



## Revisiting Multi-Step Nonlinearity Compensation with Machine Learning

Downloaded from: <https://research.chalmers.se>, 2021-09-24 18:22 UTC

Citation for the original published paper (version of record):

Häger, C., Pfister, H., Büttler, R. et al (2019)

Revisiting Multi-Step Nonlinearity Compensation with Machine Learning

European Conference on Optical Communication (ECOC)

<http://dx.doi.org/10.1049/cp.2019.0947>

N.B. When citing this work, cite the original published paper.

# REVISITING MULTI-STEP NONLINEARITY COMPENSATION WITH MACHINE LEARNING

Christian Häger<sup>1,2\*</sup>, Henry D Pfister<sup>2</sup>, Rick M Bütler<sup>3</sup>, Gabriele Liga<sup>3</sup>, Alex Alvarado<sup>3</sup>

<sup>1</sup>Department of Electrical Engineering, Chalmers University, Gothenburg, Sweden

<sup>2</sup>Department of Electrical and Computer Engineering, Duke University, Durham, USA

<sup>3</sup>Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands

\*E-mail: christian.haeger@chalmers.se

**Keywords:** MACHINE LEARNING BASED DSP, DEEP LEARNING, LOW-COMPLEXITY DIGITAL BACK-PROPAGATION, SUBBAND PROCESSING, POLARIZATION MODE DISPERSION

## Abstract

For the efficient compensation of fiber nonlinearity, one of the guiding principles appears to be: fewer steps are better and more efficient. We challenge this assumption and show that carefully designed multi-step approaches can lead to better performance-complexity trade-offs than their few-step counterparts.

## 1 Introduction

Mitigating fiber nonlinearity is a significant challenge in high-speed fiber-optic communication systems. In principle, digital backpropagation (DBP) based on the split-step Fourier method (SSFM) offers ideal compensation of deterministic propagation impairments including nonlinear effects. On the other hand, several authors have highlighted the large computational burden associated with a real-time digital signal processing (DSP) implementation and proposed various techniques to reduce the complexity [1–9]. In many of these works, the number of steps (or compensation stages) is used not only to quantify complexity but also as a general measure of the quality for the proposed complexity-reduction method. In a nutshell, one gets the impression that fewer steps are better and more efficient.

While previous work has indeed demonstrated that complexity savings are possible by reducing steps, the main purpose of this paper is to highlight the fact that fewer steps are not more efficient *per se*. In fact, recent progress in machine learning suggests that deep computation graphs with many steps (or layers) tend to be more parameter efficient than shallow ones that use fewer steps. In this paper, we illustrate through various examples how this insight can be applied in the context of fiber-nonlinearity compensation in order to achieve low-complexity and hardware-efficient DBP.

## 2 Supervised Machine Learning

We start by reviewing the standard supervised learning setting for feed-forward neural networks (NNs). A feed-forward NN with  $M$  layers defines a mapping  $\mathbf{y} = f_{\theta}(\mathbf{x})$  where the input vector  $\mathbf{x} \in \mathcal{X}$  is mapped to the output vector  $\mathbf{y} \in \mathcal{Y}$  by alternating between affine transformations  $\mathbf{z}^{(i)} = \mathbf{W}^{(i)}\mathbf{x}^{(i-1)} + \mathbf{b}^{(i)}$  and pointwise nonlinearities  $\mathbf{x}^{(i)} = \phi(\mathbf{z}^{(i)})$  with  $\mathbf{x}^{(0)} = \mathbf{x}$  and  $\mathbf{x}^{(M)} = \mathbf{y}$ . The parameter vector  $\theta$  comprises all elements of the weight matrices  $\mathbf{W}^{(1)}, \dots, \mathbf{W}^{(M)}$  and vectors  $\mathbf{b}^{(1)}, \dots, \mathbf{b}^{(M)}$ . Given a training set  $S \subset \mathcal{X} \times \mathcal{Y}$  that contains

a list of desired input–output pairs, training proceeds by minimizing the empirical loss  $\mathcal{L}_S(\theta) \triangleq \frac{1}{|S|} \sum_{(\mathbf{x}, \mathbf{y}) \in S} \ell(f_{\theta}(\mathbf{x}), \mathbf{y})$ , where  $\ell(\hat{\mathbf{y}}, \mathbf{y})$  is the per-sample loss associated with returning the output  $\hat{\mathbf{y}}$  when  $\mathbf{y}$  is correct. When the training set is large, one typically optimizes  $\theta$  using a variant of stochastic gradient descent (SGD). In particular, mini-batch SGD uses the parameter update  $\theta_{t+1} = \theta_t - \alpha \nabla_{\theta} \mathcal{L}_{B_t}(\theta_t)$ , where  $\alpha$  is the step size and  $B_t \subseteq S$  is the mini-batch used in the  $t$ -th step.

Supervised machine learning is not restricted to NNs and learning algorithms such as SGD can be applied to other function classes as well. In this paper, we do not further consider NNs, but instead focus on approaches where the function  $f_{\theta}$  results from parameterizing model-based algorithms. In fact, prior to the current revolution in machine learning, communication engineers were quite aware that system parameters (such as filter coefficients) could be learned using SGD. It was not at all clear, however, that more complicated parts of the system architecture could be learned as well. For example, in the linear operating regime, polarization mode dispersion (PMD) can be compensated by choosing the function  $f_{\theta}$  as the convolution of the received signal with the impulse response of a linear multiple-input multiple-output (MIMO) filter, where  $\theta$  corresponds to the filter coefficients. For a suitable choice of the loss function  $\ell$ , applying SGD then maps into the well-known constant modulus algorithm [10]. More details and extensions of this approach are discussed in Sec. 3.3.

## 3 Multi-Step Digital Backpropagation

Signal propagation in an optical fiber is described by the nonlinear Schrödinger equation (NLSE). For efficient DBP, the task is to approximate the solution of the NLSE using as few computational resources as possible. The SSFM computes a numerical solution by alternating between linear filtering steps accounting for chromatic dispersion (CD) and nonlinear phase rotation steps accounting for the optical Kerr effect [11, p. 40]. This is quite similar to the functional form of a deep NN [9].

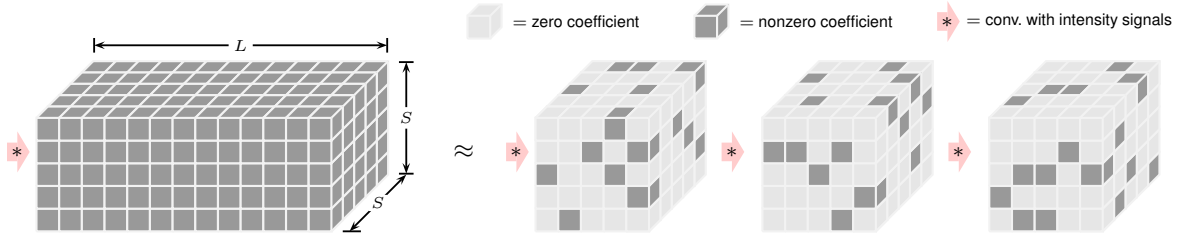


Fig. 2: Tensor representation of an  $L$ -tap  $S \times S$  MIMO filter for DBP based on subband processing, where  $S$  is the number of subbands (left); learned multi-step decomposition with sparse subfilters (right).

### 3.1 Short Chromatic-Dispersion Filters

Real-time DBP based on the SSFM is widely considered to be impractical due to the complexity of the fast Fourier transforms (FFTs) commonly used to implement frequency-domain (FD) CD filtering. To address this issue, time-domain (TD) filtering with finite impulse response (FIR) filters has been suggested in, e.g., [1, 8, 12, 13]. In these works, either a single filter or filter pair is designed and then used repeatedly in each step. However, using the same filter multiple times is suboptimal in general and the filter coefficients in the entire DBP algorithm should be optimized jointly. To that end, one can use supervised learning based on SGD by letting the function  $f_\theta$  be the SSFM, where the linear steps are implemented with FIR filters and  $\theta$  corresponds to the filter coefficients used in *all* steps [9, 14].

As an example, consider single-channel DBP of a 10-Gbaud signal over  $25 \times 80$  km of standard single-mode fiber using the SSFM with one step per span. For this scenario, Ip and Kahn have shown that 70-tap filters are required to obtain acceptable accuracy [1]. This assumes that the filters are designed using FD sampling and that the same filter is used in each step. The resulting hardware complexity was estimated to be over 100 times larger than for linear equalization. On the other hand, with jointly optimized filters, it is possible to achieve similar accuracy by alternating between filters that are as short as 5 and 3 taps [14]. This reduces the complexity by almost two orders of magnitude, making it comparable to linear equalization. Fig. 1 shows the DSP architecture assuming a folded implementation that takes advantage of filter symmetry.

At first glance, it may not be clear why multi-step DBP can benefit from a joint parameter optimization. After all, the standard SSFM applies *the same* CD filter many times in succession, without the need for any elaborate optimization. The explanation is that in the presence of practical imperfections such as finite-length filter truncation, applying the same *imperfect* filter multiple times can be detrimental because it magnifies any weakness. To achieve a good combined response of neighboring filters and a good overall response, the truncation of each filter needs to be delicately balanced.

### 3.2 Sparse MIMO Filters for Subband Processing

The complexity of DBP with TD filtering is largely dominated by the total number of required CD filter taps in all steps and this increases quadratically with bandwidth, see, e.g., [15, 16]. Thus, efficient TD-DBP of wideband signals is challenging.

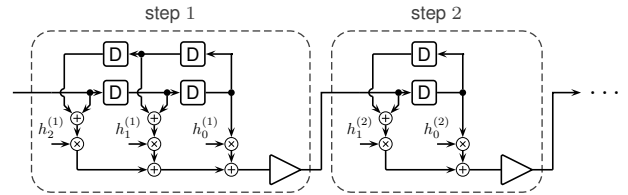


Fig. 1: Folded CD filters for efficient multi-step DBP. Triangles represent the (memoryless) nonlinear phase rotations.

One possible solution is to employ subband processing and split the received signal into  $S$  parallel signals using a filter bank [15–22]. A theoretical foundation for DBP based on subband processing is obtained by inserting the split-signal assumption  $u = \sum_{i=1}^S u_i$  into the NLSE. This leads to a set of coupled equations which can then be solved numerically. We focus on the modified SSFM proposed in [23] which is essentially equivalent to the standard SSFM for each subband, except that all sampled intensity waveforms  $|u_1|^2, \dots, |u_S|^2$  are jointly processed with a MIMO filter prior to each nonlinear phase rotation step. This accounts for cross-phase modulation between subbands but not four-wave mixing because no phase information is exchanged.

MIMO filters can be relatively demanding in terms of hardware complexity. As an example, consider the scenario in [22] where a 96-Gbaud signal is split into  $S = 7$  subbands. For a filter length of 13, the MIMO filter in each SSFM step can be represented as a  $7 \times 7 \times 13$  tensor with 637 real coefficients which is shown in Fig. 2 (left). The resulting complexity per step and subband would be almost 6 times larger than that of the CD filters used in [22]. The situation can be improved significantly by decomposing each MIMO filter into a cascade of sparse filters as shown in Fig. 2 (right). For a cascade of 3 filters, it was shown that a simple  $L_1$ -norm regularization applied to the filter coefficients during SGD leads to a sparsity level of round 8%, i.e., 92% of the filter coefficients can be set to zero with little performance penalty. Note that this filter decomposition happens *within* each SSFM step. In other words, complexity is reduced by further increasing the depth of the multi-step DBP computation graph.

### 3.3 Distributed PMD Compensation

Besides CD and fiber nonlinearity, PMD is another propagation impairment that can be compensated using DSP. PMD is

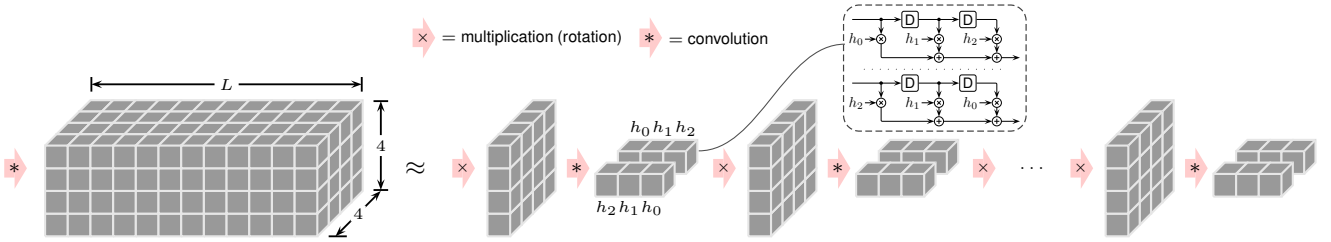


Fig. 3: Tensor representation of an  $L$ -tap  $4 \times 4$  MIMO filter for PMD compensation (left); multi-step decomposition where 4-D rotations are alternated with short fractional-delay (FD) filters accounting for DGD (right). Each FD filters is applied to both the real and imaginary part.

modeled by dividing the fiber into  $M$  sections, where the  $i$ -th section is described by a frequency-dependent Jones matrix  $\mathbf{R}^{(i)}\mathbf{J}^{(i)}(\omega)$ . Here,

$$\mathbf{J}^{(i)}(\omega) = \begin{pmatrix} e^{-j\omega\frac{\tau_i}{2}} & 0 \\ 0 & e^{j\omega\frac{\tau_i}{2}} \end{pmatrix} \quad (1)$$

is a first-order PMD matrix with differential group delay (DGD)  $\tau_i$  and  $\mathbf{R}^{(i)}$  is a complex unitary rotation matrix with determinant one. In the linear regime, the effect of PMD for the entire link is described by the matrix  $\mathbf{J}(\omega) = \prod_{i=1}^M \mathbf{R}^{(i)}\mathbf{J}^{(i)}(\omega)$  for sufficiently large  $M$ . PMD compensation (and polarization demultiplexing) then amounts to finding and applying the inverse  $\mathbf{J}^{-1}(\omega)$  to the received signal.

Unlike CD, PMD is a time-varying effect and thus requires adaptive compensation. In practical systems, adaptive PMD compensation is typically performed after CD compensation, e.g., using an  $L$ -tap MIMO filter that tries to approximate  $\mathbf{J}^{-1}(\omega)$ . Fig. 3 (left) shows the corresponding tensor representation assuming a real-valued  $4 \times 4$  filter that is applied to the separated real and imaginary parts of both polarizations [24]. An efficient multi-step decomposition of this filter is suggested in Fig. 3 (right), which essentially mimics the reverse propagation model by alternating memoryless rotations and first-order PMD. Here, the matrices (1) are approximated with two real-valued fractional-delay (FD) filters employing symmetrically flipped filter coefficients for different polarizations. The FD filters can be very short provided that the expected DGD per step is sufficiently small (i.e., many steps are used).

Initial results suggest that the multi-step PMD architecture can be effectively trained using SGD, similar to the conventional MIMO equalizer. However, more research is needed to fully characterize the training behavior, e.g., in terms of convergence speed for adaptive compensation. The main advantage of the multi-step architecture is that it can be easily integrated into DBP allowing for distributed PMD compensation with low hardware complexity. Combining PMD compensation with DBP has been previously studied for example in [25–27].

### 3.4 Coefficient Quantization and ASIC Implementation

Fixed-point requirements and other DSP hardware implementation aspects for DBP have been investigated in [8, 28–31]. A potential benefit of multi-step architectures is that they empirically tend to have many “good” parameter configurations that lie relatively close to each other. This implies that even if the

optimized parameters are slightly perturbed (e.g., by quantizing them) there may exist a nearby parameter configuration that exhibits similarly good performance to mitigate the resulting performance loss due to the perturbation.

Numerical evidence for this phenomenon can be obtained by considering the joint optimization of CD filters in DBP including the effect of filter coefficient quantization. This has been studied in [28] and the approach relies on applying so-called “fake” quantizations to the filter coefficients, where the gradient computations and parameter updates during SGD are still performed in floating point. Compared to other quantization methods, this jointly optimizes the responses of quantized filters and can lead to significantly reduced fixed-point requirements. For the scenario in [28], it was shown for example that the bit resolution can be reduced from 8–9 coefficient bits to 5–6 bits without adversely affecting performance. Furthermore, hardware synthesis results in 28-nm CMOS show that multi-step DBP based on TD filtering with short FIR filters is well within the limits of current ASIC technology in terms of chip area and power consumption [28, 29].

## 4 Conclusion

We have illustrated through various examples how machine learning can be used to achieve efficient fiber-nonlinearity compensation. Rather than reducing the number of steps, it was highlighted that complexity can also be reduced by carefully designing and optimizing multi-step methods, or even by increasing the number of steps and decomposing complex operations into simpler ones, without losing performance. We also avoided the use of neural networks as universal (but sometimes poorly understood) function approximators, and instead relied on parameterizing existing model-based algorithms.

## 5 Acknowledgements

This work is part of a project that has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 749798. The work of H. D. Pfister was supported in part by the National Science Foundation (NSF) under Grant No. 1609327. The work of A. Alvarado and G. Liga has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No. 757791). Any opinions, findings, recommendations, and conclusions expressed in this material are those of the authors and do not necessarily reflect the views of these sponsors.

## 6 References

- [1] E. Ip and J. M. Kahn, "Compensation of dispersion and nonlinear impairments using digital backpropagation," *J. Lightw. Technol.* **26**, 3416–3425 (2008).
- [2] L. B. Du and A. J. Lowery, "Improved single channel backpropagation for intra-channel fiber nonlinearity compensation in long-haul optical communication systems," *Opt. Express* **18**, 17,075–17,088 (2010).
- [3] D. Rafique, M. Mussolin, M. Forzati, et al., "Compensation of intra-channel nonlinear fibre impairments using simplified digital back-propagation algorithm," *Opt. Express* **19**, 9453–9460 (2011).
- [4] A. Napoli, Z. Maalej, V. A. J. M. Sleiffer, et al., "Reduced complexity digital back-propagation methods for optical communication systems," *J. Lightw. Technol.* **32** (2014).
- [5] A. M. Jarajreh, E. Giacoumidis, I. Aldaya, et al., "Artificial neural network nonlinear equalizer for coherent optical OFDM," *IEEE Photon. Technol. Lett.* **27**, 387–390 (2015).
- [6] E. Giacoumidis, S. T. Le, M. Ghanbarisabagh, et al., "Fiber nonlinearity-induced penalty reduction in CO-OFDM by ANN-based nonlinear equalization," *Opt. Lett.* **40**, 5113–5116 (2015).
- [7] M. Secondini, S. Rommel, G. Meloni, et al., "Single-step digital backpropagation for nonlinearity mitigation," *Photon. Netw. Commun.* **31**, 493–502 (2016).
- [8] C. Fougstedt, M. Mazur, L. Svensson, et al., "Time-domain digital back propagation: Algorithm and finite-precision implementation aspects," in "Proc. Optical Fiber Communication Conf.," (Los Angeles, CA, 2017).
- [9] C. Häger and H. D. Pfister, "Nonlinear interference mitigation via deep neural networks," in "Proc. Optical Fiber Communication Conf.," (San Diego, CA, 2018).
- [10] S. J. Savory, "Digital filters for coherent optical receivers," *Opt. Express* **16**, 804–817 (2008).
- [11] G. P. Agrawal, *Nonlinear Fiber Optics* (Academic Press, 2006), 4th ed.
- [12] L. Zhu, X. Li, E. Mateo, and G. Li, "Complementary FIR filter pair for distributed impairment compensation of WDM fiber transmission," *IEEE Photon. Technol. Lett.* **21**, 292–294 (2009).
- [13] G. Goldfarb and G. Li, "Efficient backward-propagation using wavelet-based filtering for fiber backward-propagation," *Opt. Express* **17**, 814–816 (2009).
- [14] C. Häger and H. D. Pfister, "Deep learning of the nonlinear Schrödinger equation in fiber-optic communications," in "Proc. IEEE Int. Symp. Information Theory," (Vail, CO, 2018).
- [15] M. G. Taylor, "Compact digital dispersion compensation algorithms," in "Proc. Optical Fiber Communication Conf.," (San Diego, CA, 2008).
- [16] K.-P. Ho, "Subband equaliser for chromatic dispersion of optical fibre," *Electronics Lett.* **45**, 1224–1226 (2009).
- [17] I. Slim, A. Mezghani, L. G. Baltar, et al., "Delayed single-tap frequency-domain chromatic-dispersion compensation," *IEEE Photon. Technol. Lett.* **25**, (2013).
- [18] M. Nazarathy and A. Tolmachev, "Subbanded DSP architectures based on underdecimated filter banks for coherent OFDM receivers: Overview and recent advances," *IEEE Signal Processing Mag.* **31**, 70–81 (2014).
- [19] E. F. Mateo, F. Yaman, and G. Li, "Efficient compensation of inter-channel nonlinear effects via digital backward propagation in WDM optical transmission," *Opt. Express* **18**, 15,144 (2010).
- [20] E. Ip, N. Bai, and T. Wang, "Complexity versus performance tradeoff for fiber nonlinearity compensation using frequency-shaped, multi-subband backpropagation," in "Proc. Optical Fiber Communication Conf.," (Los Angeles, CA, 2011), p. OThF4.
- [21] T. Oyama, H. Nakashima, T. Hoshida, et al., "Complexity reduction of perturbation-based nonlinear compensator by sub-band processing," in "Proc. Optical Fiber Communication Conf.," (2015).
- [22] C. Häger and H. D. Pfister, "Wideband time-domain digital backpropagation via subband processing and deep learning," in "Proc. European Conf. Optical Communication," (Rome, Italy, 2018).
- [23] J. Leibrich and W. Rosenkranz, "Efficient numerical simulation of multichannel WDM transmission systems limited by XPM," *IEEE Photon. Technol. Lett.* **15**, 395–397 (2003).
- [24] D. E. Crivelli, M. R. Hueda, H. S. Carrer, et al., "Architecture of a single-chip 50 Gb/s DP-QPSK/BPSK transceiver with electronic dispersion compensation for coherent optical channels," *IEEE Trans. Circuits Syst. I: Reg. Papers* **61**, 1012–1025 (2014).
- [25] K. Goroshko, H. Louchet, and A. Richter, "Overcoming performance limitations of digital back propagation due to polarization mode dispersion," in "Proc. Int. Conf. Transparent Optical Networks," (2016).
- [26] C. B. Czegledi, G. Liga, D. Lavery, et al., "Digital backpropagation accounting for polarization-mode dispersion," *Opt. Express* **25**, 1903–1915 (2017).
- [27] G. Liga, C. Czegledi, and P. Bayvel, "A PMD-adaptive DBP receiver based on SNR optimization," in "Proc. Optical Fiber Communication Conf.," (2018).
- [28] C. Fougstedt, C. Häger, L. Svensson, et al., "ASIC implementation of time-domain digital backpropagation with deep-learned chromatic dispersion filters," in "Proc. European Conf. Optical Communication," (Rome, Italy, 2018).
- [29] C. Fougstedt, L. Svensson, M. Mazur, et al., "ASIC implementation of time-domain digital back propagation for coherent receivers," *IEEE Photon. Technol. Lett.* **30**, 1179–1182 (2018).
- [30] C. S. Martins, L. Bertignono, A. Nespola, et al., "Efficient time-domain DBP using random step-size and multi-band quantization," in "Proc. Optical Fiber Communication Conf.," (San Diego, CA, 2018).
- [31] T. Sherborne, B. Banks, D. Semrau, et al., "On the impact of fixed point hardware for optical fiber nonlinearity compensation algorithms," *J. Lightw. Technol.* **36**, 5016–5022 (2018).