Deep Learning Based Digital Back Propagation with Polarization State Rotation & Phase Noise Invariance

Bertold Ian Bitachon[1*], Amirhossein Ghazisaeidi[2], Benedikt Baeuerle[1,3], Marco Eppenberger[1], Juerg Leuthold[1]

[1]*ETH Zurich, Institute of Electromagnetic Fields (IEF), 8092 Zürich, Switzerland*
[2]*Nokia Bell Labs, Paris-Saclay, route de villejust, Nozay, 91620, France*
[3]*Polariton Technologies AG, 8038 Zürich, Switzerland*
**bertold.bitachon@ief.ee.ethz.ch*

Abstract: A new deep learning training method for digital back propagation (DBP) is introduced. It is invariant to polarization state rotation and phase noise. Applying the method one gains more than 1 dB over standard DBP.

OCIS codes: (060.0060) Fiber optics and optical communications; (060.2330) Fiber optics communications

1. Introduction

Nonlinear impairment mitigation techniques have become an essential part in high-capacity communications links. Although efficient algorithms have been developed over the past years, many suffer from energy-hungry processing. The ideal nonlinear mitigation technique should be both efficient and require the least possible calculation effort.

Recently, neural networks have achieved record-breaking performance in various machine learning tasks, mainly due to their ability to approximate arbitrary functions [1]. This advancement in computational power has made neural networks an interesting candidate for non-linearity compensation (NLC). Several studies have shown that neural networks can mitigate fiber non-linearity impairments reasonably well at a much lower complexity than the well-established digital backpropagation (DBP) algorithm [2, 3]. In [4], neural networks were used to approximate nonlinear perturbations, leading to a record-breaking spectral efficiency (SE)-distance product of 66102 b/s/Hz-km within an experiment. Most of these algorithms work on the very end of the receiver digital signal processing (DSP) chain. However, it was shown [2] that a neural network could be used as a direct replacement of DBP and be placed at the very beginning of the receiver DSP chain. Doing so demands a modification to the symbol level training method, since the symbols are not available before performing conventional DSP at the receiver end.

In this paper, we propose a polarization state rotation and phase noise-invariant training method, which enables neural networks to compensate self-phase modulation (SPM) while yielding more than 1 dB gain with respect to one step per span DBP and reducing computational cost.

2. Theory

2.1. Complex-valued parametric optimization via neural network

Neural networks are very well suited for a direct DBP replacement, since neural networks have a similar structure to DBP. In neural networks, we alternate between linear and non-linear operations. The linear operation comes in either as a matrix multiplication or a convolution while the non-linear operation comes in the form of an activation function. The same also can be said of DBP. In DBP, we alternate between undoing the dispersion, which is a linear operation, and fiber non-linearity. Therefore, by choosing $f(x) = x \cdot \exp\left(j\zeta(|x|^2 + |y|^2)\right)$ as the activation function we can represent the DBP in the neural network framework [2]. x, y are the signal in the x- and y-polarizations respectively and ζ is the rotational strength. Due to this similarity, we can jointly optimize the parameters, i.e., the weights of the hidden layers $\boldsymbol{w}$ and ζ's using neural network's backpropagation algorithm. Since the weight is complex valued, we need to use the complex valued version of the backpropagation, such as described in [5], which leads to $\zeta \in \mathbb{C}$. This implies that the activation function also attenuates. The chosen activation function is analytic with respect to ζ, but nonanalytic in both x and y. To accommodate these conditions, we incorporate Wirtinger derivatives into our training algorithm [6].

2.2. Phase and Polarization State Rotation Insensitive Parametric Optimization

We are using a neural network (NN) as a direct replacement for DBP. Consequently, the position of our NN should as well be in the very beginning of the receiver DSP chain. Since the network is trained on the symbol level, several static layers must be added during the training loop to ensure that the network is still trainable. These static layers demultiplex the dual polarization signal and compensate phase rotation. The latter arises mainly due to the laser linewidth and the frequency offset between the transmitter and receiver laser. The static layers are not trained, but inferred from the conventional receiver DSP (or a "linear DSP scheme" performing chromatic dispersion compensation only, polarization rotation compensation and phase estimation), and not updated during the training loop. This way, the network only trains for NLC. Furthermore, this ensures that the trained NN is blind to polarization state rotation and phase noise. Since polarization demultiplexing is just a linear operation using a 2-by-2 FIR filter, we incorporate it into the network by using an additional hidden layer with a linear activation function. We integrate the static phase rotation layer by adding another hidden layer with a linear activation function. The weight in this layer is fixed to a constant complex number determined by the estimated phase.

We separate the training phase into two stages, shown in Fig. 1(a). In the 1st stage, we estimate the parameter of the static layer, i.e. the weight of the 2-by-2 FIR polarization demultiplexing filter and phase correction, using conventional receiver DSP on the training set. In the 2nd stage, we train the weights and rotational strengths of our networks using the complex-valued backpropagation and the previously estimated static layers. When the trained network infers on different, unseen datasets, we remove the static layers (in green), and compensate the polarization and phase noise using conventional receiver DSP.

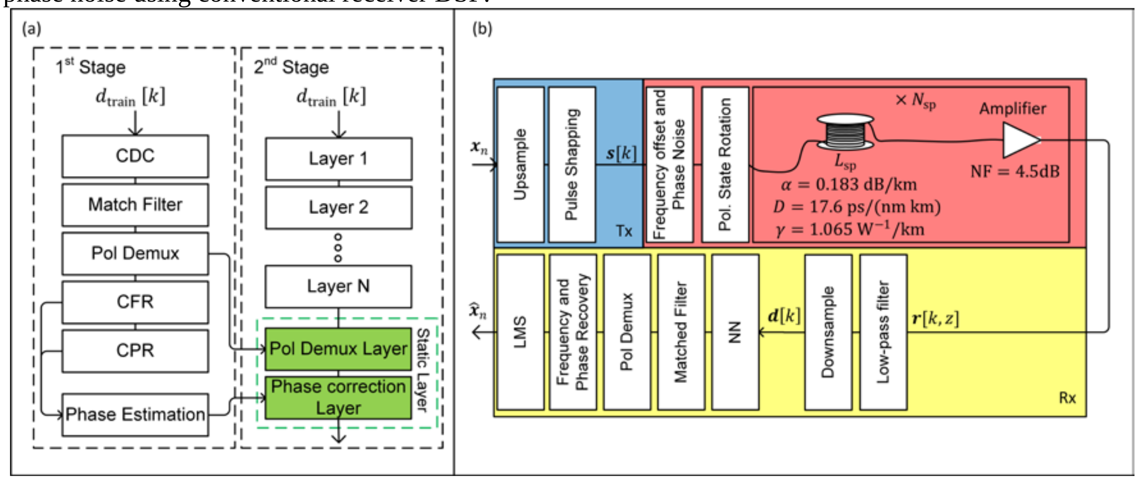

Fig. 1 (a) Training phase of our network. In the first stage, bedsides doing the whole DSP processing stages we estimate the polarization rotation and phase offset, and then we use those estimations as input to a static layer in the second stage. The training loop then takes the estimations into account during the second stage. (b) The system setup used for simulation. The neutral network (NN) investigated here, is at the begin of the receiver processing chain.

3. System and Networks Setup

To verify our training method, we first performed a numerical simulation. The transmission setup for the numerical experiment can be seen in Fig. 1(b). A dual polarized 32GBd 16QAM signal is encoded first. Prior to transmission, the symbols are up-sampled and passed through a root-raised cosine pulse-shaping filter with a roll-off factor of 0.05. Before launching the output of the shaping filter $\boldsymbol{s}[k]$ through the fiber, we add frequency offset, phase noise and polarization state rotation. The up-sampled pulse $\boldsymbol{s}[k]$ propagates through the fiber according to the manakov form of the nonlinear Schrödinger equation (NLSE) in the $\frac{\partial \boldsymbol{A}(t,z)}{\partial z} = \left(-\frac{\alpha}{2} - j\left(\frac{\beta_2}{2}\right)\frac{\partial^2}{\partial t}\right)\boldsymbol{A}(t,z) + \frac{j8}{9}\boldsymbol{\gamma}|\boldsymbol{A}(t,z)|^2\boldsymbol{A}(t,z)$, where $\boldsymbol{\alpha}, \boldsymbol{\beta_2}, \boldsymbol{\gamma}$ are attenuation, dispersion and nonlinearity parameter 2-by-2 matrices, respectively. In this paper, the NLSE is solved using the split step Fourier method (SSFM)[7]. Furthermore, we did not take into account any polarization dependent impairment, so that $\boldsymbol{\alpha}, \boldsymbol{\beta_2}, \boldsymbol{\gamma}$ are diagonal matrices. The signal $\boldsymbol{d}[k]$ is then passed through the proposed NN-based NLC, a matched filter and a 2-by-2 demultiplexing filter. Finally, the phase and frequency offset are recovered to get the approximate transmitted symbol $\boldsymbol{\hat{x}_n}$. To simulate the link, we used an oversampling of 6 and a 1 km step in the forward propagation direction of the SSFM while an oversampling of 2 is used in the receiver. The length of the span L_{sp} has been set to 100 km and the number of span repetitions N_{sp} is 12. Per span, we used a fiber that has dispersion parameters of 17.6ps/(km nm), 0.183dB/km of attenuation and 1.065W^{-1}km^{-1} of nonlinearity and an ideal amplifier that only compensates the fiber loss which has a noise figure of 4.5dB. Lastly, we added a random polarization state rotation, a random frequency offset and choose 100kHz as the laser linewidth. After passing through the channel, the signal goes through the NLC, the polarization demultiplexer, the frequency offset estimator, the phase noise estimator, and lastly a LMS filter. The LMS filter is used to correct any leftover polarization state rotation.

We fixed the number of hidden layers to 12 and varied the number of filter taps in the layers. This is equivalent to 1 step per span DBP. We initialized the weights in the hidden layers using the impulse response of the dispersion filter [8] and the rotational strength as zero. We used a 21 taps 2-by-2 FIR filter for the polarization demultiplexing. The filter taps were trained using the constant modulus (CMA) and the radius directed equalizers (RDE) algorithms. For the phase correction layer, we used the 4th power algorithm to estimate the frequency offset and the blind phase search (BPS) to estimate the laser phase noise. During inference, we applied the same algorithms, (RDE, CMA, 4th power, and BPS) to estimate the polarization state rotation, frequency offset and phase noise on the unseen dataset. Finally, we utilized a complex number compatible ADAM optimizer [9], online training and 1 million symbols to train the networks.

4. Result and Discussion

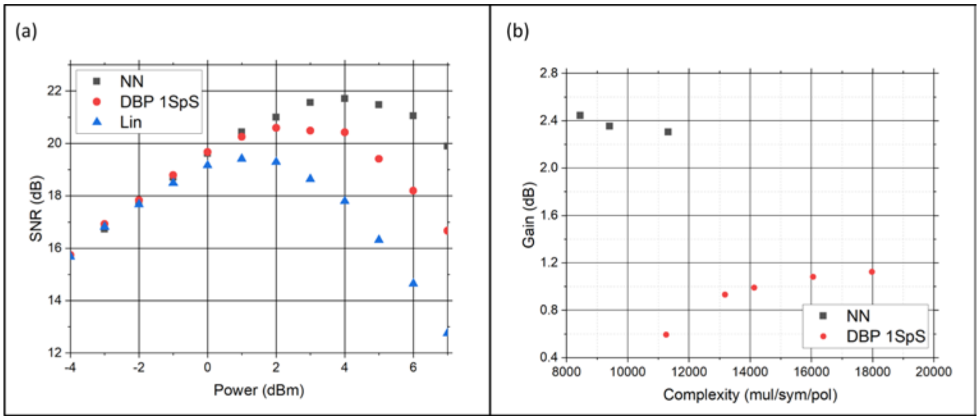

Fig. 2 (a) Average of digitally estimated SNRs as a function of launch power for three different techniques, namely the neural network (NN), the digital backpropagation (DBP) with 1 step per span (SpS), and a linear DSP scheme (Lin). (b) Received SNR Gain using the neural network (NN) or DBP over a linear compensation scheme plotted as a function of complexity

Fig. 2 (a) shows the average SNR of digitally estimated SNR in x- and y-polarization at the end of receiver DSP chain as a function of launch power in each fiber span. We repeat the training for each launch power and compare the results with a linearly compensated received SNR and a 1 step per span (SpS) DBP compensated received SNR. We show that an improvement of >2dB compared over the linear compensation method and a >1 dB advantage compared over a 1 SpS DBP is possible with the proposed training method. We quantify the improvement in inference complexity by plotting the number of multiplications (mul) needed to process one symbol per polarization. Based on [10], the number of multiplication per symbol per polarization for DBP is $(4 \cdot N_{\text{taps}} + 9) \cdot \frac{f_s}{R} \cdot N_{\text{sp}}$, whereas for our network is $(4 \cdot N_{\text{taps}} + 12) \cdot \frac{f_s}{R} \cdot N_{\text{sp}}$, where N_{taps} is the dispersion filter length per span, N_{sp} is the number of span and $\frac{f_s}{R}$ is the oversampling ratio. Fig. 2(b) shows the gain with respect to the linear compensation as a function of complexity. To obtain a gain of more than 2 dB, our network can use a complexity below 10000 mul/sym/pol. At low complexity, DBP is unable to achieve the same gain, as the chosen chromatic dispersion filter cannot optimally compensate the dispersion and non-linearity simultaneously. Meanwhile, our proposed network can jointly optimize all weights and every ζ in different layers, especially when the complexity is limited. By varying the N_{taps}, we indicated that a variation in complexity does not affect the gain significantly.

5. Conclusion

We develop a brand new training method, which allows neural networks to compensate self-phase modulation (SPM) in the presence of polarization state rotation and phase noise. The network yields more than 1 dB gain with respect to one step per span DBP at a reduced computational cost.

Acknowledgements

The authors acknowledge ERC PLASILOR (670478) for financial support, and Mikael Mazur for his work on nonlinear fiber simulation.

6. References

[1] I. Goodfellow *et al.*, *Deep Learning*. MIT Press, 2016.
[2] C. Häger *et al.*, 2018, p. W3A.4: Optical Society of America.
[3] O. Sidelnikov *et al.*, ECOC, Rome, 2018.
[4] V. Kamalov *et al.*, 2018, Th4D.5: Optical Society of America.
[5] T. Kim *et al.*, *Journal of VLSI signal processing systems for signal, image and video technology,* 32, pp. 29-43, 2002.
[6] T. Adali *et al.*, *IEEE Trans. on Signal Process.*, 59, pp. 5101-5125, 2011.
[7] E. Ip, *JLT,* 28, pp. 939-951, 2010.
[8] S. J. Savory, *Opt. Express,* 16, pp. 804-817, 2008.
[9] A. Sarroff, "Complex Neural Networks for Audio," Doctor of Philosophy Thesis, Dartmouth College, 2018.
[10] E. Ip *et al., JLT,* 26, pp. 3416-3425, 2008.