

Compensation of Nonlinear Impairments Using Inverse Perturbation Theory with Reduced Complexity

Alexey Redyuk, Evgeny Averyanov, Oleg Sidelnikov, Mikhail Fedoruk, and Sergei Turitsyn, *Fellow, OSA*

Abstract—We propose a modification of the conventional perturbation-based approach of fiber nonlinearity compensation that enables straight-forward implementation at the receiver and meets feasible complexity requirements. We have developed a model based on perturbation analysis of an inverse Manakov problem, where we use the received signal as the initial condition and solve Manakov equations in the reversed direction, effectively implementing a perturbative digital backward propagation enhanced by machine learning techniques. To determine model coefficients we employ machine learning methods using a training set of transmitted symbols. The proposed approach allowed us to achieve 0.5 dB and 0.2 dB Q^3 -factor improvement for 2000 km transmission of 11x256 Gbit/s DP-16QAM signal compared to chromatic dispersion equalization and one step per span two samples per symbol digital back-propagation technique, respectively. We quantify the trade-off between performance and complexity.

Index Terms—optical communication system, nonlinear signal distortions, Manakov equations, fiber nonlinearity compensation, perturbation-based detection technique, machine learning.

I. INTRODUCTION

NONLINEAR transmission impairments are a fundamental limiting factor for further improving the capacity and reach of modern fiber-optic communication systems. To overcome this problem, a number of methods have been proposed, such as digital back-propagation (DBP) [1], [2], Volterra function-based methods [3], [4], perturbation pre-distortion and post-equalization [5], [6], [7], phase conjugation methods [8], nonlinear interference noise approach [9], machine learning methods [10], [11] (space limitation prevent an overview of all important papers here, for more details, see e.g. [12], [13], [14] and references therein). In particular, digital back-propagation is a powerful tool for nonlinearity compensation that is considered as a benchmark to evaluate performance of other emerging techniques. However, there is always a trade-off between performance improvement and implementation complexity of the nonlinearity mitigation methods.

Practical implementation of nonlinearity compensation methods in real-time processing can potentially be achieved with the help of machine learning (ML) techniques, which have developed rapidly in the field of optical communications

A. Redyuk, E. Averyanov, O. Sidelnikov, M. Fedoruk, and S. Turitsyn are with Novosibirsk State University, Novosibirsk, 630090, Russia.

A. Redyuk, and M. Fedoruk are with Institute of Computational Technologies, Siberian Branch, Russian Academy of Sciences, Novosibirsk, 630090, Russia, e-mail: alexey.redyuk@gmail.com.

S. Turitsyn is with Aston Institute of Photonic Technologies, Aston University, Birmingham, B4 7ET, UK.

in the past few years (see [10] and references therein). ML can potentially enhance system performance while requiring acceptable implementation complexity. In addition, ML-based equalizers can be retrained on demand, which makes them suitable for dynamic reconfigurable transmission environments.

As a master model to describe nonlinear propagation of a dual-polarized signal down an optical fiber, we use a well-known system of Manakov equations [15]:

$$\frac{\partial A_{\mathbf{x}/\mathbf{y}}}{\partial z} = \left[-\frac{\alpha}{2} - i\frac{\beta_2}{2}\frac{\partial^2}{\partial t^2} + i\gamma\frac{8}{9}(|A_{\mathbf{x}}|^2 + |A_{\mathbf{y}}|^2) \right] A_{\mathbf{x}/\mathbf{y}}, \quad (1)$$

where $A_{\mathbf{x}/\mathbf{y}}(z, t)$ is the complex field envelopes for x- and y-polarization, α , β_2 and γ are attenuation, dispersion and Kerr coefficients, respectively. This makes it possible to consider the initial value problem for Manakov equations (1) using launched in the fiber signal as initial condition in the form

$$A_{\mathbf{x}/\mathbf{y}}(z = 0, t) = \sum_k a_{\mathbf{x}/\mathbf{y}}[k]f(t - kT), \quad (2)$$

where $a_{\mathbf{x}/\mathbf{y}}[k]$ are complex transmitted symbols in x- and y-polarization, respectively, k is a number of time slot, T is a symbol interval, and $f(t)$ is a waveform of a carrier pulse.

A first-order perturbation analysis of the problem (1)–(2) yields a known expression for nonlinear three-symbol interactions of different polarizations, which can be written as

$$b_{\mathbf{x}}[k] = a_{\mathbf{x}}[k] + \sum_{m,n} C_{mn} a_{\mathbf{x}}[k+m] a_{\mathbf{x}}[k+n] a_{\mathbf{x}}^*[k+m+n] + \sum_{m,n} C_{mn} a_{\mathbf{x}}[k+m] a_{\mathbf{y}}[k+n] a_{\mathbf{y}}^*[k+m+n], \quad (3)$$

where $b_{\mathbf{x}}[k]$ are received complex symbols in x-polarization, C_{mn} are complex perturbation coefficients, and a^* denotes complex conjugation of a [7], [16]. Nonlinear interactions for y-polarization symbols $b_{\mathbf{y}}[k]$ can be obtained from (3) by replacing x indices with y indices and vice versa. In trivial cases there are analytic expressions to calculate C_{mn} [17]; however, it is generally necessary to calculate thousands of the perturbation coefficients given by multidimensional overlapping integrals [18].

Recent works [7], [19] have proposed and studied receiver-side perturbation-based machine learning equalizer for nonlinearity compensation in SDM and WDM systems, where instead of a time-consuming numerical integration, robust regression methods have been employed to calculate perturbation coefficients C_{mn} . Here, we revise this concept

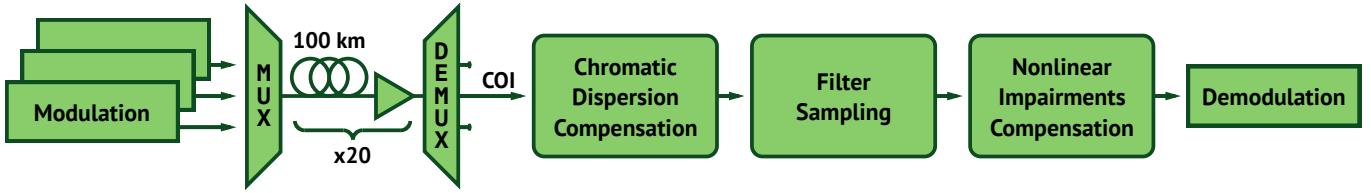


Fig. 1. Principle scheme of transmission link under investigation.

and propose a modified receiver-side perturbation-based post-equalization approach (PPE) based on perturbation analysis of *inverse Manakov problem*, when we consider the output signal as initial condition similar to DBP concept. The key goal of this work is to demonstrate how complexity of the proposed method can be reduced to the level of practical applicability.

II. PERTURBATION-BASED POST EQUALIZATION

A first-order perturbation analysis of the problem (1)–(2) for Manakov equations gives an analytical connection (3) between transmitted and received symbols. Machine learning techniques such as robust regression methods can be applied to the known training set of symbols to identify perturbation coefficients C_{mn} instead of cumbersome analytical or resource-intensive numerical integration. However, to apply model (3) on a receiver side for nonlinearity compensation, it is necessary to know the transmitted symbols $a_{x/y}$ in advance to calculate three-symbol products. An obvious way to approximate these triplets is to replace transmitted symbols by received ones [5], [7], [19], which introduces an additional approximation error in the prediction of the model.

In this paper we propose a simple alternative approach. Main new point of the proposed approach is to present solution of the Manakov equations (or received signal in case of experiment) at the receiver side after compensation of linear effects by the the following expression

$$D^{-1} [A_{x/y}(z=L, t)] \approx \sum_k b_{x/y}[k] f(t - kT), \quad (4)$$

where $D[\cdot]$ is a linear dispersion operator. We would like to stress that $b_{x/y}[k]$ here have meaning of an expansion of the received signal in the basis of $f(t - kT)$, rather then any approximation of the input signal. It then becomes straightforward to perform first-order perturbation analysis of the inverse Manakov problem, treating the signal in the form (4) as an initial condition, inverting the sign of the fiber parameters in (1), and taking reciprocal gain of the inline amplifiers (we discuss in more detail this question in Appendices A and C). Following this approach, one can derive:

$$\begin{aligned} a_x[k] = & C b_x[k] + \sum_{m,n} \tilde{C}_{mn} b_x[k+m] b_x[k+n] b_x^*[k+m+n] \\ & + \sum_{m,n} \tilde{C}_{mn} b_x[k+m] b_y[k+n] b_y^*[k+m+n]. \end{aligned} \quad (5)$$

To find coefficients C and \tilde{C}_{mn} we used linear regression model with a known training set of transmitted and received symbols (for more details of the machine learning methods that are used in the paper, see Appendix B).

At first glance, model (5) only slightly differs from the original model (3). However, after computing coefficients, we observed that perturbation kernels C_{mn} and \tilde{C}_{mn} have a perceptible difference. The main advantage of the model (5) is the possibility of straightforward use of the received symbols $b_{x/y}[k]$ for recovering transmitted symbols $a_{x/y}[k]$ without additional approximations. Furthermore, we use ML methods to find \tilde{C}_{mn} and feed the model data distorted by deterministic nonlinearity and stochastic nonlinear signal-noise interactions during the training process. As a result, obtained kernel \tilde{C}_{mn} reflects both types of distortions. Analytical estimation of C_{mn} , considering signal-noise interactions, can be found in [20].

III. TRANSMISSION SYSTEM MODEL

A range of numerical modeling experiments were performed to evaluate the performance of the proposed method. We considered 11-channel 2000 km WDM transmission of a dual-polarization 16QAM signal. We observed that further increase of the WDM channel number has a negligible impact on the central channel of interest (COI) performance. RRC pulses with a roll-off factor of 0.1 at 32 GBAud and channel spacing of 37.5 GHz have been used as data carriers. A lumped amplification scheme with span length of 100 km and EDFA noise figure of 4.5 dB was studied. At the receiver, after demultiplexing WDM channels, we performed ideal chromatic dispersion compensation (CDC) and matched filter sampling and nonlinear impairments compensation for central channel. Before the demodulation, we consistently applied the linear equalizer, based on least mean square algorithm, to compensate for the constellation phase deviation. We also assume an ideal carrier phase and frequency recovery. The schematics of a transmission link are depicted in Fig. 1 and all parameters are summarized in Table I.

TABLE I
TRANSMISSION MODEL PARAMETERS

Parameter	Value	Parameter	Value
Attenuation	0.2 dB/km	COI Wavelength	1550 nm
Dispersion	17 ps/nm/km	RRC Roll-off	0.1
Nonlinearity	1.4 1/W/km	Symbol Rate	32 GBAud
Distance	20 × 100 km	Channel Spacing	37.5 GHz
Noise Figure	4.5 dB	Channel Data Rate*	256 Gbit/s

*including FEC overhead

Numerical simulation was performed using a standard second-order symmetrical split-step Fourier method with oversampling factor of 64 and logarithmic step size selection [21].

IV. RESULTS AND COMPLEXITY REDUCTION

The performance of the proposed scheme was evaluated by comparing it with: (i) a previously studied perturbation-based model (Forward PPE) [19], based on the scheme (3); (ii) a linear phase shift equalizer (CDC), which recovers only the phase of the received signal; and (iii) the single-carrier digital back-propagation algorithm with a different number of steps per span and oversampling factor of 2. Figure 2 shows the BER of a central channel as a function of a launch power per channel for different nonlinear impairments compensation schemes without any restrictions on computational complexity.

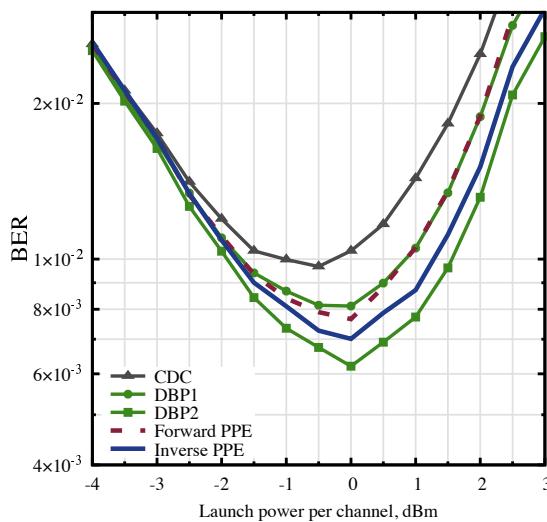


Fig. 2. Performance comparison of the proposed Inverse PPE scheme and known nonlinearity compensation methods.

It can be noted that the proposed method provides performance between DBP1 (one step per span, two samples per symbol) and DBP2 (two steps per span, two samples per symbol) levels, exceeding at the optimal power the performance of linear equalizer (CDC) and forward approach (Forward PPE) by 33% and 8% in terms of BER (0.5 dB and 0.1 dB in terms of Q^2 -factor), respectively.

As mentioned above, for various nonlinearity compensation methods, implementation complexity is an important practical issue, along with performance improvement. Therefore, we next estimated the total number of complex multiplications per recovered symbol required for the application of the proposed scheme with learned perturbation coefficients \tilde{C}_{mn} .

A. Basic Complexity Estimation

Number of perturbation terms in (5) with summation constraint $|m| + |n| \leq M$ is $n_1 = (M+1)^2 + M^2$, where M is a channel memory parameter corresponding to the number of used neighboring symbols at each side of the symbol under consideration. The total number of required complex multiplications is given by $C_{\text{base}} = 3n_1 + 1$, showing quadratic growth rate $O(M^2)$ that, from a practical point of view, can make implementation non-acceptable.

Thus, the application of the proposed method in a straightforward manner is not sufficiently effective in terms of the implementation complexity. In order to bring it closer to practical applicability, we reduce the computational complexity by utilizing a combination of techniques capable decreasing the number of required complex multiplications per symbol.

B. Perturbation Coefficients Selection

We start by reducing the number of terms in the right side of the model (5). It was already observed that, in the perturbation coefficients matrix in Fig. 3, coefficients with the largest values are concentrated under hyperbolic curves $|mn| < M$ [16], [22]. The total number of these coefficients can be estimated as $n_2 = 4 \sum_{k=1}^M \lfloor \frac{M-1}{k} \rfloor + 4M + 1$. Then, performing such selection by importance and dropping negligible coefficients, new complexity estimation is given by $C_{\text{selection}} = 3n_2 + 1$, scaling as $O(M \log M)$.

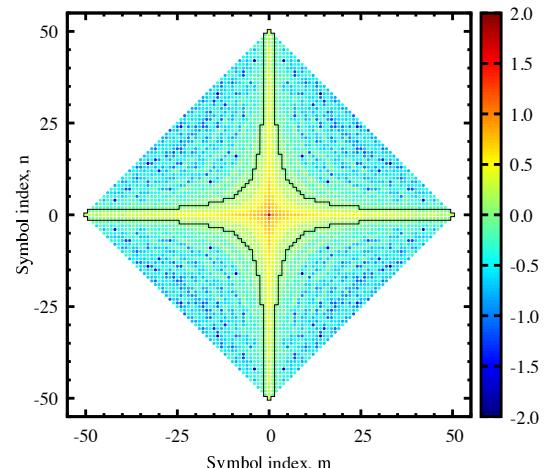


Fig. 3. Learned perturbation matrix $\log_{10} |\tilde{C}_{mn}|$, $|m| + |n| \leq 50$. The black internal line highlights the area where coefficients satisfy $|mn| < 50$.

C. Quantization of Perturbation Coefficients

Further complexity reduction can be achieved with vector quantization of coefficients \tilde{C}_{mn} . We combine multiple terms with similar perturbation coefficients in the model (5) and approximate these coefficients with a certain complex number called centroid. It has been shown that quantization of \tilde{C}_{mn} can effectively reduce the number of multiplications without any significant loss of performance [22], [23]. For this purpose, we applied the k -means clustering algorithm. All the perturbation coefficients are divided into a number of clusters with corresponding centroids so as to minimize the inertia in Euclidean norm (for more details see Appendix D).

Figure 4 shows the perturbation coefficients and clusters centroids. Four major clusters stand out clearly here: left wing with negative real part, right wing with positive real part, center stem with mostly negative imaginary part, and coefficient $C_{0,0}$ is a singleton cluster that has the largest magnitude and did not fit in the figure. This observation provides a good heuristic estimate for the optimal number of quantization levels

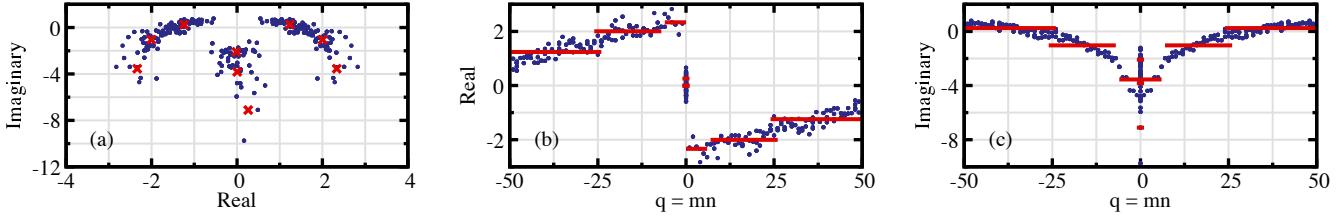


Fig. 4. (a) Learned perturbation coefficients \tilde{C}_{mn} (blue dots) along with quantization centroids (red crosses) in a complex plane with $M = 50$ and $l = 10$, (b)/(c) real/imaginary parts of perturbation coefficients \tilde{C}_{mn} and quantization levels as function of $q = mn$.

to be $l = 3s + 1$, $s \in \mathbb{N}$, as we assign one centroid to $C_{0,0}$ and divide the rest equally between wings and stem. In our numerical experiments $s \approx \lfloor \frac{M+14}{16} \rfloor$ was enough to greatly reduce number of coefficients with less than 1% performance loss. It should be noted that the centroids of the clusters fully cover all the perturbation coefficients and the symmetry properties are also preserved after quantization procedure. The resulting computational complexity yields $C_{\text{vq}} = 2n_2 + l + 1$.

D. Allocation of Cyclic Buffer

If we consider one perturbation term for fixed m and n from sum in (5) and factor out $\tilde{C}_{mn}b_x[k+m]$ we will get

$$b_x[k+n]b_x^*[k+m+n] + b_y[k+n]b_y^*[k+m+n]. \quad (6)$$

Let us closely consider the set of indices $I_k = \{(k+n, k+m+n) : |m| + |n| \leq M, |mn| < M\}$. It can be seen that successive set I_{k+1} has many common elements with I_k . Assume that DSP unit has internal memory, that are able to store intermediate results in cyclic buffer. In this case only $|I_{k+1} \setminus I_k| = 2M + 1$ new multiplications are needed to compute the entire set of (6) with indices from I_{k+1} . This gives us the final complexity estimate $C_{\text{buffer}} = n_2 + l + 2M + 2$.

Applying the above-described techniques significantly reduces a number of coefficients \tilde{C}_{mn} and a required number of multiplication per symbol. For $M = 70$, we reduce the computational complexity from 30,000 to 1750 multiplications, which is more than one order of magnitude. Figure 5 shows the dependence of the implementation complexity on a memory parameter M for different complexity estimation scenarios. It should be noted that, by reducing the computational complexity, we degrade the accuracy of the model. However, we verified that the accuracy can be restored by a slight increase of the parameter M without a significant increase in complexity.

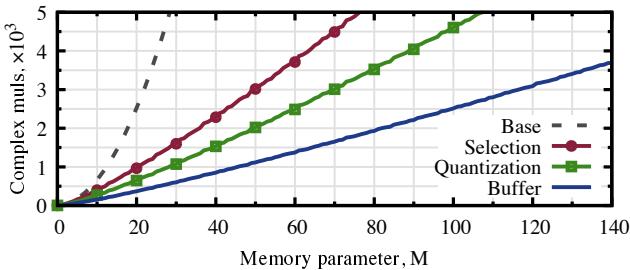


Fig. 5. Complexity vs. memory parameter M for different complexity estimation scenarios.

Figure 6 shows the achieved Q^2 -factor improvement in comparison with the linear equalizer CDC for the system under consideration in terms of computational complexity. It can be seen that complexity can be reduced by up to 1750 complex multiplications per symbol without any significant degradation of performance. It should also be noted that 0.3 dB Q^2 -factor improvement compared with linear equalization can be achieved using less than 600 complex multiplications.

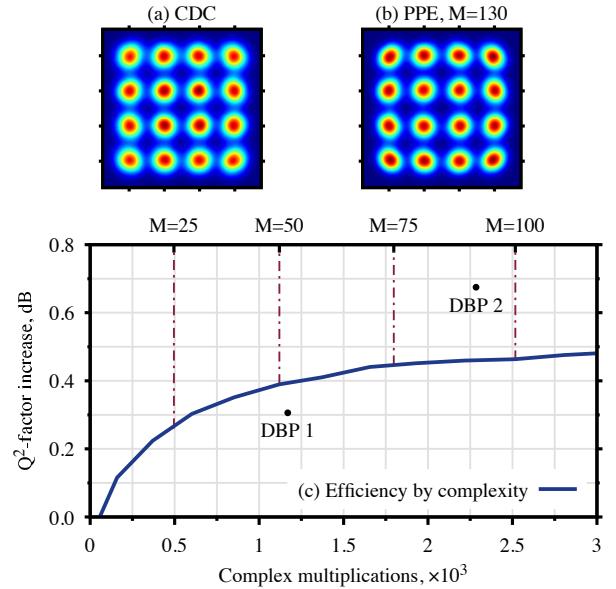


Fig. 6. (a) Received constellation diagram, (b) constellation diagram after inverse PPE with $M = 130$, (c) Q^2 -factor increase with the given limit of complex multiplications.

To compare the method's complexity, we added DBP1 and DBP2 complexity estimation to Fig. 6c, in terms of the total number of complex multiplications per recovered symbol, using the expression obtained by A. Napoli et al. in [2]

$$C_{\text{DBP}} = N_{\text{steps}}(C_{\text{FDE}} + n) + C_{\text{FDE}},$$

$$C_{\text{FDE}} = \frac{N(\log_2 N + 1)}{N - N_D + 1} n, \quad (6)$$

where N is the size of FFT-block, N_D is the number of samples used to account for boundary dispersion effects, n is the oversampling factor, and N_{steps} is the total number of computational steps. To estimate DBP complexity we use following parameters: $N = 1024$, $N_D = 606$, $n = 2$ and

$N_{\text{steps}} = 20$ and 40 for DBP1 and DBP2, respectively, which leads to 1170 and 2285 complex multiplications, required by DBP1 and DBP2 algorithms.

Related to hardware implementation, it should be noted that the forward approach (3) is probably more robust against a precision loss of the received sequence $b[k]$ than the proposed backward approach (5). Potentially, it can reduce performance of (5) in case of the low fixed-point resolution of $b[k]$. However, this point requires a separate study.

V. CONCLUSIONS

In this work we revisited conventional perturbation-based concept for fiber nonlinearity compensation and proposed modified receiver-side approach with low implementation complexity derived from an inverse Manakov problem. By reformulating the underlying model and applying the machine learning techniques to determine model coefficients, the proposed method more effectively deals with fiber nonlinearity consisting of deterministic nonlinear impairments and stochastic nonlinear signal-noise interactions. We compared the performance of the proposed method with a linear phase equalizer that recovers the phase of the received signal and the digital back propagation. In case of multi-channel transmission of 11×256 Gbit/s DP-16QAM signal, we demonstrated 0.5 dB and 0.2 dB Q^2 -factor improvement, compared with linear equalization and one step per span and two samples per symbol one-channel DBP, respectively. We showed that implementation complexity can be significantly reduced by up to 1750 complex multiplications by symbol without significant degradation of performance.

APPENDIX A

DETAILS OF THE PERTURBATION MODEL

A. Basic theory

The perturbation theory has been discussed in many publications and is well-known. However, for the benefit of readers who are not working in the field, this Appendix reviews standard perturbation theory for the nonlinear Shrödinger equation (NLSE). Generalization to the Manakov equations is straightforward. The following notation for forward and inverse Fourier transforms will be used below:

$$\hat{f}(\omega) = \int f(t) e^{i\omega t} dt, \quad f(t) = \int \frac{d\omega}{2\pi} \hat{f}(\omega) e^{-i\omega t}.$$

The NLSE for a complex envelope $A(z, t)$ reads:

$$\partial_z A = -\frac{\alpha}{2} A - i \frac{\beta_2}{2} \partial_t^2 A + i\gamma |A|^2 A. \quad (7)$$

We can rewrite (7) using the following well-known substitution: $A(z, t) = \sqrt{P_0} \exp(-\frac{\alpha z}{2}) U(z, t)$:

$$\partial_z U = -i \frac{\beta_2}{2} \partial_t^2 U + i\gamma P_0 e^{-\alpha z} |U|^2 U. \quad (8)$$

Applying Fourier transform to (8), after straightforward manipulations using inverse transformation of \hat{U} for U in the nonlinear term, we arrive at the expression:

$$\begin{aligned} \partial_z \hat{U} &= i \frac{\beta_2 \omega^2}{2} \hat{U} + i\gamma P_0 e^{-\alpha z} \iiint \frac{d\omega_1}{2\pi} \frac{d\omega_2}{2\pi} \frac{d\omega_3}{2\pi} dt \times \\ &\quad \times \hat{U}_{\omega_1} \hat{U}_{\omega_2} \hat{U}_{\omega_3}^* e^{-it(\omega_1+\omega_2-\omega_3-\omega)}. \end{aligned} \quad (9)$$

To save space, we shall write \hat{U}_ω in place of $\hat{U}(z, \omega)$. Next we introduce the change of variables to reduce the number of integrations:

$$\begin{cases} \omega_1 = \nu_1 + \nu_3, \\ \omega_2 = \nu_2 + \nu_3, \\ \omega_3 = \nu_1 + \nu_2 + \nu_3, \end{cases} \quad \left| \det \left(\frac{\partial \omega}{\partial \nu} \right) \right| = 1. \quad (10)$$

Using substitution (10) and a definition of the Dirac delta function, we can rewrite nonlinear term of (9):

$$\begin{aligned} &\iiint \frac{d\omega_1}{2\pi} \frac{d\omega_2}{2\pi} \frac{d\omega_3}{2\pi} dt \hat{U}_{\omega_1} \hat{U}_{\omega_2} \hat{U}_{\omega_3}^* e^{-it(\omega_1+\omega_2-\omega_3-\omega)} \\ &= \iint \frac{d\nu_1}{2\pi} \frac{d\nu_2}{2\pi} \hat{U}_{\nu_1+\omega} \hat{U}_{\nu_2+\omega} \hat{U}_{\nu_1+\nu_2+\omega}^*. \end{aligned}$$

Thus, we have the following standard form of (9) with symbols ν reassigned back to ω :

$$\begin{aligned} \partial_z \hat{U} &= i \frac{\beta_2 \omega^2}{2} \hat{U} + i\gamma P_0 e^{-\alpha z} \iint \frac{d\omega_1}{2\pi} \frac{d\omega_2}{2\pi} \times \\ &\quad \times \hat{U}_{\omega_1+\omega} \hat{U}_{\omega_2+\omega} \hat{U}_{\omega_1+\omega_2+\omega}^*. \end{aligned} \quad (11)$$

Transform $\hat{U}(z, \omega) = \exp(i \frac{\beta_2 \omega^2}{2} z) \hat{u}(z, \omega)$ yields:

$$\begin{aligned} \partial_z \hat{u} &= i\gamma P_0 e^{-\alpha z} \iint \frac{d\omega_1}{2\pi} \frac{d\omega_2}{2\pi} \times \\ &\quad \times \hat{u}_{\omega_1+\omega} \hat{u}_{\omega_2+\omega} \hat{u}_{\omega_1+\omega_2+\omega}^* e^{-i\beta_2 z \omega_1 \omega_2}. \end{aligned} \quad (12)$$

Integrating (12) by z from 0 to L , we can obtain the following:

$$\begin{aligned} \hat{u}(L, \omega) - \hat{u}(0, \omega) &= i\gamma P_0 \int_0^L dz e^{-\alpha z} \iint \frac{d\omega_1}{2\pi} \frac{d\omega_2}{2\pi} \times \\ &\quad \times \hat{u}_{\omega_1+\omega} \hat{u}_{\omega_2+\omega} \hat{u}_{\omega_1+\omega_2+\omega}^* e^{-i\beta_2 z \omega_1 \omega_2}. \end{aligned} \quad (13)$$

Next we assume that initial conditions are

$$\begin{aligned} A(0, t) &= \sqrt{P_0} \sum_k a[k] f(t - kT), \\ \hat{A}(0, \omega) &= \sqrt{P_0} \hat{f}(\omega) \sum_k a[k] e^{i\omega kT}, \end{aligned}$$

where signal pulse shape $f(t)$ satisfies Nyquist intersymbol interference criterion. Therefore, provided the convolution theorem is used we can perform filtered sampling by multiplying equation by matched filter $\hat{f}^*(\omega)$ in the frequency domain and taking the inverse Fourier transformation:

$$\int \frac{d\omega}{2\pi} \hat{f}^*(\omega) \hat{u}(0, \omega) e^{-i\omega t} = \sum_k a[k] F(t - kT),$$

where $F(t)$ is an auto-correlation function of $f(t)$. Doing something similarly to the receiver side function $u(L, t)$ may be problematic because there is no way to represent it in such a way as the transmitter side function $u(0, t)$. Therefore, we define received complex amplitudes $b[k]$ as follows:

$$b[k] = G(kT), \quad G(t) = \int \frac{d\omega}{2\pi} \hat{f}^*(\omega) \hat{u}(L, \omega) e^{-i\omega t}, \quad k \in \mathbb{Z}. \quad (14)$$

Let us rewrite (13) after filter sampling:

$$\begin{aligned} G(t) - \sum_k a[k] F(t - kT) &= i\gamma P_0 \int_0^L dz e^{-\alpha z} \times \\ &\quad \times \iint \frac{d\omega}{2\pi} \frac{d\omega_1}{2\pi} \frac{d\omega_2}{2\pi} \hat{f}^* \hat{u}_{\omega_1+\omega} \hat{u}_{\omega_2+\omega} \hat{u}_{\omega_1+\omega_2+\omega}^* e^{-i\beta_2 z \omega_1 \omega_2 - i\omega t}. \end{aligned}$$

After substituting $t = kT$ we derive a discrete model:

$$b[k] = a[k] + i\gamma P_0 \int_0^L dz e^{-\alpha z} \iiint \frac{d\omega}{2\pi} \frac{d\omega_1}{2\pi} \frac{d\omega_2}{2\pi} \times (15) \\ \times \hat{f}_\omega^* \hat{u}_{\omega_1+\omega} \hat{u}_{\omega_2+\omega} \hat{u}_{\omega_1+\omega_2+\omega}^* e^{-i\beta_2 z \omega_1 \omega_2 - i\omega kT}.$$

To apply perturbation theory and obtain first-order approximation, we shall use $\hat{u}(0, \omega)$ instead of $\hat{u}(z, \omega)$ in the right-hand side of the (15), because $\hat{u}(0, \omega)$ is the solution of (12) in case of $\gamma = 0$:

$$\hat{f}_\omega^* \hat{u}_{\omega_1+\omega} \hat{u}_{\omega_2+\omega} \hat{u}_{\omega_1+\omega_2+\omega}^* e^{-i\beta_2 \omega_1 \omega_2 z - i\omega kT} \approx \\ \approx \sum_{p,q,r} a[p] a[q] a^*[r] \hat{f}_\omega^* \hat{f}_{\omega_1+\omega} \hat{f}_{\omega_2+\omega} \hat{f}_{\omega_1+\omega_2+\omega}^* \times \\ \times e^{-i\beta_2 \omega_1 \omega_2 z + i\omega_1(p-r)T + i\omega_2(q-r)T + i\omega(p+q-r-k)T} =$$

using the indices shift $p = k+n$, $q = k+m$, $r = k+h$ we get

$$= \sum_{m,h,n} a[k+m] a[k+n] a^*[k+h] \hat{f}_\omega^* \hat{f}_{\omega_1+\omega} \hat{f}_{\omega_2+\omega} \hat{f}_{\omega_1+\omega_2+\omega}^* \times \\ \times e^{-i\beta_2 \omega_1 \omega_2 z + i\omega_1(n-h)T + i\omega_2(m-h)T + i\omega(m+n-h)T}.$$

Finally, we can write

$$b[k] = a[k] + \sum_{m,h,n} C_{m,h,n} a[k+m] a[k+n] a^*[k+h], \quad (16)$$

where

$$C_{m,h,n} = i\gamma P_0 \int_0^L dz e^{-\alpha z} \iiint \frac{d\omega}{2\pi} \frac{d\omega_1}{2\pi} \frac{d\omega_2}{2\pi} \hat{f}_\omega^* \hat{f}_{\omega_1+\omega} \hat{f}_{\omega_2+\omega} \times \\ \times \hat{f}_{\omega_1+\omega_2+\omega}^* e^{-i\beta_2 \omega_1 \omega_2 z + i\omega_1(n-h)T + i\omega_2(m-h)T + i\omega(m+n-h)T}.$$

In order to reduce complexity of the model, it might be useful to consider the “pulse-matching condition,” which implies $h = m + n$. This simplification can be explained by the fact that almost all the “energy” of the three-dimensional tensor $C_{m,h,n}$ is concentrated at a single matrix slice $C_{m,m+n,n}$. As a part of the study, we verified that the performance efficiency of the proposed and the original methods is provided mainly by the slice $C_{m,m+n,n}$. However, the problem of usage of coefficients from other slices requires a separate investigation. Thereby, we have

$$b[k] = a[k] + \sum_{m,n} C_{m,n} a[k+m] a[k+n] a^*[k+m+n], \quad (17)$$

where

$$C_{m,n} = i\gamma P_0 \int_0^L dz e^{-\alpha z} \iiint \frac{d\omega}{2\pi} \frac{d\omega_1}{2\pi} \frac{d\omega_2}{2\pi} \times \\ \times \hat{f}_\omega^* \hat{f}_{\omega_1+\omega} \hat{f}_{\omega_2+\omega} \hat{f}_{\omega_1+\omega_2+\omega}^* e^{-i\beta_2 \omega_1 \omega_2 z - i\omega_1 m T - i\omega_2 n T}.$$

There is a trivial extension of the obtained model (17) for the case of a multispan transmission link with inline amplifiers and dual-polarized signal resulting in model (3).

Considering that the received signal $A_{x/y}(L, t)$ after compensation of linear effects can be presented by the form (4) it is possible to perform first-order perturbation analysis of the inverse Manakov problem, treating the received signal in the

form (4) as the initial conditions, inverting the sign of the fiber parameters in Manakov equations (1), and taking the reciprocal gain of the inline amplifiers. Following this technique, one can obtain model (5) and modified analytic expressions of the perturbation coefficients \tilde{C}_{mn} .

B. Machine Learning Methods

The perturbation model equation (5) has a form of weighted sum and the linear regression model can be applied to obtain the perturbation coefficients \tilde{C}_{mn} . Using known training sequences of transmitted $a[k]$ and received $b[k]$ symbols, one can construct system $\mathbf{X}\beta = y$, where

$$\mathbf{X} = \begin{bmatrix} \vdots & & \vdots \\ b[k] & \dots & b[k+m] b[k+n] b^*[k+m+n] & \dots \\ \vdots & & \vdots & \end{bmatrix}, \quad (18)$$

$$\beta = [C \dots \tilde{C}_{mn} \dots]^\top, \quad y = [\dots a[k] \dots]^\top. \quad (19)$$

This system can be solved as an ordinary least squares (OLS) problem with a mean squared error (MSE) objective function:

$$\text{MSE}(\beta) = \frac{1}{N} \|\mathbf{X}\beta - y\|_2^2 = \frac{1}{N} \sum_{k=1}^N |X_k \beta - y_k|^2. \quad (20)$$

One way to solve the OLS problem is to use the first Gauss transformation $\mathbf{X}^\dagger \mathbf{X} \beta = \mathbf{X}^\dagger y$, where \mathbf{X}^\dagger is the Hermitian conjugation of \mathbf{X} , and singular value decomposition to invert matrix $\mathbf{X}^\dagger \mathbf{X}$. The solution is then written as

$$\beta = (\mathbf{X}^\dagger \mathbf{X})^{-1} \mathbf{X}^\dagger y. \quad (21)$$

It is worth noting that, in order to get correct evaluation metrics, one should use another known sequence of data that was not involved in the training process. In this paper, sample size for training and testing sets was 2^{16} for each polarization. This size was chosen based on two requirements: to provide enough samples for the convergence of the optimization method, and to obtain accurate estimation of the BER.

C. Approximation of the received signal

Developed approach is based on presenting the received signal (after compensation of linear distortions) in the form:

$$D^{-1}[A_{x/y}(z=L, t)] \approx \sum_k b_{x/y}[k] f(t - kT). \quad (22)$$

Similar expansion is satisfied by default at the transmitter side, because we generate it in this form. We would like to stress that this presentation should not be considered as a recovery of the transmitted data, but rather as a sampling and representation of the received signal in a specific basis. In case of the complete orthogonal basis formed by $f(t - kT)$, any practical bandwidth limited function can be re-presented in this way. However, the same approximate approach can be also used when a carrier pulse $f(t)$ is well localised within a symbol slot and its overlap with pulses from other time slots is minimal. To verify applicability of the proposed approach, we measured a level of inaccuracy by comparison with direct numerical modeling. We performed numerical

experiments with ideal DBP (the same number of steps as for forward propagation and the same oversampling factor) in two scenarios. The first is using the received signal as is for initial data. The second is using a digitally regenerated signal by (22), adding the required level of chromatic dispersion. After DBP we compared symbols $a_{x/y}[k]$ with recovered symbols $\bar{a}_{x/y}[k]$.

Figure 7 shows that, in terms of error vector magnitude (EVM), there is no significant difference between ideal DBP with original signal and regenerated one. Basing on this, we conclude that it is reasonable to perform perturbation analysis of the inverse Manakov problem, treating regeneration of the received signal as the initial condition and inverting the sign of the fiber parameters in Manakov equations (1).

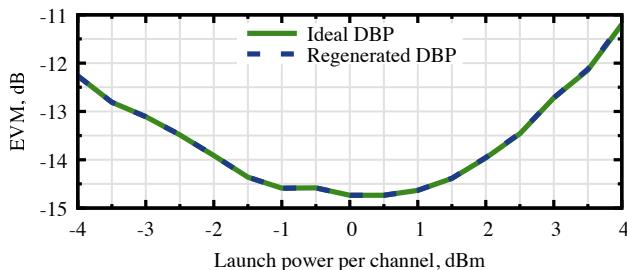


Fig. 7. Comparison of ideal one-channel DBP with raw received and digitally regenerated signal.

D. Used Metrics

$$\text{Inertia}(\boldsymbol{\mu}; \mathbf{X}) = \frac{1}{N} \sum_{i=1}^N \min_{k=1, \dots, K} |X_i - \mu_k|^2,$$

where \mathbf{X} is the data matrix, each line of which corresponds to one point, and $\boldsymbol{\mu}$ is the cluster centroid matrix.

$$\text{EVM [dB]} = 10 \log_{10} \frac{\frac{1}{N} \sum_{k=1}^N |a[k] - b[k]|^2}{\frac{1}{M} \sum_{k=1}^M |q_k|^2},$$

where $a[k]$, $b[k]$ are transmitted and received symbols, and q_k is the ideal constellation points of chosen M -ary modulation.

$$Q^2 [\text{dB}] = 20 \log_{10} \left[\sqrt{10} \operatorname{erfc}^{-1} \left(\frac{8\text{BER}}{3} \right) \right].$$

ACKNOWLEDGMENT

The work was supported by the Russian Science Foundation (Grant No. 17-72-30006). The work of SKT was supported by the EPSRC Programme grant TRANSNET (EP/R035342/1).

REFERENCES

- [1] E. Ip, "Nonlinear compensation using backpropagation for polarization-multiplexed transmission," *Journal of Lightwave Technology*, vol. 28, no. 6, pp. 939–951, Mar. 2010.
- [2] A. Napoli, Z. Maalej, V. A. J. M. Sleiffer, M. Kuschnerov, D. Rafique, E. Timmers, B. Spinnler, T. Rahman, L. D. Coelho, and N. Hanik, "Reduced complexity digital back-propagation methods for optical communication systems," *Journal of Lightwave Technology*, vol. 32, no. 7, pp. 1351–1362, Apr. 2014.
- [3] F. P. Guiomar and A. N. Pinto, "Simplified volterra series nonlinear equalizer for polarization-multiplexed coherent optical systems," *Journal of Lightwave Technology*, vol. 31, no. 23, pp. 3879–3891, Dec 2013.
- [4] L. Liu, L. Li, Y. Huang, K. Cui, Q. Xiong, F. N. Hauske, C. Xie, and Y. Cai, "Intrachannel nonlinearity compensation by inverse volterra series transfer function," *Journal of Lightwave Technology*, vol. 30, no. 3, pp. 310–316, Feb. 2012.
- [5] T. Oyama, H. Nakashima, S. Oda, T. Yamauchi, Z. Tao, T. Hoshida, and J. C. Rasmussen, "Robust and efficient receiver-side compensation method for intra-channel nonlinear effects," in *OFC 2014*, Mar. 2014, paper Tu3A.3.
- [6] Y. Gao, J. Cartledge, A. Karar, S. S.-H. Yam, M. O'Sullivan, C. Laperle, A. Borowiec, and K. Roberts, "Reducing the complexity of perturbation based nonlinearity pre-compensation using symmetric edc and pulse shaping," *Opt. Express*, vol. 22, no. 2, pp. 1209–1219, Jan 2014.
- [7] M. Sorokina, S. Sygletos, and S. Turitsyn, "Sparse identification for nonlinear optical communication systems: Sino method," *Opt. Express*, vol. 24, no. 26, pp. 30433–30443, Dec. 2016.
- [8] A. D. Ellis, M. E. McCarthy, M. A. Z. Al-Khateeb, and S. Sygletos, "Capacity limits of systems employing multiple optical phase conjugators," *Opt. Express*, vol. 23, no. 16, pp. 20381–20393, Aug 2015.
- [9] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "Inter-channel nonlinear interference noise in WDM systems: Modeling and mitigation," *Journal of Lightwave Technology*, vol. 33, no. 5, pp. 1044–1053, Mar. 2015.
- [10] D. Zibar, M. Piels, R. Jones, and C. G. Schaeffer, "Machine learning techniques in optical communication," *Journal of Lightwave Technology*, vol. 34, no. 6, pp. 1442–1452, Mar. 2016.
- [11] O. S. Sidelnikov, A. A. Redyuk, and S. Sygletos, "Dynamic neural network-based methods for compensation of nonlinear effects in multi-mode communication lines," *Quantum Electronics*, vol. 47, no. 12, pp. 1147–1149, Dec. 2017.
- [12] A. Amari, O. A. Dobre, R. Venkatesan, O. S. S. Kumar, P. Ciblat, and Y. Jaouen, "A survey on fiber nonlinearity compensation for 400 gb/s and beyond optical communication systems," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 4, pp. 3097–3113, Jun. 2017.
- [13] D. Rafique, "Fiber nonlinearity compensation: commercial applications and complexity analysis," *Journal of Lightwave Technology*, vol. 34, no. 2, p. 544–552, 2016.
- [14] J. C. Cartledge, F. P. Guiomar, F. R. Kschischang, G. Liga, and M. P. Yankov, "Digital signal processing for fiber nonlinearities," *Opt. Express*, vol. 25, no. 3, pp. 1916–1936, Feb 2017.
- [15] D. Marcuse, C. Manyuk, and P. Wai, "Application of the manakov-pmd equation to studies of signal propagation in optical fibers with randomly varying birefringence," *Journal of Lightwave Technology*, vol. 15, no. 9, pp. 1735–1746, Sep. 1997.
- [16] Z. Tao, L. Dou, W. Yan, L. Li, T. Hoshida, and J. C. Rasmussen, "Multiplier-free intrachannel nonlinearity compensating algorithm operating at symbol rate," *Journal of Lightwave Technology*, vol. 29, no. 17, pp. 2570–2576, Sep. 2011.
- [17] A. Ghazisaeidi and R. Essiambre, "Calculation of coefficients of perturbative nonlinear pre-compensation for nyquist pulses," in *2014 The European Conference on Optical Communication (ECOC)*, Sep. 2014.
- [18] A. Mecozzi, "A unified theory of intrachannel nonlinearity in pseudolinear transmission," in *Impact of Nonlinearities on Fiber Optic Communications*. Springer New York, 2011, pp. 253–291.
- [19] E. Averyanov, A. Redyuk, O. Sidelnikov, M. Sorokina, M. Fedoruk, and S. Turitsyn, "Perturbative machine learning technique for nonlinear impairments compensation in WDM systems," in *2018 European Conference on Optical Communication (ECOC)*, Sep. 2018, paper Th2.39.
- [20] A. Ghazisaeidi, "A theory of nonlinear interactions between signal and amplified spontaneous emission noise in coherent wavelength division multiplexed systems," *Journal of Lightwave Technology*, vol. 15, no. 23, pp. 5150–5175, Dec. 2017.
- [21] G. Bosco, A. Carena, V. Curri, R. Gaudino, P. Poggiolini, and S. Benedetto, "Suppression of spurious tones induced by the split-step method in fiber systems simulation," *IEEE Photonics Technology Letters*, vol. 12, no. 5, pp. 489–491, May 2000.
- [22] M. Malekiha, I. Tseleniker, and D. V. Plant, "Efficient nonlinear equalizer for intra-channel nonlinearity compensation for next generation agile and dynamically reconfigurable optical networks," *Opt. Express*, vol. 24, no. 4, pp. 4097–4108, Feb. 2016.
- [23] Z. Li, W. Peng, F. Zhu, and Y. Bai, "MMSE-based optimization of perturbation coefficients quantization for fiber nonlinearity mitigation," *Journal of Lightwave Technology*, vol. 33, no. 20, pp. 4311–4317, Oct. 2015.