Evaluation of Frequency-Domain Learned Digital Back-Propagation Nonlinear Compensation for Unrepeatered Optical Links

Lucas Silva Schanner

***Abstract*— We have implemented a technique for nonlinear compensation in optical transmission based on neural network optimization applied to digital back-propagation and evaluated its performance with experimental data from an unrepeatered link, sweeping the parameters most relevant to computational complexity. This technique enabled mutual information gains over 0.1 bit/symbol in all tested scenarios when compared with the non-optimized counterpart, or 0.15 bit/symbol when compared with similar complexity linear compensation.**

***Keywords*— digital back-propagation, artificial intelligence, nonlinear compensation, unrepeatered systems.**

I. Introduction

Optical fiber networks constitute the backbone of modern internet infrastructure. As demand for data traffic continues to grow exponentially, so must the capacity of the underlying networks. As such, there is a constant search for innovative and more efficient transmission systems capable of delivering increasingly higher data rates. One of the key limiting factors of achievable data rates in optical fiber communications is the presence of fiber nonlinearities [1], most notably due to the Kerr effect. The Kerr effect is a nonlinear distortion that varies in magnitude with the optical power in the fiber, creating a correlation between the signals phase and the optical power density on the fiber core. As a result of this phenomenon, after a certain threshold, increasing the transmission power in an optical fiber will only degrade performance, instead of enhancing it, as would be expected in a linear system impacted only by additive noise.

There is, therefore, a great interest in developing methods to mitigate this distortion, as it would allow for higher transmission power, higher signal-to-noise ratios and consequently higher data rates and/or link reach. One of the most well established nonlinear compensation techniques is the digital back-propagation (DBP) algorithm [2]. It is based on the nonlinear differential equation that models the propagation of light through optical fiber, known as the non-linear Schröedinger equation (NLSE), essentially aiming to provide a numerical solution to the equation.

Noticing the structural similarity between DBP and artificial neural networks (ANN), recent works [3, 4] have proposed combining data-driven optimization techniques developed for ANN and the structure of DBP, creating an architecture that is a hybrid between data-driven and model-driven methods. Utilizing this approach, it was possible to significantly improve performance, reduce complexity, and remove requirements of accurate knowledge of link parameters, when compared with traditional DBP. This method was named Learned-DBP (LDBP).

Such method could be specially useful for unrepeatered systems, whose links employ no in-line active components for amplification of the optical signal. They are of great interest to be installed over remote or geographically inaccessible regions without the need for electricity sources along the link. They represent a unique challenge for nonlinear compensation techniques, because since they employ higher optical powers and complex hybrid amplification schemes to allow propagation over longer distances, they are more heavily impacted by nonlinear distortion.

In this work, we describe the LDBP algorithm and evaluate the performance of a frequency domain implementation in the particular case of unrepeatered links, sweeping the metaparameters most relevant for determining complexity. The method is validated using experimental data acquired from the unrepeatered transmission of 17 $\times$ 200-Gb/s wavelength-multiplexed channels, with its performance in terms of mutual information (MI) being compared to the reference set by traditional DBP.

II. Methodology

A. The Nonlinear Schröedinger Equation

Most well established methods for compensating the Kerr-related phenomena are based on the differential equation that describes the lightwave propagating through the fiber, known as the nonlinear Schröedinger equation (NLSE). In its simplified form, the NLSE can be written as:

$$\frac{\partial A(z,t)}{\partial z} = -\frac{\alpha}{2}A - \beta_1\frac{\partial A}{\partial t} + j\frac{\beta_2}{2}\frac{\partial^2 A}{\partial t^2} - j\gamma|A|^2 A \quad (1)$$

where $A(z,t)$ is the complex electric field at position z and instant t, α is the attenuation of the fiber, β_1 and β_2 correspond to group velocity and differential group velocity, respectively, and γ is the fiber nonlinear parameter.

It should be firstly noted that, if we take γ to be 0, the resulting equation is linear and admits a closed form solution in frequency domain:

$$\tilde{A}(z,\omega) = \tilde{A}(0,\omega)e^{-j(\beta_1\omega+\frac{\beta_2}{2}\omega^2)z}e^{-\frac{\alpha}{2}z} \quad (2)$$

Lucas Silva Schanner, Optical Communication Solutions, CPQD, Campinas-SP, e-mail: schanner@cpqd.com.br;

This work was partially supported by the Brazilian MCTI,and FUNTTEL-Finep.

where the term $e^{-j(\beta_1\omega)z}$ represents a time shift and is usually irrelevant for communication systems. The $e^{-j\frac{\beta_2}{2}\omega^2 z}$ term represents what is called chromatic dispersion (CD) and must be compensated in practically all high-capacity optical communication systems, either in the optical or digital domain.

A more complete version of Eq. 1 should describe $A(z,t)$ as a vector with two components, corresponding to two possible polarizations of light in the fiber. For the purposes of nonlinear compensation, the polarization components can be considered independent, save for the modulus operation, which takes in account both polarizations. Usually, after nonlinear compensation, a 2x2 MIMO equalization stage is employed to correct for crosstalk between polarization components.

B. Digital Back-Propagation

In Eq. 1, if we take $\gamma \neq 0$, there is no closed form solution, only numeric approximations of the result. One possible solution is through the split-step Fourier method (SSFM), which assumes that over short enough steps, the linear and nonlinear components of a differential equation can be considered independent, and therefore separated into a linear step, followed by a nonlinear step. In the case of the NLSE, the linear step is a CD compensating filter $H(w)$, and the nonlinear step is a point-wise phase shift proportional to instantaneous power:

$$A(z,t) = A`(z,t)e^{-j\gamma|A`(z,t)|^2 L_{\text{eff}}}$$

where the effective length L_{eff} is a function of the length of the step l and the fiber attenuation α. $A`(z,t)$ is an intermediate representation between the linear and nonlinear stage of a step.

The SSFM solution to a differential equation is the interleaved application of linear and nonlinear steps. Digital back-propagation (DBP) proposes to solve the NLSE backwards in the spatial direction using SSFM (Fig 1). It remains one of the most well established methods for nonlinearity compensation today, although its high computational cost has hindered its adoption in commercial systems.

The linear stages are described by $H(\omega)$, which is a function of the dispersion parameter D of the fiber at a given wavelength, as well as the length of the step. The nonlinear steps are function of the optical power at a given point $P(z)$, and also of the nonlinear parameter γ and of L_{eff}, which is a function of the length of the step and the fiber attenuation parameter α. Conventional DBP is sensible to uncertainty in these parameters, as small deviations from the expected values can incur in severe performance penalties. Furthermore, fixed step sizes are usually sub-optimal, as steps in regions of low optical power are less relevant for performance than steps in regions of high optical power [5].

Although time-domain CD compensation is possible, it has limited performance for short spans of fiber [6, 7] and is, therefore, impractical for conventional DBP.

C. Neural Networks and LDBP

Neural networks [8] have been demonstrated as highly competent in image recognition, speech transcription, natural

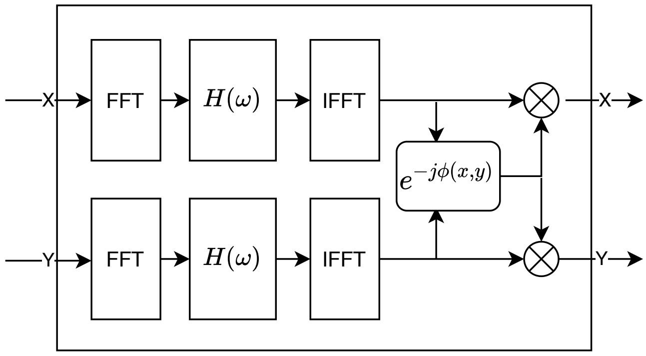

Fig. 1. Block diagram representing one step of DBP. Each step is composed by the application of a linear filter in the frequency domain, followed by nonlinear phase rotation. The intensity of the nonlinear phase rotation is a function of both the x and y components.

language processing and many other tasks. In their most general form, feed forward neural networks are mappings between input vector x and a output vector y, represented by:

$$y = \rho^{(N)}(B^{(N)}(...\rho^{(1)}(B^{(1)}(x))))$$

where N is the number of layers, $B^{(1)}, B^{(2)}, ..., B^{(N)}$ are linear or affine functions, and $\rho^{(1)}, \rho^{(2)}, ..., \rho^{(3)}$ are nonlinear functions. $B^{(k)}$ functions usually correspond to $B^{(k)}(x) = W^{(k)}x + b^{(k)}$, where the weights $W^{(k)}$ and the vectors $b^{(k)}$ are the weights and biases of the layer, respectively. The nonlinear functions $\rho^{(k)}$ are usually differentiable point-wise functions, like the logistic function, the sigmoidal function or rectified linear function [9]. The values of the weights $W^{(k)}$ and the biases $b^{(k)}$ are normally adjusted using an optimization algorithm to minimize some cost function, in what is referred to as "training" of the neural network.

As it is possible to notice, neural networks and DBP share the same fundamental structure, which is a sequence of linear steps interleaved with nonlinear steps. The coefficients in the linear filters $H(\omega)$, as well as the parameters related to the nonlinear steps $P(z)$, L_{eff}, γ, can be left as free weights and trained to minimize some cost function using optimization algorithms developed for neural networks, characterizing the previously discussed LDBP. [3].

Although [4] has demonstrated that joint optimization of the filters can make time-domain filtering effective for LDBP, with a desirable complexity reduction for hardware implementations, we have chosen to restrict the analysis of this work to frequency-domain implementation of LDBP, so its performance can be directly compared with the equivalent untrained conventional DBP, with no discussion on the complexity of the different architectures. This choice of implementation also allows us to use unsupervised learning, which greatly simplifies the training procedure and improves robustness of possible real-time implementations.

D. Unrepeatered Optical Systems

Unrepeatered systems are a subclass of optical transmission systems that employ no active components for amplification of the optical signal between along the link. They are particularly interesting for application of nonlinear compensation techniques because they typically employ higher optical powers to enable longer reach, which makes nonlinear distortion more relevant as an impairment. Moreover, the power is usually concentrated in a certain region at the beginning of the link, which makes the task of nonlinear compensation much easier.

Whereas for repeatered systems the number of steps in DBP is often discussed in terms of steps per span [2], making DBP tens of times more computationally expensive than purely linear compensation for links with a large number of spans, unrepeatered systems can have very substantial gains with only a few steps concentrated in the high propagation power regions, implying in computational complexity within the same order of magnitude of linear compensation.

However, while for repeatered systems the optimal step locations can be easily calculated [5], in the unrepeatered case it must be carefully optimized depending on the link characteristics. Usually this kind of link employs hybrid amplification schemes based on distributed Raman amplification (DRA) and/or remote optically pumped amplifiers (ROPA), resulting in irregular or uncertain power profiles. As the use of analytical approaches to optimize step locations are harder and less effective, data-driven optimization techniques are even more attractive as an alternative.

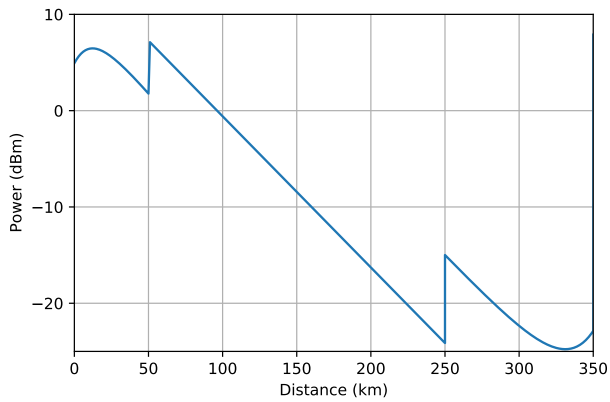

Fig. 2. Power profile of the link used in the experiment, estimated from nominal parameters of the fibers and amplifiers.

III. Experimental setup and Results

An unrepeatered link was employed to experimentally evaluate the performance of the nonlinear compensation technique, in which 17 channels (spaced by 50 GHz) were interleaved in two independent modulators resulting in 32-GBd 16-symbols quadrature amplitude modulated (16QAM) optical carriers. The 350-km link was composed by 100km of large effective area fibers (EX2000) followed by 250km of low-loss single mode fibers (LL-SMF), supported by hybrid amplification schemes compounded by erbium-doped fiber amplifiers (EDFA), distributed Raman amplifiers (DRA), and remote optically pumped amplifiers (ROPA), placed at 50km and 250km away from the transmitter. Finally, on the reception, the channel under test is filtered and coherently received being acquired by a real-time scope and processed offline.

The conventional DBP processing was carried out considering the nominal values of the fiber dispersion and nonlinear parameters, as well as the power profile estimated from nominal values of parameters from the amplifiers and fibers, as shown in Fig. 2. DBP was applied to the first 100 km of fiber, with the remaining 250 km being compensated linearly, with frequency domain equalization. We chose to apply DBP and LDBP only to the first 100 km of fiber because it is the region were most of nonlinear distortion occurs, due to higher optical powers, but also because it allows a fairer comparison between LDBP and constant step size DBP, since the power profile stays relatively flat.

The frequency domain LDBP tests were conducted using the Tensorflow [10] framework, which allows automatic differentiation. In all scenarios, the LDBP was initialized with the same weights as the conventional DBP to ensure initial performance sufficient for self training. Then, the captures from our unrepeatered system setup were divided into one training capture and four validation captures, of equal size. Random windows of 10000 samples were selected from our training capture, for which equalizer coefficients could be assumed constant. For each window, timing error was first estimated and corrected in pre-processing, to avoid differentiation of timing error correction algorithms, then 2x2 multiple-input multiple-output (MIMO) equalizer coefficients were estimated for the window and set as constant, to allow simple gradient propagation. The window is passed through the DSP chain, encompassing partial CD compensation (CDC), LDBP, 2x2 MIMO equalization, frequency offset recovery and phase recovery. Then, the error vector magnitude (EVM) is calculated from the recovered symbols and used as cost function. The gradients to the coefficients in the LDBP are computed and one step of the optimization algorithm is performed. Adam [11] was chosen for it's superior performance compared to SGD and relative simplicity. This whole cycle is repeated 500 times for each experiment. Finally, we estimate mutual information (MI) using the average EVM from the four validation captures.

Figure 3 shows achieved mutual information as a function of the number of steps and number of taps for LDBP and conventional DBP. In the cases where the impact of the number of steps were under test, a relatively large (1024) filter size was considered and in the cases where the impact of the size of the filters were under test, a sufficiently large number of steps (5) was considered. This was done so that the impact of these two parameters could be studied separately. Both graphs in Fig. 3 show mutual information achieved with only linear compensation, performed with a filter of size 1024, as performance reference.

Computational complexity (in terms of operations per sample) of the hardware implementation of these nonlinear compensation algorithms is largely determined by these two parameters. The number of operations grows in complexity $\Theta(n)$ with the number of steps, and $\Theta(n \log n)$ with the size

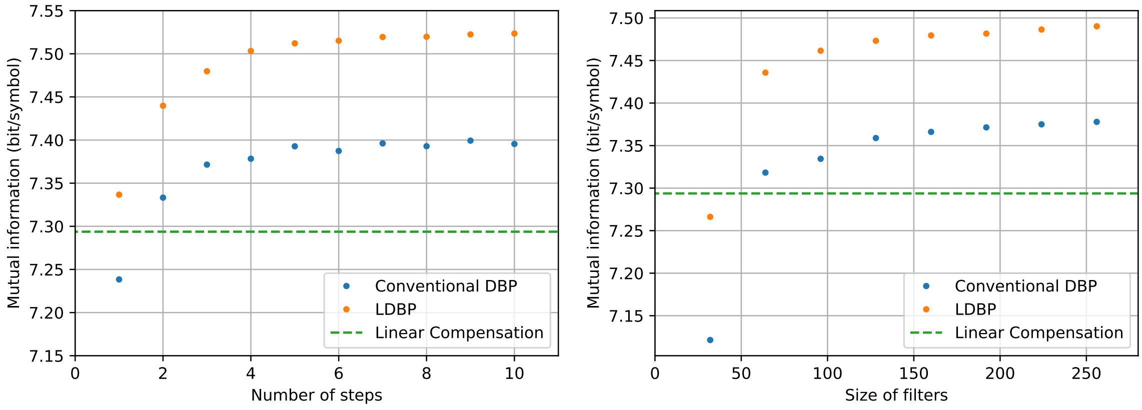

Fig. 3. Mutual information achieved with DBP and LDBP, varying the number of steps (left) and filter size (right). In the image varying the number of steps, a constant filter size = 1024 was considered. In the image varying the size of the filters a constant number of steps = 5 was considered.

of the filters [12].

Note that, in all cases, LDBP outperforms DBP by at least 0.1 bit/symbol. For one step, conventional DBP does not perform better than purely linear compensation, indicating the need for optimization of the nonlinear step. It is also noteworthy that it only takes two steps of LDBP to outperform 10 steps of conventional DBP. Both algorithms approximate maximum performance at around 5 steps, indicating that no more than five steps are needed for this particular link. This indicates that at most, nonlinear compensation would be only 5 times more computationally expensive than linear compensation, while delivering additional 0.2 bits/symbol for LDBP and 0.1 bit/symbol for DBP.

Complexity can be reduced even further by reducing the size of the filters at each step. Figure 3 also shows that filter sizes can be reduced down to 128 taps with mutual information penalties lower than 0.05 bit/symbol. In this case the complexity of the nonlinear compensation is comparable with linear compensation, while delivering 0.15 bit/symbol mutual information gain with LDBP and 0.05 bit/symbol gain with conventional DBP.

IV. Conclusions

We were able to evaluate the performance of learned digital back-propagation (LDBP) with experimental data from an unrepeatered optical link, attesting that in all scenarios the application of optimization to the linear and nonlinear coefficients can improve performance or, conversely, reduce required complexity in all cases, making digital back-propagation nonlinear compensation more attractive and viable. In all tested scenarios performance was increased by at least 0.1 bit/symbol when compared with the equivalent untrained DBP. We have shown that for this particular unrepeatered link, LDBP was capable of delivering 0.15 bit/symbol mutual information gain at similar complexity when compared with purely linear compensation.

References

[1] G. P. Agrawal, "Nonlinear fiber optics," in *Nonlinear Science at the Dawn of the 21st Century*. Springer, 2000, pp. 195–211.

[2] E. Ip and J. M. Kahn, "Compensation of dispersion and nonlinear impairments using digital backpropagation," *J. Lightwave Technol.*, vol. 26, no. 20, pp. 3416–3425, 10 2008.

[3] C. Häger and H. D. Pfister, "Nonlinear interference mitigation via deep neural networks," in *2018 Optical Fiber Communications Conference and Exposition (OFC)*, 3 2018, pp. 1–3.

[4] V. Oliari, S. Goossens, C. Häger, G. Liga, R. M. Bütler, M. v. d. Hout, S. v. d. Heide, H. D. Pfister, C. Okonkwo, and A. Alvarado, "Revisiting efficient multi-step nonlinearity compensation with machine learning: An experimental demonstration," *Journal of Lightwave Technology*, vol. 38, no. 12, pp. 3114–3124, 2020.

[5] C.-Y. Lin, R. Asif, M. Holtmannspoetter, and B. Schmauss, "Step-size selection for split-step based nonlinear compensation with coherent detection in 112-gb/s 16-qam transmission," *Chinese Optics Letters*, vol. 10, no. 2, p. 020605, 2012.

[6] T. Xu, G. Jacobsen, S. Popov, J. Li, E. Vanin, K. Wang, A. T. Friberg, and Y. Zhang, "Chromatic dispersion compensation in coherent transmission system using digital filters," *Opt. Express*, vol. 18, no. 15, pp. 16 243–16 257, Jul 2010. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-18-15-16243

[7] T. Xu, G. Jacobsen, S. Popov, J. Li, A. T. Friberg, and Y. Zhang, "Digital chromatic dispersion compensation in coherent transmission system using a time-domain filter," in *Asia Communications and Photonics Conference and Exhibition*, 2010, pp. 132–133.

[8] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, May 2015.

[9] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," *arXiv preprint arXiv:1505.00853*, 2015.

[10] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, "Tensorflow: A system for large-scale machine learning," in *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, 2016, pp. 265–283.

[11] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[12] C. K. Pun, S. Chan, K. Yeung, and K. Ho, "On the design and implementation of fir and iir digital filters with variable frequency characteristics," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 49, no. 11, pp. 689–703, 2002.