-FNN has knowledge of both the past \((t_n+k, -K/2 \leq k < 0)\) and the future \((t_n+k, 0 < k \leq K/2)\) to provide the present prediction \((t_n)\). After the input layer, multiple hidden layers \((N_{HL} = 1\) in this work, see Fig. 1(a)) composed of \(N_{HN}\) hidden nodes with nonlinear activation functions approximate the desired system and are followed by the output layer providing the desired prediction.

TD-FNN can be trained with standard methods such as back-propagation as the time-dynamics of the network are not inherent of its architecture, but rather forced by considering a time-window at the input layer.

As the memory of a TD-FNN depends directly on the size of the input time-window, large memory inevitably requires a large network with a consequent increase of the number of weights to be trained. In the following analysis of TD-FNN-based equalizers, the input window for the TD-FNN is fixed to 5 symbols.

2.2 Reservoir computing

Reservoir computing (RC) has been introduced recently in\textsuperscript{17} to unify into a single framework the concepts of echo-state networks\textsuperscript{18} and liquid state machines.\textsuperscript{19} In short, the idea behind RC is to consider highly redundant RNNs in which the input and hidden parts of the network (reservoir) is left untrained. This allows to avoid the training challenges of standard RNNs, and to effectively train the full network by optimizing only the output weights with a single linear regression step.

The architecture of RC considered in this work is shown in Fig. 1(b), and consists of the same three blocks as for FNN. The hidden layer is renamed reservoir for its similarity with nonlinear dynamical systems. The inputs \(u[n]\) are injected into the reservoir through the interconnection \(W_{in}\) and drive the reservoir states \(x[n]\) characterized by recurrent interconnections (within the reservoir, \(W_{res}\)) and output interconnections (\(W_{out}\)).
Figure 1. Neural network architectures: (a) Time-Delay Neural Network and (b) Reservoir Computing

Here, we also consider direct interconnections between input and output layers $W_{\text{out}}^{\text{in}}$. The system outputs are collected at the output layer as $y[n]$. Both $W_{\text{out}}^{\text{in}}$ and $W_{\text{out}}^{\text{res}}$ weights are trained using ridge regression, whereas input and reservoir weights ($W_{\text{in}}$, $W_{\text{res}}$) are initialized from a random distribution and kept fixed. To decrease the reservoir complexity from the perspective of hardware implementation, only sparse connectivity with an average of 5 interconnections per node is considered in this work.

The characteristics of the reservoir, both in terms of number of nodes and of the spectral radius $\rho$ of the reservoir matrix $W_{\text{res}}$, contribute to defining its memory and thus its ability to equalize ISI.

2.2.1 Reservoir memory

In Fig. 2 we analyze the reservoir memory considering two alternative definitions: settling time derived from control theory, which is defined as the time it takes for the reservoir states to reach within 2% of their steady state values after a step-function applied at the input layer (single input node); 95%-memory capacity based on the concept of linear memory capacity, i.e. the ability to recall a random sequence seen at the input after a time delay $n$. For the latter definition, the memory is considered as the maximum delay for which the normalized mean square error (NMSE) of the recall task is still below 5%.

Reservoir memory considering (a) the reservoir settling time when driven with a step as input, and (b) 95%-memory capacity for the $n$-step linear recall, i.e. the longest instant for which the recall NMSE is still below 0.05. All results are averaged over 100 reservoir realization and follow the reservoir properties of.

For the analysis, we consider the average performance of 1000 randomly initialized reservoirs with an average of 5 interconnections per node, and a single input (plus bias) and output node. The input weights are drawn from a uniform distribution between 0 and 1, and the reservoir weights are drawn from a normal distribution with zero mean and variance $\sigma_{\text{res}}^2 = 1$. The spectral radius is then varied by re-scaling the reservoir matrix, see.
For fair comparison between the two definitions, no directed input-to-output interconnections are considered for the 95%-memory capacity, as is the case for the settling time.

In Fig. 2(a) the settling time is shown to increase with the reservoir size, approaching a saturation for reservoir sizes beyond 300 nodes. The settling time increases super-linearly with the spectral radius, even for values of $\rho > 1$ which are known to lead to unstable reservoirs and generally poor performance.\textsuperscript{20}

By considering the 95%-memory capacity, instead, the results are shown in Fig. 2(b). A similar increase and saturation of the memory with the reservoir size is observed, but the memory clearly drops as the reservoir becomes strongly unstable, i.e. $\rho > 1$.

These memory values will be qualitatively linked to the required memory to equalize ISI for IM/DD communications in the following section.

### 3. IM/DD TRANSMISSION SYSTEM

![Figure 3. Transmission system under investigation.](image)

In this section, we will describe the IM/DD transmission system used for the numerical and experimental setup to compare different algorithms for equalization - including the ones detailed in the previous sections.

The transmission system considered in this investigation is shown in Fig. 3. At the transmitter, a random stream of bits is generated and pulse shaped with root-raised cosine (RRC, 0.1 roll-off) to construct a 32-GBd OOK signal. The electrical signal is encoded onto the optical field by a Mach-Zehnder modulation (MMZ). The optical signal is then injected into the transmission channel consisting of standard single-mode fiber (SSMF). An additive white Gaussian noise (AWGN) source, e.g. a receiver pre-amplifier, is used to degrade the signal. 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/or transmission distance. At the receiver, two optical receivers are considered: a standard single photodetector (PD) receiver and a spectrally-sliced receiver. The first is the conventional single-PD receiver used as a reference. The second is the spectrally-sliced receiver proposed and discussed in.\textsuperscript{9–11} We have proposed the spectrally-sliced receiver to reduce the effect of power-fading in the IM/DD systems. This effect is avoided by filtering the received signal in smaller sub-bands with optical filters, before detection. The slicing process should cover the whole bandwidth of the received signal to not lose information. After either receiver front-end, different digital equalization schemes are then benchmarked both numerically and experimentally. Note that, for the spectrally-sliced receiver, the equalizer should not only mitigate the residual chromatic dispersion but recombine the frequency slices to recreate the full signal and recover the transmitted information.

#### 3.1 Numerical comparison of equalizers

In the numerical analysis, in order to keep the focus only on the chromatic dispersion and the square-law detection from the PD, the MZM, DAC, and ADC were not modeled and the SSMF transmission is modeled only as a linear lossless dispersive element with a dispersion parameter $D = 16.4$ ps/nm/km. The system is simulated at 8 samples/symbol. More details can be found in.\textsuperscript{10} The performance metric considered in this analysis is the SNR penalty. It is defined as the difference in required SNR for a BER at the KP4 threshold\textsuperscript{24} with reference to a 0-km transmission system employing a single PD receiver and no equalization. The results are shown in Fig. 4 as a function of the transmission length.
As can be seen, for the single-PD receiver, the use of TD-FNN (star green curve) and RC with a 500-node reservoir (square dark-red curve) allows an increase in transmission reach of $\approx 10$ km. A comparable distance with what is achievable with an MLSE receiver with memory 7 (circle yellow curve). It is worth highlighting that the effectiveness of RC is explored by using only one sample per time in the input of the reservoir, i.e. forcing the equalization to rely on the memory present in the recurrent connections of the reservoir. The single input RC equalizer performs similarly to the 5-symbol (40-samples at 8 sps) input TD-FNN, i.e. allows a decrease in input (and training) complexity for similar performance.

By moving to the spectrally-sliced receiver, a significant improvement in transmission distance is achieved (here shown only for an RC equalizer). This results in requiring 4 samples per time in parallel inputs to the RC equalizers (instead of one sample per time as in the single-PD receiver), however decreasing the required memory by a similar factor as the GVD-induced broadening scales linearly with the spectral bandwidth. The GVD of CD can be calculated as: $\Delta t = D \cdot L \cdot \Delta \lambda$, where $\Delta t$ is the spreading over time of the CD, $D$ is the CD coefficient, $\Delta \lambda$ is the wavelength distance for the received signal. Filtering the received signal with four different slices, reduces the $\Delta \lambda$, and thus the spreading over time, by a factor 4. The reduced time broadening reduces the total required amount of memory for the CD equalization. However, the four detected slices still needs to be regrouped and mitigate the residual CD. The RC is able to that with a fast training stage yielding an achievable transmission approaching 45 km.

![Figure 4](image_url)

**Figure 4.** Numerical results for the single-PD and spectrally-sliced receiver.

### 3.2 Experimental validation

For the experimental validation, a MZM driven by a 88-GSa/s digital-to-analog converter encodes the OOK signal onto the optical carrier. After SSFM transmission, the optical signal is then amplified by an erbium-doped fiber amplifier (EDFA), irrespective of the receiver used. For the spectrally-sliced PD, a wavelength selective switch (WSS) spectrally slices the signal with second-order Gaussian filters with a 3-dB bandwidth of 16 GHz and a 16-GHz shift in their central frequency, resulting in an overlap of about 8 GHz between spectral slices. The signal is then detected independently by four PDs and different equalization schemes are considered. For the single-PD reference receiver, the WSS is set as a bandpass filter to suppress out-of-band noise. More details can be found in.

The performance metric considered in the experimental results is the SNR defined by the system. Different from the numerical analyses, the SNR is not varied deliberately. It is defined by the gain of the receiver pre-amplifier, which compensates the varying losses of the inserted fiber spools and is set to keep a constant input power into the receiver. For the results we achieve BER equal to zero (no errors), we replaced to a floor value of $10^{-5}$, i.e. the lower bound on the BER we can reliably estimate with our measurements.
Fig. 5 shows the experimental results for different fiber length. They were varied considering different fiber spools, therefore the receiver SNR does vary only approximately linearly with the fiber length.\textsuperscript{11}

For the single-PD receiver, the results from all tested equalizers (FFE, TD-FNN, and RC) show similar performance, with only a marginal benefit of the TD-FNN.

For the spectrally-sliced receiver, as for the numerical prediction, a substantial increase in transmission distance is shown. The lower noise loading allows achieving BER below the KP4 threshold for up to 80-km transmission by using either the RC or the TD-FNN equalizer. The TD-FNN was used to show that similar performance can also be achieved with the different algorithms in the spectrally-sliced receiver. This is to highlight the reduction of the power fading effect in this setup. However, notice that for a spectrally-sliced receiver with TD-FNN equalizer, the processing of four parallel slices increases the complexity of the TD-FNN even further, resulting in up to 160 input nodes (plus 4 biases). We can also note that the experimental setup increased further the distance than the numerical setup for the spectrally-sliced receiver. The reason is twofold, from the one side, the noise added by the receiver per-amplifier is lower than the amount considered for the numerical analysis, from the other side, the MZM was not modeled in the numerical analysis. For OOK systems, the MZM helps the equalization as it makes the tails of the RRC filter non-negative, because of its finite extinction ratio.\textsuperscript{11}

4. CONCLUSIONS

Equalization for IM/DD transmission systems is discussed focusing on the higher gains provided by more advanced/complex equalization techniques based on machine learning. In particular, we discussed the use of TD-FNN and RC as nonlinear equalizers. The numerical results show similar performance by using both equalizers yielding a potential advantage of RC due to its simpler training procedure and inherent memory (smaller input layer). The experimental validation confirms our numerical findings, highlighting the potential of RC for equalization of IM/DD transmission systems, especially when combined with novel receiver architectures such as the recently proposed spectrally-sliced receiver.

Acknowledgments

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 766115, the European Research Council through the ERCCoG FRECOM project (grant No 771878) and Villum Fonden through the Villum Young Investigator program OPTIC-AI project (grant No 29344).
REFERENCES


