



# Advancing theoretical understanding and practical performance of signal processing for nonlinear optical communications through machine learning

Qirui Fan <sup>1</sup>✉, Gai Zhou<sup>1</sup>, Tao Gui<sup>1</sup>, Chao Lu<sup>2</sup> & Alan Pak Tao Lau <sup>1</sup>

In long-haul optical communication systems, compensating nonlinear effects through digital signal processing (DSP) is difficult due to intractable interactions between Kerr nonlinearity, chromatic dispersion (CD) and amplified spontaneous emission (ASE) noise from inline amplifiers. Optimizing the standard digital back propagation (DBP) as a deep neural network (DNN) with interleaving linear and nonlinear operations for fiber nonlinearity compensation was shown to improve transmission performance in idealized simulation environments. Here, we extend such concepts to practical single-channel and polarization division multiplexed wavelength division multiplexed experiments. We show improved performance compared to state-of-the-art DSP algorithms and additionally, the optimized DNN-based DBP parameters exhibit a mathematical structure which guides us to further analyze the noise statistics of fiber nonlinearity compensation. This machine learning-inspired analysis reveals that ASE noise and incomplete CD compensation of the Kerr nonlinear term produce extra distortions that accumulates along the DBP stages. Therefore, the best DSP should balance between suppressing these distortions and inverting the fiber propagation effects, and such trade-off shifts across different DBP stages in a quantifiable manner. Instead of the common ‘black-box’ approach to intractable problems, our work shows how machine learning can be a complementary tool to human analytical thinking and help advance theoretical understandings in disciplines such as optics.

<sup>1</sup> Photonics Research Center, Department of Electrical Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China.

<sup>2</sup> Photonics Research Center, Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China. ✉email: [remi.qr.fan@gmail.com](mailto:remi.qr.fan@gmail.com)

Optical communications are the backbone of all forms of information technology infrastructure in our modern society. As global Internet traffic grows by 60% per year<sup>1</sup>, research breakthroughs in optical communication speeds are much needed to meet the connectivity demands in the future. Fiber Kerr nonlinearity has long been the fundamental bottleneck for long-haul optical communications. While signal propagation dynamics in nonlinear optical fibers is well-known and governed by the nonlinear Schrödinger equation (NLSE), the interactions between fiber nonlinearity-induced self-phase modulation (SPM), CD, and inline optical amplifier noise prove to be very difficult to statistically characterize and compensate. The best and most common DSP to date for fiber nonlinearity compensation is the class of digital back-propagation (DBP)<sup>2</sup> algorithm and its variants<sup>3–5</sup> that provide reasonable performance gain (measured in terms of the improvement in bit error ratio (BER) or the corresponding quality (Q) factor in comparison with the case without nonlinearity compensation) in single-channel single-polarization systems. In practical polarization-division-multiplexed (PDM) wavelength-division-multiplexed (WDM) environments, however, single-channel DBP algorithms show negligible improvements due to interchannel cross-phase modulation (XPM) effects<sup>6,7</sup>. In addition, as different wavelength channels might have traveled through different links in a mesh network before arriving at the same receiver, interchannel nonlinear effects cannot be fully extracted from the received signals that significantly reduce the effectiveness of any joint-channel DBP that includes all the signals from neighboring channels, and attempts to compensate both SPM and XPM effects. Consequently, DBP is not widely deployed in commercial transceivers at present. Fundamentally, the interplay between CD, SPM, and amplified spontaneous emission (ASE) noise from inline amplifiers renders the system intractable and makes it difficult to derive optimal DSP algorithms.

ML has recently gained a lot of attention as a powerful tool in science and engineering problems that are virtually impossible to explicitly formulate. In optics, ML has been applied to enhance resolution of microscopy<sup>8</sup>, identify anthrax spores through quasi-phase imaging (QPI)<sup>9</sup>, and predict Internet network traffic<sup>10</sup>. ML is also applied to fiber nonlinearity compensation in optical communications. Various ML techniques, such as expectation maximization (EM)<sup>11</sup>, support vector machine (SVM)<sup>12,13</sup>, and message-passing algorithms<sup>5</sup>, were studied, but they show meaningful gains only for dispersion-managed links or OFDM signals, both of which are not default choices of technology in current long-haul digital coherent systems. For single-carrier systems, Kamalov et al.<sup>14</sup> conducted a field-trial demonstration using neural networks with information symbol triplets as inputs, but the performance is inferior to standard DBP. On the other hand, Häger and Pfister<sup>15–17</sup> considered the linear and nonlinear steps of DBP as a deep neural network (DNN) where preliminary simulation studies for single-channel single-polarization systems are presented. However, practical transmission impairments, such as laser-phase noise, laser-frequency offsets, polarization, and WDM effects, have not been studied. In addition, many ML applications in optical communications are impressive “black-box” data-driven models with unparalleled performance, but they contribute little additional insights into the problem concerned.

Here, we show a unique example of how ML can not only produce readily implementable algorithms advancing system performance, but also complement analytical thinking to develop deeper mathematical insights into fiber nonlinearity compensation. We first demonstrate performance gain in realistic experimental settings using the deep neural network-based digital back-propagation (DNN-based DBP) architecture<sup>15–17</sup>. To do so, we take into account the time series and dispersive nature of the

received signals, and appropriately integrate the DNN with other essential non-ML DSP blocks. We show that a low-complexity implementation of such DNN-based DBP demonstrates a 0.9-dB Q-factor gain compared with optimal DBP performance with arbitrary complexity for single-channel 28-GBaud 16-QAM systems. We further extended DNN-based DBP to PDM and WDM systems with dynamic polarization-state estimation. Low-complexity DNN-based DBP demonstrates a Q-factor gain of 0.6 dB and 0.25 dB over arbitrarily complex DBP for single-channel PDM and WDM PDM 28-GBaud 16-QAM systems, respectively. It should be emphasized that the DNN-based DBP is a single-channel DSP algorithm that provides performance gain in WDM environments, which serves as a key stepping stone toward practically implementable nonlinear compensation algorithms in a WDM environment. In addition, the optimized DNN-based DBP configurations reveal subtle mathematical structures that guide us to analyze the interplay between CD, nonlinearity, and noise. Such machine-learning-inspired analysis leads to a deeper insight that the optimal DSP should balance between compensating transmission impairments and additional distortions generated by the DBP itself. This is in contrast with typical ML applications in optical communications, which propose high-performance algorithms and bypass the need to further analyze the system at hand. The work is an example of the emerging area of interpretable machine learning<sup>18</sup> in the ML community where qualitative and human-understandable insights are gained from examining optimized ML configurations, which in turn help advance the theoretical understandings of the field concerned.

## Results

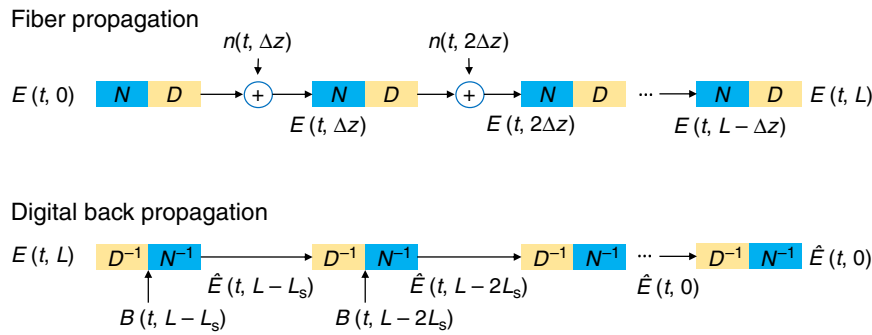
**Digital back propagation.** Let  $E(z, t)$  at  $z = 0$  be the electric field of a signal at the transmitter. In the simplest form, signal propagation in optical fibers is described by the stochastic scalar nonlinear Schrödinger equation (NLSE)

$$\begin{aligned} \frac{\partial E(z, t)}{\partial z} &= \underbrace{\left(-\frac{1}{2}\alpha - j\beta_2 \frac{1}{2} \frac{\partial^2}{\partial t^2}\right)}_{\mathbf{D}} E(z, t) + \underbrace{j\gamma |E(z, t)|^2}_{\mathbf{N}} E(z, t) + n(z, t) \\ &= (\mathbf{D} + \mathbf{N})E(z, t) + n(z, t), \end{aligned} \quad (1)$$

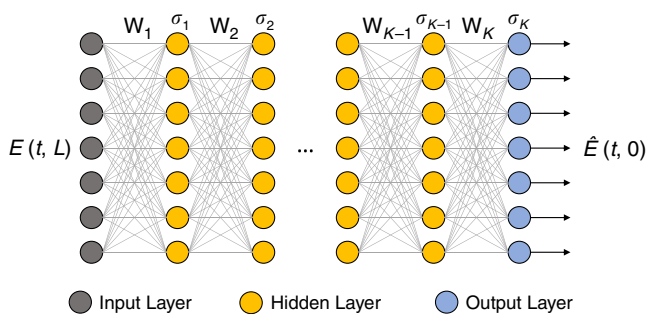
where  $n(z, t)$  are distributed ASE noise from inline amplifiers, which can be modeled as additive white Gaussian noise (AWGN) with zero mean and autocorrelation  $\mathbb{E}[n(z, t)n^*(z', t')] = \sigma^2 \delta(t - t')\delta(z - z')$ , where  $\delta$  is the Kronecker delta function. In this formulation,  $\mathbf{D}$  and  $\mathbf{N}$  are the linear and nonlinear operators, respectively, and  $\alpha$ ,  $\beta_2$ , and  $\gamma$  denote attenuation, group velocity dispersion, and fiber nonlinearity coefficient.

Signal propagations in optical fibers can be numerically simulated using the split-step Fourier method (SSFM) that interleaves the effect of CD/loss and nonlinear-phase rotation over a small length  $\Delta z$  of fiber. By separately applying the effect of CD/loss and nonlinearity to the signal, each step is analytically tractable. With digital coherent receivers, DBP is the standard technique to compensate fiber nonlinearity and is shown in Fig. 1. The DBP algorithm is a cascade of CD compensation filters  $\mathbf{D}^{-1}$  with transfer function  $H(\omega) = e^{-j\omega^2 \beta_2 L_s / 2}$ , where  $L_s$  is the step size of the DBP. The nonlinear-phase derotation operation  $\mathbf{N}^{-1}$  is defined by  $\sigma_k(x) = xe^{-j\gamma L_{\text{eff}} \xi_k |x|^2}$ , where  $L_{\text{eff}} = (1 - e^{-\alpha L_s})/\alpha$  is the effective length and  $\xi_k$  is a scaling factor. The operators  $\mathbf{D}^{-1}$  and  $\mathbf{N}^{-1}$  attempt to undo the linear and nonlinear effects of NLSE during fiber transmission.

In standard DBP implementations,  $\xi_k$  is the same for each stage of the DBP and is empirically optimized using brute-force approaches. The optimal value  $\xi_{\text{opt}}$  depends on the noise level,



**Fig. 1 Fiber propagation model and DBP structure.** The DBPs are interleaving operations of CD compensation and nonlinear-phase derotation. The choice of step size  $L_s$  is a trade-off between transmission performance and computational complexity.



**Fig. 2 DBP as DNN.** The digital back-propagation algorithm can be viewed as a deep neural network with interleaving linear and nonlinear operations  $W_k$  and  $\sigma_k$ , respectively. The parameters can be optimized by machine learning.

dispersion map among other factors. The choice of the step size  $L_s$  is a trade-off between complexity and performance, and is well-documented in the literature<sup>2</sup>. DBP has since been extended to the case of polarization-multiplexed transmissions<sup>19</sup>, stochastic DBP<sup>5</sup>, and joint-channel DBP<sup>20</sup> for WDM transmissions, and numerous simplification techniques have been proposed. Unfortunately, as of today, DBP is still relatively complex<sup>3,4,21,22</sup>, and for WDM systems, single-channel DBP provides marginal performance improvements.

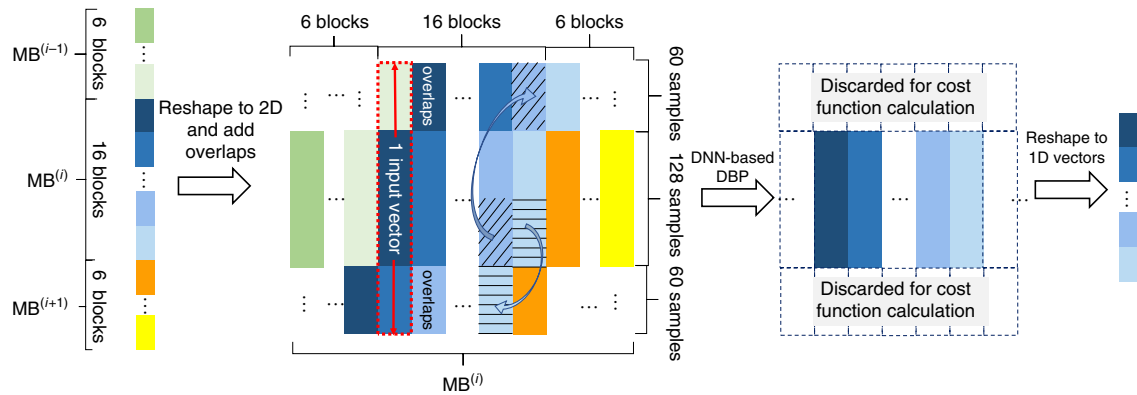
With the advent of machine learning and especially deep learning in recent years, one can put the familiar DBP algorithm in the lens of deep learning. Specifically, the interleaving linear and nonlinear steps of DBP can be seen as the linear and nonlinear operations of a multilayer neural network as shown in Fig. 2, in which the input is the received signal sample and the output is the estimated symbol sequence<sup>15</sup>. In this case, the nonlinear operation is basically the nonlinear-phase derotation. For a communication channel with linear effects, such as loss and chromatic dispersion, the linear operators resemble the effect of linear filters, which essentially implies that  $W_k$  are Toeplitz matrices. In this case, all the filter taps  $\mathbf{s}_k = [s_{k1} \ s_{k2} \ s_{k3} \ \dots]$  and  $\xi_k$  become parameters that can be optimized by machine-learning techniques.

**Training of DNN-based DBP.** While the concept of DNN-based DBP has been studied in the literature<sup>15,16</sup>, we propose several necessary modifications to enable performance gain in practical transmission experiments using DNN-based DBP. The input to the DNN-based DBP is derived from the coherently detected signal with sampling rate of 2 samples/symbol. As our random symbols are generated by 25 repetitions of pseudo random bit sequence (PRBS) with a period of 65,536 symbols, we use the first

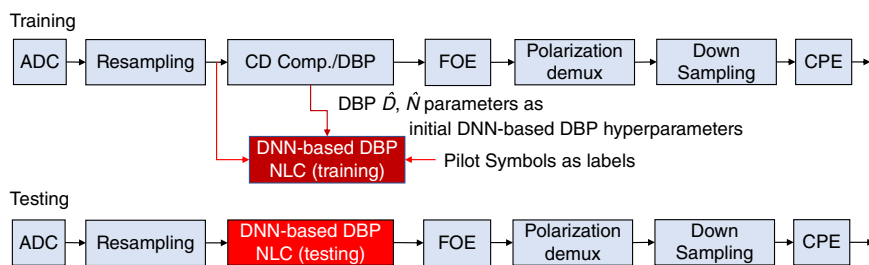
32,768 symbols (65,536 samples) for training and 25 copies of the other 32,768 symbols for testing to avoid repeating of training data in the testing set. The testing set contains a total of 819,200 symbols (3,276,800 bits for 16-QAM signals) for BER and Q-factor calculation. We apply a DNN model to our application by taking into account the time series and dispersive nature of the received signals, and use adjacent input vectors and training mini batches with overlapping signals, and use overlap-and-save<sup>23</sup> in the linear step of our implementation. In particular, we divide the 65,536 training samples into blocks of neural network inputs and jointly tuned the input vector size, linear filter length, and the initial- and final learning rates of AdaBound optimizer (to be described in more detail below). The results show that 121 taps and 128 samples per input, i.e., 512 total input vectors are optimal settings across most experimental setups studied in our work. With the amount of CD-induced pulse broadening in the transmission link, we append 60 neighboring samples to the two ends of each input to appropriately incorporate pulse-broadening effects in the overlap-and-save method, thus resulting in 512 input vectors of length 248 samples into the neural network as shown in Fig. 3. The number of layers in the deep neural network is equivalent to the number of DBP steps. The loss function, or objective function, we chose to optimize is the error vector magnitude (EVM) defined as the mean of  $\frac{|\hat{E}(t,0) - E(t,0)|^2}{|E(t,0)|^2}$  that is closely related to mean-squared error (MSE).

It should be noted that the signal-to-noise ratio (SNR) is essentially proportional to  $1/\text{EVM}$  or  $1/\text{MSE}$  and can be used as loss function in principle. We use EVM/MSE as the cost function as they are common in both optical communications and machine-learning community<sup>24,25</sup>. Typically, optimization of neural networks uses real-valued parameters, while communication signals are complex-valued. Therefore, we first convert all computations into real multiplications and additions so that the optimization task is supported by common deep-learning frameworks like Tensorflow<sup>26</sup>. Standard DBP configurations are used as the initializations of the deep neural network, i.e., CD compensation filter as the linear step with  $\xi_k = \xi_{\text{opt}}$  for all  $k$ . The batch size is optimized to 16. Since the correct overlapping waveform is needed for overlap-and-save in each linear step, we append 6 neighboring inputs before and after each batch of 16, so that the overlapping parts outside the batch are updated along with each batch at each step. This arrangement is similar to having overlapping data in neighboring mini batches. The whole batch of  $6 + 16 + 6 = 28$  inputs are processed together, while the output of the middle 16 vectors is used to calculate the cost function and update the parameters of the neural network as shown in Fig. 3.

We will first pass the input into standard DSP blocks, such as CD compensation or DBP, followed by the laser-frequency-offset compensation (FOE) by using the periodogram of the 4th power



**Fig. 3 Organizing the received signal sequence into data blocks and mini batches for DNN-based DBP.** For each block that contains 128 samples, 60 samples from both ends are appended to form an input vector. For 16 input vectors, 6 neighboring data blocks from both ends are appended to form the  $i$ th mini batch ( $MB^{(i)}$ ) containing 28 input vectors. After DNN-based DBP, only the middle 16 data blocks are used for cost function calculation during the training phase.



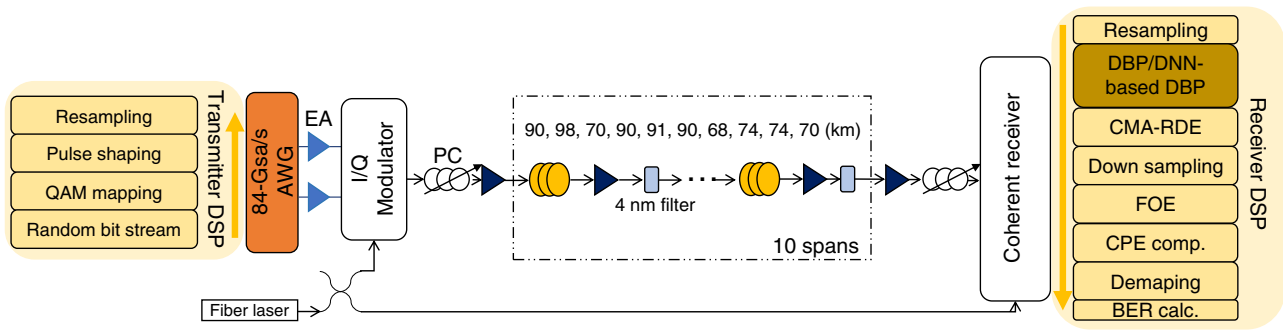
**Fig. 4 Detailed DSP structure for DNN-based DBP.** The training and testing phase of DNN-based DBP and other DSP blocks in digital coherent receiver are shown. FOE frequency-offset estimation, CPE carrier-phase estimation.

of the received QAM signal<sup>27</sup>, polarization demultiplexing (for polarization-multiplexed signals), and carrier-phase estimation (CPE) using blind-phase search (BPS) technique<sup>28</sup> as shown in Fig. 4. At the testing phase, DBP is replaced by the trained DNN-based DBP and the CPE output is used for BER calculation. For training the neural network, the Adam<sup>29</sup> optimizer is a popular choice, thanks to its rapid converging speed compared with stochastic gradient descent (SGD) and is adopted by Häger and Pfister<sup>15</sup>. However, we observed in our work that the optimized parameter settings using Adam are highly specific to each training dataset and generalized poorly for different experimental setups. In this connection, we chose the recently proposed AdaBound<sup>30</sup> optimizer that outperforms Adam in convergence stability, neural network performance, and generalizability to different datasets. This is achieved by specifying the desired initial and final learning rate so that the actual learning rate is bounded and smoothly transitioned between these two values in the AdaBound optimizer. In addition, we extended the neural network model in Häger and Pfister<sup>15</sup> to PDM systems, conduct PDM and WDM experiments, appropriately integrate the neural network with non-ML laser and polarization impairment compensation algorithms, and highlight how the optimized neural network configurations in turn help advance theoretical understandings of nonlinear signal–noise interactions during propagation and receiver signal processing, as will be shown next.

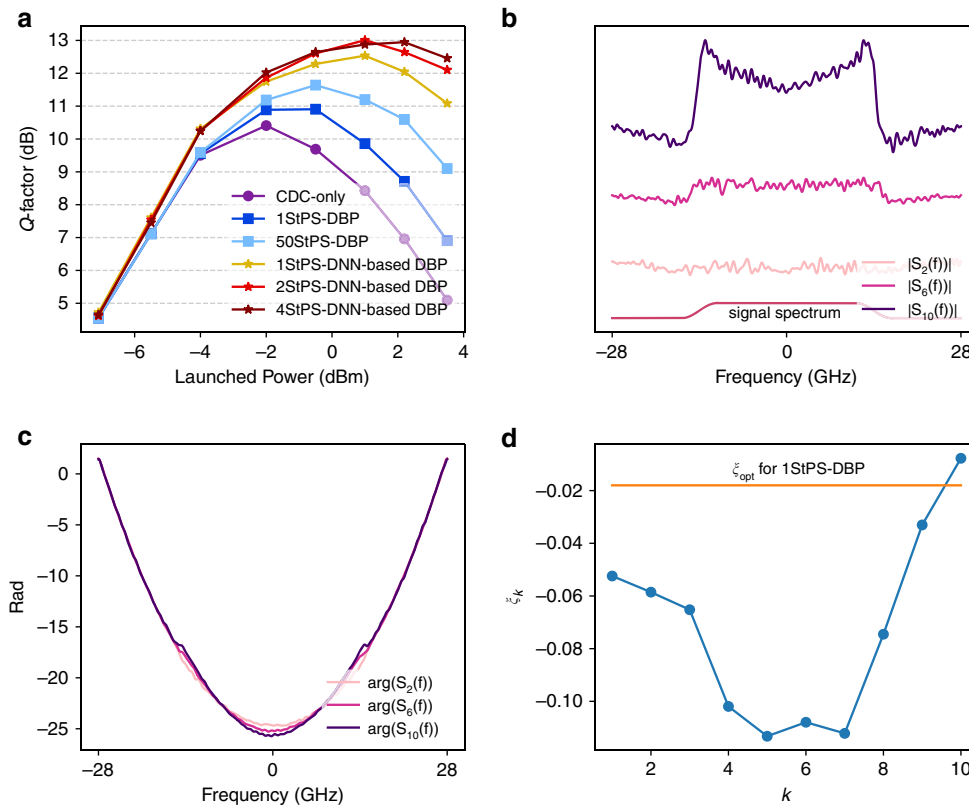
**DNN-based DBP for single-channel transmissions.** Experiments are conducted to determine the effectiveness of DNN-based DBP in practice. The experimental setup is shown in Fig. 5. At the transmitter side, a 92GSa/s arbitrary waveform generator (AWG) is used to generate 28-GBaud 16-QAM symbols shaped by squared raised cosine filter with roll-off factor of 0.2. The

electrical waveforms first go through SHF-807 high-bandwidth electrical amplifiers followed by I/Q modulator to modulate the optical signals. The modulated optical waveform was amplified and launched into the fiber link. A flat-top optical filter with 3-dB bandwidth of 4 nm is used in each span to suppress the out-of-band amplified spontaneous emission (ASE) noise to maximize the optical signal-to-noise ratio (OSNR). NKT Koheras ADJUSTIK fiber laser with linewidth around 100 Hz is used at both transmitter and receiver side. Since each span length is different, and our erbium-doped fiber amplifiers (EDFAs) have minimal gain of 20 dB (to compensate the loss of approximately 100 km of fiber), we fixed all EDFA gains to be 20 dB and use tunable attenuator integrated in each EDFA to ensure the proper equalization of the fiber loss incurred over all the spans. After 815-km transmission and polarization alignment between the signal and LO using a polarization controller, the optical waveform was coherently detected and sampled by an 80 GSa/s digital oscilloscope with 33-GHz electrical bandwidth. The sampled signals are then processed by offline DSP whose structure is shown in Fig. 5. The received signal is first resampled to two samples/symbol and digitally filtered to remove out-of-band noise before DBP/DNN-based DBP is applied. The Constant Modulus Algorithm (CMA) is used to compensate the residual linear distortion followed by downsampling to one sample/symbol and CPE for laser-phase noise compensation.

Figure 6a plots the comparison of  $Q$  factors (calculated from bit error ratio (BER) through  $Q = 20 \log_{10}(\sqrt{2} \text{erfc}^{-1}(2\text{BER}))$ ) as a function of signal-launched power using DNN-based DBP and DBP with different step sizes together with CD compensation (CDC) only. We consider 50 steps-per-span (StPS)–DBP as the benchmark for optimal DBP performance with arbitrary complexity as no further improvements are obtained beyond



**Fig. 5 Experimental setup for single-channel transmission.** The link consists of 10 spans of standard single-mode fibers (SSMF), each with an inline-lumped EDFA and a flat-top optical filter with 3-dB bandwidth of 4 nm. The span length varies from 68 km to 98 km with a total length of 815 km. EA electronic amplifier, AWG arbitrary waveform generator, FOE frequency-offset equalization, CPE carrier-phase estimation.



**Fig. 6 Results for single-channel single-polarization transmissions over 815 km.** **a** Q-factor vs. signal-launched power for CD compensation only, DBP, and DNN-based DBP with different computational complexities. **(b)** Optimized amplitude spectra of linear operators at different stages of 1-step-per-span (StPS) DNN-based DBP at 2.2-dBm signal-launched power. The spectra are vertically shifted from each other for visual clarity, and the signal spectrum is also shown. **(c)** Optimized phase response of linear operators at different DNN-based DBP stages at -1.4-dBm signal-launched power. **(d)** Optimized  $\xi_k$  at different DNN-based DBP stages at -1.4-dBm signal-launched power. While the optimized phase responses are close to quadratic that resembles CD compensation, the amplitude spectra are “M”-shaped. The optimized  $\xi_k$  is not the same for each step, and exhibits a “U”-shaped structure. Source data are provided as a Source Data file.

50 steps-per-span. It can be seen that 1-StPS DNN-based DBP already outperforms 1-StPS and 50-StPS DBP by 1.6 dB and 0.9 dB, respectively. The gain over 50-StPS DBP is further improved to about 1.3 dB when 2-StPS is used. Note that the performance improvements of DNN-based DBP originate from optimizing the parameters in DBP and it incurs no additional computational complexity i.e., 1-StPS DNN-based DBP has the same complexity as 1-StPS DBP, which is 50 times simpler than 50-StPS DBP. Note also that such comparisons do not take into account the complexity of training the DNN-based DBP since the

training is typically performed offline. In addition, as DNN-based DBP implicitly learns the best DSP parameters for a given link, it can readily be applied to links with heterogeneous spans, and hence DNN-based DBP is a link-agnostic learning algorithm particularly suitable for practical deployed systems. The DNN-based DBP is optimized by using 200 epochs with step size 0.01, and the converged linear filter spectra  $S_k(f)$  and phase derotation coefficients  $\xi_k$  are shown in Fig. 6b–d at 2.2-dBm signal-launched power. It can be seen that the phase responses largely resemble quadratic phase for CD compensation. However, the converged

amplitude responses exhibit an “M”-shaped feature and become more apparent at later stages of the DNN-based DBP. Furthermore, the optimal  $\xi_k$  learnt by machine learning are not constant across  $k$  and not equal to  $\xi_{\text{opt}}$  in general. Rather,  $\xi_k$  have a “U”-shaped structure so that the nonlinear-phase derotation is larger in the middle stages. Finally, it should be noted that similar performance gains and features are obtained when using external cavity lasers (ECL) with linewidth  $\sim 100$  kHz at the transmitter and receiver by modifying the cost function to that of radius-directed equalization (RDE)<sup>31</sup>.

**Mathematical insights into fiber nonlinearity compensation as interpretable ML.** Why are the amplitude responses of the linear filter “M”-shaped and why is the optimal  $\xi_k$  “U”-shaped? In most applications, ML are “black-box” models with excellent predictive power, but it is difficult to inquire how they work exactly and why they produce such good results. However, in our case, the “M”-shape and “U”-shape features are clear mathematical structures that strongly suggest certain hidden dynamics of DBP yet to be better understood. Specifically, the “M”-shaped filter indicates that the optimal linear filter exhibits some high-pass feature that becomes more pronounced at later stages of the DNN-based DBP. A plausible explanation for this phenomenon is that there exists an additional undesired term with a “ $\cap$ ”-shaped spectrum that grows with the DNN-based DBP stages such that the “M”-shaped filter tries to compensate. On the other hand, the “U”-shaped  $\xi_k$  suggests that the nonlinear-phase derotation is small at the beginning and toward the end of the DBP, but can be larger in the middle stages. This can imply that the beginning and end stages of DBP are more prone to noise and distortions, and hence phase derotation based on instantaneous signal power may not be effective. The overall optimized DNN-based DBP configuration from ML seems to suggest that noise and distortion accumulations play a hidden yet pivotal role in the nature of nonlinear DSP and effectiveness of DNN-based DBP.

We proceed to analyze noise and distortion accumulations in DBP to try to explain the optimal DNN-based DBP configurations discovered by machine learning. For simplicity, we will remove the variable  $t$  and abbreviate  $E(z, t)$  as  $E_z$  and let the step size  $L_s = \Delta z$  in the following analysis. Referring to Fig. 1, the received signal is given by

$$\begin{aligned} E_L &= \text{CD}^{\Delta z} \left( E_{L-\Delta z} e^{j\gamma\Delta z |E_{L-\Delta z}|^2} \right) + n_L \\ B_{L-\Delta z} &= E_{L-\Delta z} e^{j\gamma\Delta z |E_{L-\Delta z}|^2} + n'_L, \end{aligned} \tag{2}$$

where  $\text{CD}^{\Delta z}(\cdot)$  denotes the effect of CD on a signal over a distance  $\Delta z$  and  $n'_L = \text{CD}^{-\Delta z}(n_L)$ . This is followed by a phase derotation proportional to  $\xi_{L-\Delta z} |B_{L-\Delta z}|^2$ . Note that as supposed to  $\xi_k$  that refers to the  $k$ th-step DNN-based DBP in the previous section,  $\xi_{L-\Delta z}$  here relates to the phase derotation to estimate the transmitted signal at  $z = L - \Delta z$ , which is given by

$$\begin{aligned} \hat{E}_{L-\Delta z} &= B_{L-\Delta z} e^{-j\gamma\Delta z \xi_{L-\Delta z} |B_{L-\Delta z}|^2} \\ &= \left( E_{L-\Delta z} e^{j\gamma\Delta z |E_{L-\Delta z}|^2} + n'_L \right) e^{j\gamma\Delta z \xi_{L-\Delta z} |E_{L-\Delta z} e^{j\gamma\Delta z |E_{L-\Delta z}|^2} + n'_L|^2} \\ &= E_{L-\Delta z} e^{j\gamma\Delta z [|E_{L-\Delta z}|^2 - \xi_{L-\Delta z} |E_{L-\Delta z} + n'_L|^2]} + n''_L, \end{aligned} \tag{3}$$

where  $n''_L = n'_L e^{-j\gamma\Delta z |E_{L-\Delta z}|^2}$  and  $n'''_L = n'_L e^{j\gamma\Delta z \xi_{L-\Delta z} |E_{L-\Delta z} e^{j\gamma\Delta z |E_{L-\Delta z}|^2} + n'_L|^2}$ . Since we are interested in the statistical properties of the estimated transmitted signal only and  $n_L, n'_L, n''_L$ , and  $n'''_L$  are all AWGN processes with the same statistical distributions, we will denote all of

them as  $n_L$  for simplicity so that

$$\begin{aligned} \hat{E}_{L-\Delta z} &= E_{L-\Delta z} e^{j\gamma\Delta z [|E_{L-\Delta z}|^2 - \xi_{L-\Delta z} |E_{L-\Delta z} + n_L|^2]} + n_L \\ &\approx E_{L-\Delta z} + j\gamma\Delta z \left[ (1 - \xi_{L-\Delta z}) |E_{L-\Delta z}|^2 E_{L-\Delta z} \right. \\ &\quad \left. - 2\xi_{L-\Delta z} \Re \left\{ E_{L-\Delta z} n_L^* \right\} E_{L-\Delta z} \right] + n_L. \end{aligned} \tag{4}$$

In this case, the extra distortions arising from fiber nonlinearity and its compensation are  $|E_{L-\Delta z}|^2 E_{L-\Delta z}$  and  $\Re \{ E_{L-\Delta z} n_L^* \} E_{L-\Delta z}$ . Note that for typical pulse shapes,  $|E_{L-\Delta z}|^2 E_{L-\Delta z}$  will have a “ $\cap$ ”-shaped spectrum due to the triple convolution of the original pulse’s spectrum. Also, the amplifier noise  $n_L$  at  $z = L$  is brought backward to the signal estimate  $\hat{E}_{L-\Delta z}$  at  $z = L - \Delta z$  through  $\Re \{ E_{L-\Delta z} n_L^* \} E_{L-\Delta z}$ . Next, we denote  $E_{L-\Delta z} = \text{CD}^{\Delta z} (E_{L-2\Delta z} e^{j\gamma\Delta z |E_{L-2\Delta z}|^2}) + n_{L-\Delta z}$  so that

$$\begin{aligned} B_{L-2\Delta z} &= \text{CD}^{-\Delta z} (\hat{E}_{L-\Delta z}) \\ &= E_{L-2\Delta z} e^{j\gamma\Delta z |E_{L-2\Delta z}|^2} \\ &\quad + j\gamma\Delta z \left[ (1 - \xi_{L-\Delta z}) \text{CD}^{-\Delta z} (|E_{L-\Delta z}|^2 E_{L-\Delta z}) \right. \\ &\quad \left. - 2\text{CD}^{-\Delta z} \left( \Re \left\{ E_{L-\Delta z} n_L^* \right\} E_{L-\Delta z} \right) \right] + v_{\Delta z}, \end{aligned} \tag{5}$$

where  $v_{\Delta z} = \text{CD}^{-\Delta z} (n_L + n_{L-\Delta z})$ . Note that the statistics of  $v_{\Delta z}$  does not depend on the CD operation, and we will also denote  $v_0 = n_L$  for notational convenience. The signal estimate at  $z = L - 2\Delta z$  is given by

$$\begin{aligned} \hat{E}_{L-2\Delta z} &= B_{L-2\Delta z} e^{-j\gamma\Delta z \xi_{L-2\Delta z} |B_{L-2\Delta z}|^2} \\ &\approx E_{L-2\Delta z} + j\gamma\Delta z \left[ (1 - \xi_{L-2\Delta z}) |E_{L-2\Delta z}|^2 E_{L-2\Delta z} \right. \\ &\quad \left. - 2\xi_{L-2\Delta z} \Re \{ E_{L-2\Delta z} v_{\Delta z}^* \} E_{L-2\Delta z} + (1 - \xi_{L-2\Delta z}) \text{CD}^{-\Delta z} \right. \\ &\quad \left. (|E_{L-\Delta z}|^2 E_{L-\Delta z}) - 2\xi_{L-\Delta z} \text{CD}^{-\Delta z} \left( \Re \{ E_{L-\Delta z} v_0^* \} E_{L-\Delta z} \right) \right] + v_{\Delta z}. \end{aligned} \tag{6}$$

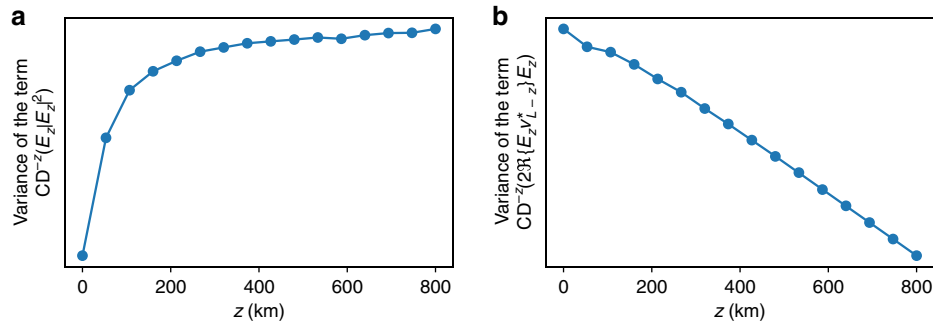
A mathematical pattern is emerging from Eq. (6). Continuing with the above derivation, the estimate of the transmitted signal using DBP is

$$\begin{aligned} \hat{E}_0 &= E_0 + j\gamma\Delta z \left[ \sum_{k=1}^K (1 - \xi_{L-k\Delta z}) \text{CD}^{-(L-k\Delta z)} (|E_{L-k\Delta z}|^2 E_{L-k\Delta z}) \right. \\ &\quad \left. - 2\xi_{L-k\Delta z} \text{CD}^{-(L-k\Delta z)} \left( \Re \{ E_{L-k\Delta z} v_{k\Delta z}^* \} E_{L-k\Delta z} \right) \right] + v_{K\Delta z}, \end{aligned} \tag{7}$$

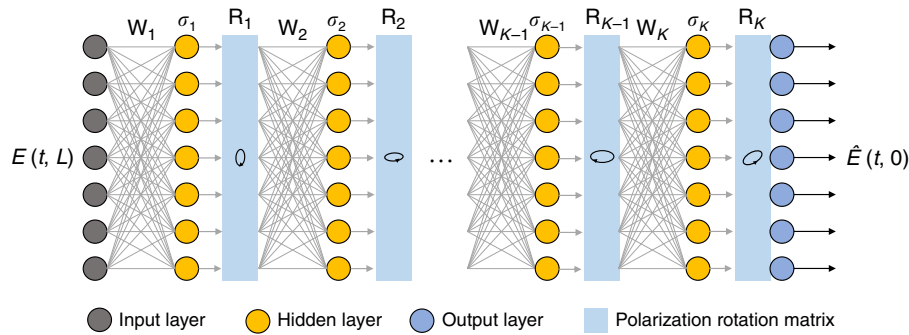
where  $K$  is the total number of DBP steps. As  $K \rightarrow \infty$  and  $\Delta z \rightarrow 0$  for distributed amplification system with arbitrarily complex DBP,

$$\begin{aligned} \hat{E}_0 &= E_0 + j\gamma \int_0^L (1 - \xi_z) \text{CD}^{-z} (|E_z|^2 E_z) \\ &\quad - 2\xi_z \text{CD}^{-z} \left( \Re \{ E_z v_{L-z}^* \} E_z \right) dz + v_L. \end{aligned} \tag{8}$$

In Eq. (8),  $v_L$  corresponds to the total ASE noise in the system and the term  $j\gamma \int_0^L \text{CD}^{-z} (|E_z|^2 E_z) dz$  actually corresponds to the nonlinear-phase shift due to the signal, and is largely compensated by carrier-phase estimation in practical systems. The other terms  $j\gamma \int_0^L \xi_z \text{CD}^{-z} (|E_z|^2 E_z) dz$  and  $j\gamma \int_0^L 2\xi_z \text{CD}^{-z} \left( \Re \{ E_z v_{L-z}^* \} E_z \right) dz$  are the major nonlinear impairments that degrade transmission performance. As the variances of the noises and distortions are typically used to characterize overall system performance, Fig. 7 shows the simulated variances of  $\text{CD}^{-z} (|E_z|^2 E_z)$  and  $2\text{CD}^{-z} \left( \Re \{ E_z v_{L-z}^* \} E_z \right)$  as a function of  $z$ , and it can be seen that



**Fig. 7 Evolution of different distortion terms in Eq. (8) with propagation distance  $z$ .** Variance of the term (a)  $CD^{-z}(|E_z|^2 E_z)$  and (b)  $2CD^{-z}(\Re\{E_z v_{L-z}^*\} E_z)$  as a function of  $z$  for a 23-GBaud 16-QAM system transmitted over 815 km obtained from simulations. The signal-launched power is  $-1.4$  dBm with 19-dB OSNR. It can be seen that the variance of  $CD^{-z}(|E_z|^2 E_z)$  is an increasing function of  $z$ , while the variance of  $2CD^{-z}(\Re\{E_z v_{L-z}^*\} E_z)$  is a decreasing function of  $z$ . Source data are provided as a Source Data file.



**Fig. 8 DBP-based DNN for polarization-division-multiplexed system.** The DNN-based DBP contains a PSP rotation matrix  $R_k$  in each stage to partially compensate polarization effects.

one of them increases with  $z$ , while the other decreases with  $z$ . As  $\xi_z$  control the relative strengths of  $CD^{-z}(|E_z|^2 E_z)$  and  $2CD^{-z}(\Re\{E_z v_{L-z}^*\} E_z)$  in the overall nonlinear distortions in  $\hat{E}_0$ , we can now appreciate why  $|\xi_z|$  is smaller at the beginning and end of the DNN-based DBP stages and larger in the middle as shown in Fig. 6d. This can be intuitively interpreted by noting that

1. The nonlinear-phase derotation at each DBP stage is not perfect due to ASE noise, and such imperfections accumulate throughout the whole DBP chain. Imperfections at the early DBP stages (corresponding to  $\xi_z$  for  $z \sim L$ ) accumulate the most in the final signal estimate  $\hat{E}_0$  and therefore  $\xi_z$  should be small for  $z \sim L$  to minimize such accumulation.
2. Toward the end of the DBP, the signal amplitude is already heavily distorted and quite different from the original signal due to noise and accumulation of imperfect compensation from preceding DBP stages. Therefore, the phase derotation at the end of the DBP stages (corresponding to  $\xi_z$  for  $z \sim 0$ ) will not be accurate, and hence  $\xi_z$  should be small for  $z \sim 0$  to prevent overcompensation and in turn produce additional distortions.

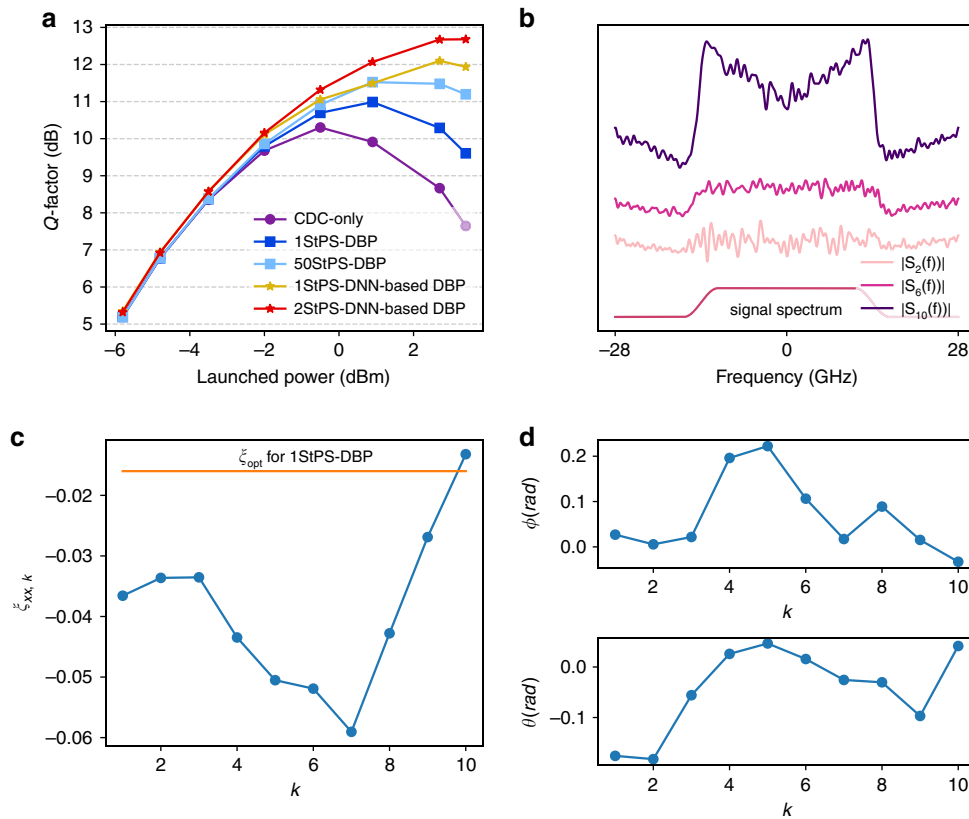
In addition, since the imperfect-phase derotations cannot completely eliminate the nonlinear distortion term  $|E_z|^2 E_z$ , they continue to grow within the DBP stages and accumulate at the end of the algorithm. As  $|E_z|^2 E_z$  has a “ $\cap$ ”-shaped spectrum, an inverted-shaped spectrum to partially equalize the distortions will be more beneficial than a pure CD compensation operation in Eq. (8). Furthermore, since  $|E_z|^2 E_z$  has three times the bandwidth of  $E_z$ , one should simply filter out the out-of-band distortions at each DBP stage. This explains why the overall linear filter of DNN-based DBP exhibits the “M”-shaped features depicted in Fig. 6b, and how the shape is more apparent toward the later stages of DNN-based DBP.

The new insights developed illustrate that the optimal linear filter of DNN-based DBP does not merely equalize CD. It is in fact a trade-off between compensating CD of the signal and mitigating the third-order nonlinear distortion term  $|E_z|^2 E_z$  and its accumulation along the DBP stages. Similarly, the optimal nonlinear-phase derotation actually attempts to strike a balance between reversing the nonlinear phase during propagation and minimizing additional phase noise accumulation along the DBP stages due to corrupted signal power levels. Overall, machine learning reveals that the original design philosophy of DBP as an iterative linear and nonlinear compensation of fiber propagation effects is not complete. Rather, the optimal DSP should undo the nonlinear channel effects, as well as manage additional distortions accumulated within the DSP itself. In our work, it should be emphasized that the new insights are inspired from the ML-optimized configurations depicted in Fig. 5. In our work, ML provided directions for theoreticians to work toward deeper analytical insights, which in turn validate the consistency and reliability of the DNN-based DBP as we are able to interpret the features with concrete mathematical arguments.

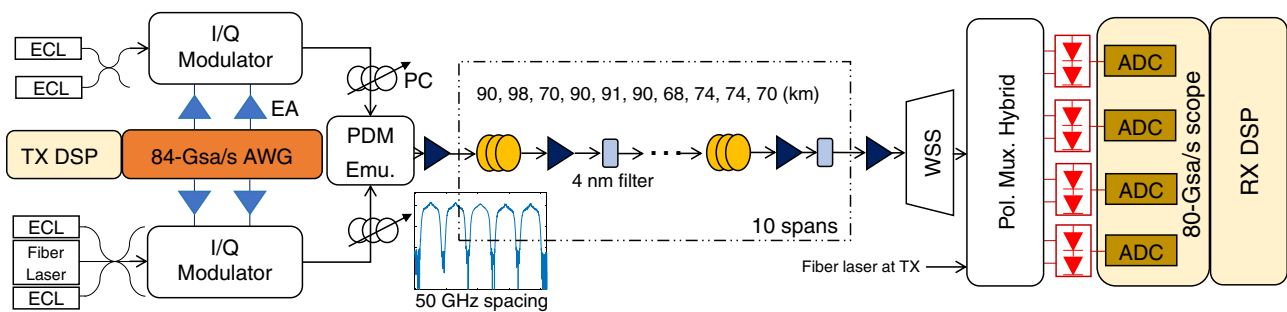
**DNN-based DBP for PDM and WDM transmissions.** For PDM transmissions with vector input  $\mathbf{E}(z, t) = [E_x(z, t) \ E_y(z, t)]^T$ , the vector Manakov–PMD equation

$$\frac{\partial \Psi}{\partial z} = \left( -\frac{1}{2} \alpha - j\beta_0 - \beta_1 \frac{\partial}{\partial t} - j\beta_2 \frac{1}{2} \frac{\partial^2}{\partial t^2} \right) \Psi + j \frac{8}{9} \gamma |\Psi|^2 \Psi + \begin{bmatrix} n_x(z, t) \\ n_y(z, t) \end{bmatrix} \quad (9)$$

governs signal propagation. In this formulation,  $\Psi = \mathbf{U}^{-1}(z)\mathbf{E}$  where  $\mathbf{U}(z)$  models the random principle states of polarization



**Fig. 9 Results for single-channel PDM transmissions over 815 km. a** Q-factor vs. signal-launched power for CD compensation only, DBP, and DNN-based DBP with different computational complexities for single-channel PDM 16-QAM transmissions over 815 km. **b** Optimized amplitude spectra of linear operators at different stages for 1-StPS DNN-based DBP at signal-launched power of 2.7 dBm. The spectra are vertically shifted from each other for visual clarity. **c** Optimized  $\xi_{xx,k}$  at different stages of DNN-based DBP at signal-launched power of 2.7 dBm. **d** Optimized rotation angles  $\theta_k$  and  $\phi_k$ . Source data are provided as a Source Data file.



**Fig. 10 Experimental setup for PDM-WDM transmissions.** Four more external cavity lasers with 50-GHz channel spacing are added to realize a five-channel WDM system. WSS: wavelength-selective switch.

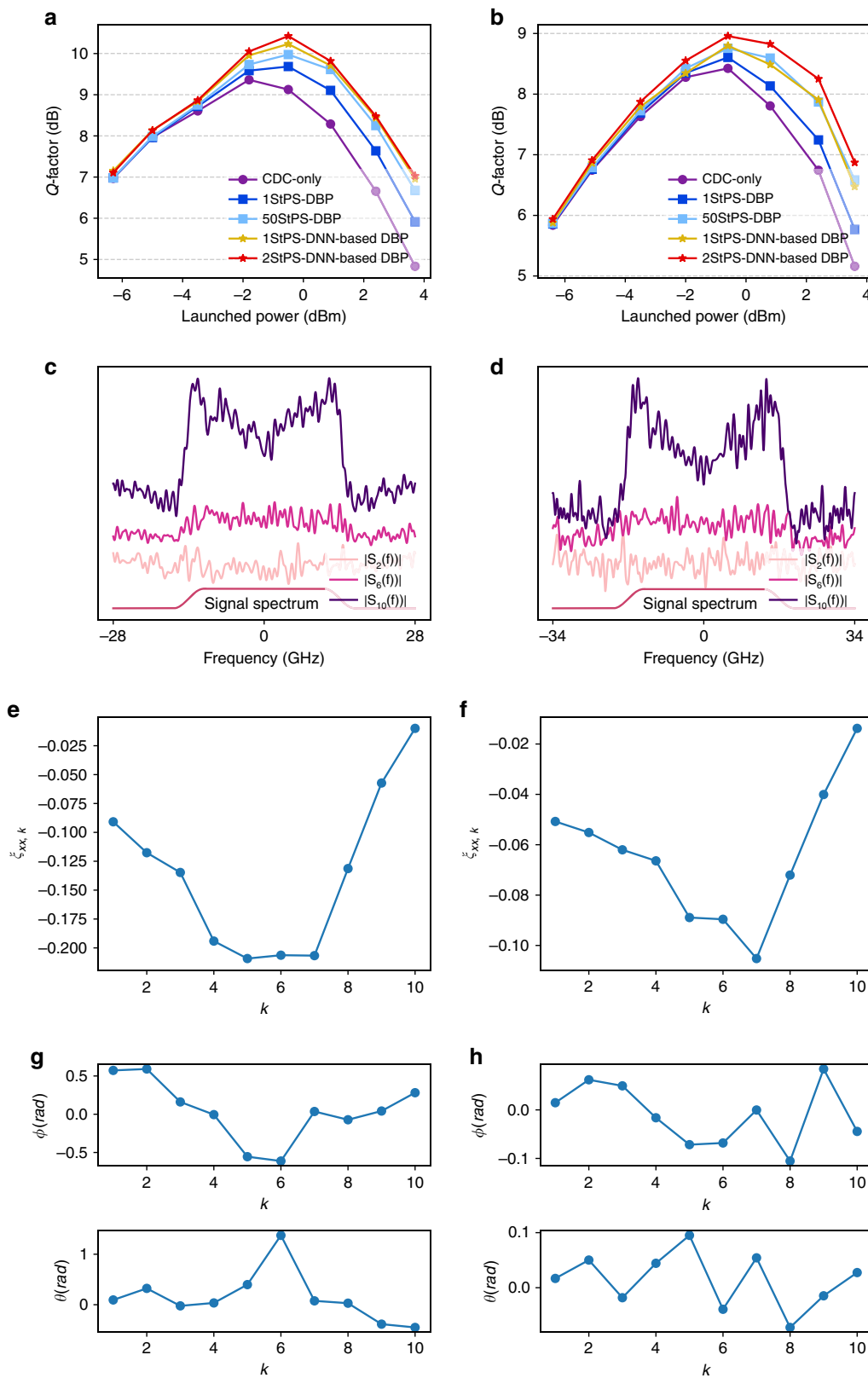
(PSP) rotation that evolve with  $z$ .  $\beta_0, \beta_1$  are  $2 \times 2$  matrices modeling birefringence (difference in refractive index of signals in the two PSPs) and polarization-mode dispersion (PMD) (difference in group refractive index induced group delay of signals in the two PSPs). Random polarization rotations and PMD lead to polarization-dependent nonlinear interactions. Although their effects in nonlinear fiber propagation are well characterized analytically, their unknown and random nature significantly reduce the effectiveness of DBP<sup>19,32</sup>. In this connection, the proposed DNN-based DBP framework can be extended to roughly estimate and compensate the polarization rotations at each DBP stage. In particular, we can express the DNN-based

DBP in a vector form and append a polarization rotation matrix

$$\mathbf{R}_k = \begin{bmatrix} \cos \theta_k \cos \phi_k - j \sin \theta_k \sin \phi_k & -\sin \theta_k \cos \phi_k + j \cos \theta_k \sin \phi_k \\ \sin \theta_k \cos \phi_k + j \cos \theta_k \sin \phi_k & \cos \theta_k \cos \phi_k + j \sin \theta_k \sin \phi_k \end{bmatrix} \quad (10)$$

at the  $k$ th stage along with the linear and nonlinear operators in each DNN-based DBP step as shown in Fig. 8. Note that  $\mathbf{R}_k$  is optimized from training data without any prior knowledge of the link PSP. In this case, the nonlinear step  $\sigma_k(\cdot)$  will consist of two optimization parameters  $\xi_{xx,k} = \xi_{yy,k}$  and  $\xi_{xy,k} = \xi_{yx,k}$ .





**Fig. 11 Results for PDM-WDM transmissions over 815 km.** Q-factor vs. signal-launched power for CD compensation only, DBP, and DNN-based DBP with different computational complexities for (a) 28-Gbaud and (b) 34-Gbaud 16-QAM signals. Optimized amplitude spectra of linear operators at different stages for 1-StPS DNN-based DBP at signal-launched power of 2.4 dBm for (c) 28-Gbaud and (d) 34-Gbaud signals. The spectra are vertically shifted for visual clarity. Optimized  $\xi_{xx,k}$  at different stages of DNN-based DBP at signal-launched power of 2.4 dBm for (e) 28-Gbaud and (f) 34-Gbaud signals. Optimized rotation angles  $\theta_k$  and  $\phi_k$  for (g) 28-Gbaud and (h) 34-Gbaud signals. Source data are provided as a Source Data file.

Figure 9a compares the Q factors as a function of signal-launched power for 28-GBaud PDM 16-QAM transmissions over 815 km using CDC-only, DNN-based DBP and DBP with different step sizes. PDM is achieved in experimental settings through a PDM emulator that splits the original signal into two polarizations with equal energy, delays one of the signals to generate delayed signal copies and recombines them. The DNN-based DBP here is trained on the PDM data. It can be seen from Fig. 9a that 1-StPS DBP can only provide a small gain of 0.7 dB compared with CDC only, and 50-StPS can provide a further gain of 0.5 dB over 1-StPS DBP. However, a mere 1-StPS DNN-based DBP can already produce an extra 0.6-dB gain over 50-StPS DBP, which shows that DNN-based DBP can not only save complexity but improve transmission performance in polarization-multiplexed systems. The 2-StPS DNN-based DBP can further improve the performance by 0.6 dB over 1-StPS DNN-based DBP. The amplitude spectra and optimized  $\xi_{xx,k}$  shown in Fig. 9b, c exhibit the “M”-shaped and “U”-shaped features as discussed previously. The phase spectra have the same quadratic shape as CD compensation filter and are not shown here. The optimized angles  $\theta_k$  and  $\phi_k$  shown in Fig. 9d do not exhibit any mathematical structures or trends, and are in agreement with theoretical expectations. It is clear that including  $R_k$  in DNN-based DBP provides new dimensions of optimization, and hence DNN-based DBP can overcome the limitations of traditional DBP by compensating polarization-dependent nonlinear effects.

Finally, we investigate DNN-based DBP performance for PDM 16-QAM WDM transmissions with different baud rates. The 5-channel 50-GHz-spaced WDM transmission setup is shown in Fig. 10. The low-linewidth fiber laser is combined with two external cavity lasers (ECL) of around 100-kHz linewidth to generate the odd-numbered channels, while two other ECLs produce the even-numbered channels. The odd and even channels are modulated by two IQ modulators driven by two independent random 16-QAM sequences with baud rates of 28 GBd or 34 GBd. The roll-off factor is 0.2. The signals from the five channels are combined into the PDM emulator to generate delayed signal copies for polarization multiplexing. The link configuration is the same as single-channel experiments described previously. At the receiver, the center channel is filtered by a wavelength-selective switch (WSS) with 3-dB bandwidth of 40 GHz, which is sampled by 80 GSa/s sampling scope samples followed by offline processing.

The Q factor of the center channel for the five-channel system is shown in Fig. 11a for 28-GBaud transmissions. The DNN-based DBP here is trained on the WDM data, and 262144 bits are used in the testing set to calculate the BER and Q factor. The performance gain of 1-StPS DBP and 50-StPS DBP over CDC only is around 0.3 dB and 0.6 dB, respectively. On the other hand, 1-StPS and 2-StPS DNN-based DBP provides an additional gain of 0.25 dB and 0.45 dB over 50-StPS DBP and a total gain of 0.85 dB and 1 dB over CDC only, respectively. For 34-GBaud transmissions, the gain of 1-StPS DNN-based DBP over 1-StPS DBP and CDC only is around 0.2 dB and 0.4 dB, respectively. 2-StPS DNN-based DBP further improves the optimal Q factor of 1-StPS DNN-based DBP by 0.15 dB. The amplitude spectra and optimized  $\xi_{xx,k}$  shown in Fig. 11c–f display the same “M”-shaped and “U”-shaped features as expected from the analytical insights developed through machine learning. Overall, the results show that DNN-based DBP represents a new design dimension in single-channel DSP algorithm for nonlinearity compensation in WDM systems without compromising computational complexity. It serves as a crucial step forward in improving practical WDM transmission performance.

## Discussion

In this paper, we experimentally demonstrate that by relating the configuration of well-known digital back-propagation algorithm into interleaving linear and nonlinear operators of a deep neural network, machine-learning techniques can optimize the network parameters and lead to dramatic performance improvements and computational savings. Applying DNN-based DBP to PDM-WDM systems reaps sizeable performance improvements compared with other single-channel DSP algorithms, thus achieving a key step in bringing nonlinearity compensation DSP into realistic WDM systems. More importantly, the optimal parameter configurations in turn guided us to analyze the interplay between CD, nonlinearity, and noise more closely and led to deeper theoretical insights that the receiver DSP should not exactly invert the linear and nonlinear steps of the fiber propagation effects. Rather, it should balance between compensating transmission impairments and suppressing the additional distortions arising from imperfect linear and nonlinear compensation steps due to inline amplifier noise. As the new analytical insights are inspired by the optimized neural network model, it shows that machine learning can actually go beyond conventional thinking to develop deeper theoretical understandings in the field of nonlinear fiber transmissions in addition to providing algorithms exceeding state-of-the-art performance. Our work serves as an example that machine-learning techniques can not only provide a detour from intractable systems and arrive at intelligent solutions or strategies, but also help elucidate the path toward deeper physical insights and the underlying mathematical structures.

## Data availability

The data files are available from the corresponding author upon reasonable request. Source data are provided with this paper.

## Code availability

Example codes for the training and testing of the deep neural networks used in the paper are available from the corresponding author upon reasonable request. Source data are provided with this paper.

Received: 21 February 2019; Accepted: 1 July 2020;

Published online: 23 July 2020

## References

1. Winzer, P. J., Neilson, D. T. & Chraplyvy, A. R. Fiber-optic transmission and networking: the previous 20 and the next 20 years. *Opt. Express* **26**, 24190–24239 (2018).
2. Ip, E. & Kahn, J. M. Compensation of dispersion and nonlinear impairments using digital backpropagation. *J. Lightwave Technol.* **26**, 3416–3425 (2008).
3. Du, L. B. & Lowery, A. J. Improved single channel backpropagation for intra-channel fiber nonlinearity compensation in long-haul optical communication systems. *Opt. Express* **18**, 17075–17088 (2010).
4. Rafique, D. et al. Compensation of intra-channel nonlinear fibre impairments using simplified digital back-propagation algorithm. *Opt. Express* **19**, 9453–9460 (2011).
5. Irukulapati, N. V., Wymeersch, H., Johannisson, P. & Agrell, E. Stochastic digital backpropagation. *IEEE Trans. Commun.* **62**, 3956–3968 (2014).
6. Savory, S. J., Gavioli, G., Torrenco, E. & Poggiolini, P. Impact of interchannel nonlinearities on a split-step intra-channel nonlinear equalizer. *IEEE Photonics Technol. Lett.* **22**, 673–675 (2010).
7. Dar, R. & Winzer, P. J. On the limits of digital back-propagation in fully loaded WDM systems. *IEEE Photonics Technol. Lett.* **28**, 1253–1256 (2016).
8. Rivenson, Y. et al. Deep learning microscopy. *Optica* **4**, 1437–1443 (2017).
9. Jo, Y. et al. Holographic deep learning for rapid optical screening of anthrax spores. *Sci. Adv.* **3**, e1700606 (2017).
10. Khan, F. N., Fan, Q., Lu, C. & Lau, A. P. T. An optical communication’s perspective on machine learning and its applications. *J. Lightwave Technol.* **37**, 493–516 (2019).

11. Zibar, D. et al. Nonlinear impairment compensation using expectation maximization for dispersion managed and unmanaged PDM 16-QAM transmission. *Opt. Express* **20**, B181–B196 (2012).
12. Li, M. et al. Nonparameter nonlinear phase noise mitigation by using M-ary support vector machine for coherent optical systems. *IEEE Photon. J.* **5**, 7800312–7800312 (2013).
13. Giacomidis, E. et al. Reduction of nonlinear intersubcarrier intermixing in coherent optical OFDM by a fast newton-based support vector machine nonlinear equalizer. *J. Lightwave Technol.* **35**, 2391–2397 (2017).
14. Kamalov, V. et al. Evolution from 8QAM live traffic to PS 64-QAM with neural-network based nonlinearity compensation on 11000 km open subsea cable. In: (eds Doverspike, R. D., Kuchta, D. & Shieh, W.) *Optical Fiber Communication Conference (OFC)* paper Th4D-5 (Optical Society of America, 2018).
15. Häger, C. & Pfister, H. D. Nonlinear interference mitigation via deep neural networks. In: (eds Doverspike, R. D., Kuchta, D. & Shieh, W.) *Optical Fiber Communications Conference and Exposition (OFC)* 1–3 (Optical Society of America, 2018).
16. Lian, M., Häger, C. & Pfister, H. D. What can machine learning teach us about communications? In: (eds Narayanan, K., Guo, D. & Vontobel, P.) *IEEE Information Theory Workshop (ITW)* 1–5 (IEEE, 2018).
17. Häger, C. & Pfister, H. D. Deep learning of the nonlinear Schrödinger equation in fiber-optic communications. In: (eds Guruswami, V., Jafar, S. A., Montanari, A. & Nair, C.) *IEEE International Symposium on Information Theory (ISIT)* 1590–1594 (IEEE, 2018).
18. Escalante, H. J. et al. Explainable and interpretable models in computer vision and machine learning. In: (eds Escalante, H. J. et al.) *Springer Series on Challenges in Machine Learning* 39–31 (Springer, 2018).
19. Ip, E. Nonlinear compensation using backpropagation for polarization-multiplexed transmission. *J. Lightwave Technol.* **28**, 939–951 (2010).
20. Mateo, E. F., Yaman, F. & Li, G. Efficient compensation of inter-channel nonlinear effects via digital backward propagation in WDM optical transmission. *Opt. Express* **18**, 15144–15154 (2010).
21. Zhu, L. & Li, G. Nonlinearity compensation using dispersion-folded digital backward propagation. *Opt. Express* **20**, 14362–14370 (2012).
22. Tanimura, T. et al. March. Semi-blind nonlinear equalization in coherent multi-span transmission system with inhomogeneous span parameters. In: (eds Hamilton, S. & Jopson, R.) *Optical Fiber Communication Conference (OFC)* paper OMR6 (Optical Society of America, 2010).
23. Fougstedt, C. et al. Time-domain digital back propagation: algorithm and finite-precision implementation aspects. In: (eds Bosco, G., Elbers, J.-P. & Schares, L.) *Optical Fiber Communication Conference (OFC)* paper W1G-4 (Optical Society of America, 2017).
24. Schmogrow, R. et al. Error vector magnitude as a performance measure for advanced modulation formats. *IEEE Photonics Technol. Lett.* **24**, 61–63 (2011).
25. Alpaydin, E. *Introduction to Machine Learning* (MIT press, Cambridge, MA, 2020).
26. Abadi, M. et al. Tensorflow: a system for large-scale machine learning. *OSDI* **16**, 265–283 (2016).
27. Selmi, M., Jaouen, Y. & Ciblat, P. Accurate digital frequency offset estimator for coherent PolMux QAM transmission systems. In: (eds Leeb, W. & Winzer, P.) *35th European Conference on Optical Communication* 1–2 (IEEE, 2009).
28. Pfau, T., Hoffmann, S. & Noé, R. Hardware-efficient coherent digital receiver concept with feedforward carrier recovery for M-QAM constellations. *J. Lightwave Technol.* **27**, 989–999 (2009).
29. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. Preprint at <https://arxiv.org/abs/1412.6980> (2014).
30. Luo, L., Xiong, Y., Liu, Y. & Sun, X. Adaptive gradient methods with dynamic bound of learning rate. Preprint at <https://arxiv.org/abs/1902.09843> (2019).
31. Fatadin, I., Ives, D. & Savory, S. J. Blind equalization and carrier phase recovery in a 16-QAM optical coherent system. *J. Lightwave Technol.* **27**, 3042–3049 (2009).
32. Gao, G., Chen, X. & Shieh, W. Influence of PMD on fiber nonlinearity compensation using digital back propagation. *Opt. Express* **20**, 14406–14418 (2012).

## Acknowledgements

The authors acknowledge the generous support of The Hong Kong Government General Research Fund under project number PolyU 152757/16E.

## Author contributions

Q.F. conducted all the experiments, derived the training, and testing dataset and other signal-processing techniques. Q.F., T.G., and C.L. helped set up the experimental testbed. G.Z. conducted simulations for the ML-inspired analysis. Q.F. and A.P.T.L. derived the ML-inspired analysis.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41467-020-17516-7>.

**Correspondence** and requests for materials should be addressed to Q.F.

**Peer review information** *Nature Communications* thanks Massimo Tornatore and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020