Inverse Regular Perturbation with ML-Assisted Phasor Correction for Fiber Nonlinearity Compensation

HUBERT DZIECIOL[1†], TOSHIAKI KOIKE-AKINO[2*], YE WANG[2], AND KIERAN PARSONS[2]

[1] Optical Networks Group, Department of Electronic and Electrical Engineering, University College London, London WC1E 7JE, UK
[2] Mitsubishi Electric Research Laboratories (MERL), 201 Broadway, Cambridge, MA, 02139, USA
[*] Corresponding author: koike@merl.com; [†] H. Dzieciol conducted this study while he was an intern at MERL.

Compiled June 14, 2022

We improve an inverse regular perturbation (RP) model using a machine learning technique. The proposed learned RP (LRP) model jointly optimizes step-size, gain and phase rotation for individual RP branches. We demonstrate that the proposed LRP can outperform the corresponding learned digital back-propagation (DBP) method based on a split-step Fourier method (SSFM), with up to 0.75 dB gain in a 800km standard single mode fiber link. Our LRP also allows a fractional step-per-span (SPS) modelling to reduce complexity while maintaining superior performance over a 1-SPS SSFM-DBP.

http://dx.doi.org/10.1364/ao.XX.XXXXXX

1. INTRODUCTION

Over the past few decades, the rapid progress of digital signal processing (DSP) technologies has enabled significantly increased throughput and reliability of coherent optical fiber communications systems. For example, linear impairments, such as chromatic dispersion (CD), polarization mode dispersion and phase noise, have been extensively studied and addressed in the digital domain [1–3]. Accordingly, Kerr fiber nonlinearity has become a dominant source that limits the maximum achievable information rates and effective signal-to-noise ratio (SNR) of the optical communications systems [3, 4].

The Manakov equation has been widely used for propagation model of a dual-polarization (DP) signal through a standard single mode fiber (SSMF). It can be numerically solved with the split-step Fourier method (SSFM), which computes the nonlinear and linear operations sequentially in small steps [4]. Alternatively, a number of models, including regular perturbation (RP), have also been studied [5–7]. Digital back propagation (DBP) is a state-of-the-art method for nonlinearity compensation (NLC) [8]. Although it offers considerably improved performance over most other methods, this comes at the cost of higher computational complexity. The main challenge to realize a practical NLC is hence to find the best balance between the complexity and performance. Recent advancements in machine learning (ML) and deep neural networks (DNNs) have revolutionized a wide variety of the scientific fields including optical fiber communications and networking [9]. Within the context of NLC, several physics-informed DNNs have been proposed for DBP [10–13]. Furthermore, a study in [14] used perturbation theory to enhance the learned DBP, whereas [15–19] used ML to improve nonlinear equalization based on the symbol triplet interactions.

In this paper, we propose an ML-aided NLC approach based on an inverted first-order RP model [6]. We show that its parallel branch structure [20] can be adjusted with a low-complexity ML technique to significantly increase the accuracy. We demonstrate that our learned RP (LRP) model can outperform conventional SSFM-DBP by up to 0.75 dB. Moreover, for LRP with fractional step-per-span (SPS) cases (i.e., where each RP branch models more than one span), we show gains of up to 0.2 dB over a 1-SPS SSFM-DBP, while decreasing the computational complexity.

2. BACKGROUND THEORY

We first discuss the underlying theory behind nonlinear signal propagation in the optical fiber with the Manakov model and RP models. We then introduce our LRP as an NLC technique.

A. Manakov Model

Signal propagation of DP complex field $\mathbf{u}(z,t)$ through SSMF with lumped optical amplification (OA) may be expressed as

$$\frac{\partial}{\partial z}\mathbf{u} = -i\frac{\beta_2}{2}\frac{\partial^2}{\partial t^2}\mathbf{u} + i\gamma\frac{8}{9}f(z)\|\mathbf{u}\|^2\mathbf{u} + \frac{1}{\sqrt{f(z)}}\mathbf{n}, \qquad (1)$$

where β_2 is a dispersion coefficient, γ is the Kerr nonlinearity factor, and $\mathbf{n}$ is the noise vector due to amplified spontaneous emission (ASE). The function $f(z) = \exp\left(-\alpha z + \alpha L_{\text{sp}}\left\lfloor z/L_{\text{sp}}\right\rfloor\right)$ represents amplification and attenuation in the fiber, where α is a fiber loss factor and L_{sp} is the span length [6, 20].

B. Regular Perturbation (RP) Model

Optical signal propagation through SSMF can be approximated with the first-order RP model, which solves the Manakov equation in the linear domain and bundles the nonlinear components into a small additive term [6]. Similar to SSFM-DBP, the direction of the RP model can be inverted by flipping the direction of $f(z)$ (i.e., switching the attenuation to amplification), and changing the sign of β_2 and γ parameters. Accordingly, the inverted RP takes the received signal $u(L,t)$ to reconstruct $\hat{u}(0,t)$ (i.e., back–propagated signal) for NLC. To achieve better results,

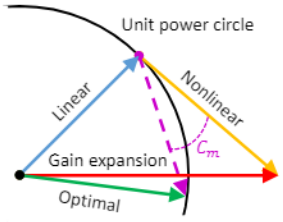

Fig. 1. Per-branch operation in the RP model versus LRP.

the γ parameter is often hand-tuned [8]. As shown in [20], this structure can be easily parallelized into separate branches and is suitable for DNNs. The first-order RP model is expressed as

$$\hat{\mathbf{u}}(0,t) = \hat{\mathbf{u}}_{\mathrm{L}}(0,t) + \hat{\mathbf{u}}_{\mathrm{NL}}(0,t) + \mathcal{O}(\gamma^2), \tag{2}$$

where $\mathcal{O}(\gamma^2)$ represents higher-order residuals to the Manakov model, L is a total link length, and the first term

$$\hat{\mathbf{u}}_{\mathrm{L}}(0,t) = \mathcal{D}_{z_N}[\mathbf{u}(L,t)] \tag{3}$$

is a linear branch, which contains a dispersion block, given by

$$\mathcal{D}_z[\cdot] = \mathcal{F}^{-1}\Big[\exp\left(i\beta_2 z\omega^2/2\right)\mathcal{F}[\cdot]\Big], \tag{4}$$

where $\mathcal{F}$ denotes the discrete Fourier transform (DFT), $\mathcal{F}^{-1}$ denotes the inverse DFT, and ω represents the angular frequency.

The second term $\mathbf{u}_{\mathrm{NL}}$ is an aggregation of N nonlinear branches as follows:

$$\hat{\mathbf{u}}_{\mathrm{NL}}(0,t) = \sum_{m=0}^{N-1} \mathcal{D}_{L-z_m}\left[\mathcal{K}_{z_m}\left[\mathcal{D}_{z_m}\left[\mathbf{u}(L,t)\right]\right]\right]. \tag{5}$$

Here, the nonlinear block $\mathcal{K}$ is given by

$$\mathcal{K}_{z_m}[\mathbf{u}(t)] = i\frac{8}{9}\gamma\Delta_m f(z_m)\|\mathbf{u}(t)\|^2\mathbf{u}(t), \tag{6}$$

where $\|\cdot\|$ denotes the 2-norm. Each branch is assigned a step size, forming an $(N+1)$-element vector $\boldsymbol{z} = \{z_0, z_1, \ldots z_N\}$. We use the last element of $\boldsymbol{z}$ as a step for the linear branch. The corresponding effective step size is calculated as

$$\Delta_m = \frac{1}{\alpha}\left(1 - \exp(-\alpha(z_m - z_{m-1}))\right), \qquad z_{-1} = 0. \tag{7}$$

Analogous to SPS in SSFM-DBP, the accuracy of RP increases with the number of branches and decreases with the optical launching power (i.e. more fiber nonlinearity) [5, 20]. Note that our model allows $z_n = \{0, L\}$ such that the pre-/post-linear block covers the entire link length. This is found to improve the performance for cases with small number of branches. However, most current literature do not consider such a branch [20].

C. ML-aided NLC based on Inverse RP model

Although the RP model can provide reasonably accurate approximation with a large number of branches, it suffers from a significant issue when the number of branches is small. As shown in Fig. 1, a simplified one-branch conventional RP model takes a linear signal (blue phasor) and adds a 90° shifted, amplitude-dependent signal (orange phasor) due to $\mathcal{K}$ operation, to provide the nonlinear phase rotation. Adding these two vectors may result in a phasor towards a wanted rotation (red). However, as the nonlinearity increases, it inevitably leads to gain expansion outside of the unit power circle (black line). Hence, the model accuracy is considerably reduced at higher power regimes.

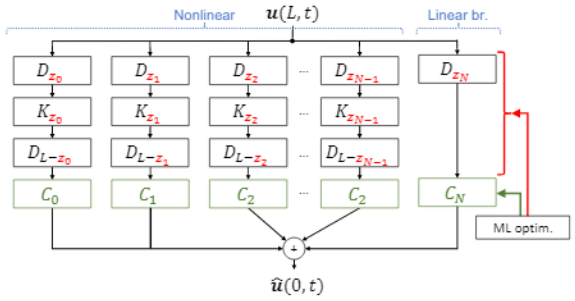

Fig. 2. LRP model with phasor correction at parallel branches.

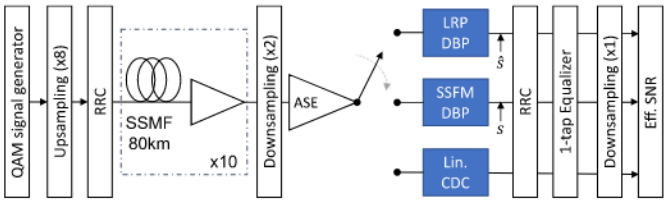

Fig. 3. System setup used for simulation.

Herein, we propose an improved RP model with a trainable phasor correction term (complex-valued scalar) $\boldsymbol{C} = \{C_0, C_1, \ldots, C_N\}$ at the end of each branch as shown in Fig. 2. The nonlinear term in (5) becomes

$$\hat{\mathbf{u}}_{\mathrm{NL}}(0,t) = \sum_{m=0}^{N-1} \mathcal{D}_{L-z_m}\left[\mathcal{K}_{z_m}\left[\mathcal{D}_{z_m}\left[\mathbf{u}(L,t)\right]\right]\right] \times C_m, \tag{8}$$

and the linear branch $\hat{\mathbf{u}}_{\mathrm{L}}(0,t)$ in (3) is modified as

$$\hat{\mathbf{u}}_{\mathrm{L}}(0,t) = \mathcal{D}_{z_N}[\mathbf{u}(L,t)] \times C_N. \tag{9}$$

With this simple modification, we can collectively adjust gain and rotation of each branch (purple arrow in Fig. 1) and find a better phasor, close to unit power (green in Fig. 1). Hence, the proposed LRP model can approximate the nonlinear effects more accurately and efficiently mitigate the unwanted gain expansion.

3. SYSTEM SETUP AND TRAINING STRATEGY

We next describe the simulation setup used to generate training/testing data, and to validate the LRP performance. We also discuss the training strategy and parameter initialisation.

A. System setup

Fig. 3 illustrates the system setup. A single-carrier DP 64-ary quadrature-amplitude modulation (QAM) signal with a symbol rate of 64Gbaud is generated at the transmitter. The generated electrical signal is pulse-shaped with a root-raised cosine (RRC) filter with a roll-off factor of 0.1. It is modulated into the optical carrier with an in-phase/quadrature modulator. The carrier is provided by an ideal laser with a central frequency of 1550nm and zero frequency offset. Similar to [10, 21], the laser induced phase noise is not considered for simplicity. Next, the analog optical signal is up-sampled to 8 sa/sym and propagated through $n_{\mathrm{span}} = 10$ spans of 80 km SSMF links. Each span is followed by an erbium-doped OA with a gain of 16dB and a noise figure of 4dB. The ASE noise is added as a lumped sum at the end

of the link. Additionally, the enhanced Gaussian noise (EGN) model [22] is used to emulate inter-channel noise of a 5-channel, 50 GHz spaced, wavelength division multiplexed (WDM) system. The SSMF has a Kerr coefficient of $\gamma = 1.2$ W/km, a loss factor of $\alpha = 0.2$ dB/km, and a dispersion factor of $\beta_2 = 17$ ps/nm/km. The received optical signal is downsampled to 2 sa/sym, and passes through either linear CD compensation (CDC), SSFM-DBP, or LRP-DBP model, to mitigate the channel impairments. The output is then shaped with a matched RRC filter. Subsequently, the received signal is normalized and de-rotated with a 1-tap equalizer and downsampled to 1 sa/sym. Lastly, the effective SNR is calculated from the residual error vector between the transmitted and the received signals.

B. Optimization strategy

The LRP model is trained by the Adam optimizer with a learning rate of 0.001 to find optimal values for C and z. We use 200 randomly generated data frames that are split 80% to 20% for training and testing, respectively. A single frame contains a snapshot of a coherently detected, digitally back-propagated signal as a target. The simulated systems are complex-valued. Each snapshot has 8,192 samples per polarization at an oversampling of 2 sa/sym. The frames, used for training, are simulated (forward and back-propagated) with 100-SPS Manakov SSFM. The forward propagated signal is downsampled to 2 sa/sym. The raw nonlinear phase rotation is not compensated prior to training and is included within the optimization, while 1-tap equalization is used after NLCs. Subsequently, the signal is passed to 100-SPS DBP and the corresponding nonlinearity compensated output is saved as a second element of the snapshot.

Although the detected signal is at 2 sa/sym, the forward propagation in the analog domain is at 8 sa/sym. Similar to [10], we use a logarithmic step distribution across each fiber span [23], and asymmetric SSFM (i.e. the nonlinear step is at the segment boundary) [24]. Note that the lumped ASE noise is disabled for training but enabled for validation, to show the best–case scenario performance. In real-life applications, where datasets are noisy, transfer learning could be applied.

We optimize each launch power separately. The mini-batch size is set to 40 frames; there are $B = 160$ batches (i.e. 160 frames) per epoch for training ($B = 40$ for validation); and the number of epochs is set to 6,500. The mean squared error (MSE) loss between the predicted $\hat{s}$ and the originally transmitted signal s is calculated for $N_s = 4096$ symbols per $P = 2$ polarizations:

$$\mathcal{L}(s,\hat{s}) = \frac{1}{B}\frac{1}{P}\frac{1}{N_s}\sum_{k=1}^{B}\sum_{j=1}^{P}\sum_{i=1}^{N_s}|s_i - \hat{s}_i|^2. \tag{10}$$

MSE in dB-scale is defined as $\text{MSE}_{\text{dB}} = 10\log_{10}[\mathcal{L}(s,\hat{s})]$. The complex-valued C is adjusted by the gradient of $\mathcal{L}(s,\hat{s})$.

C. Parameter initialization

Appropriate initialization of the trainable parameters is a key aspect to achieve higher performance and faster convergence. We initialize z as a logarithmic step distribution across each span [23], whereas C is initialized as a vector filled with ones on real axis and zeroes on imaginary axis. We also found that the model provides much better convergence when Δ_m is decoupled from z and the initial value of $f(z)$.

4. RESULTS AND DISCUSSION

Fig. 4 shows the effective SNR performance of the trained LRP model across launch powers between 0 to 8 dBm. Each of the data points was validated with 100 randomly generated symbol sequences (8,192 samples each) and compared with learned SSFM-DBP [8, 23] (i.e., γ-scalar and step size adjusted) with a similar number of SPS calculated as N/n_{span}. For reference, we show a linear channel (black dashed curve) and a linear CDC.

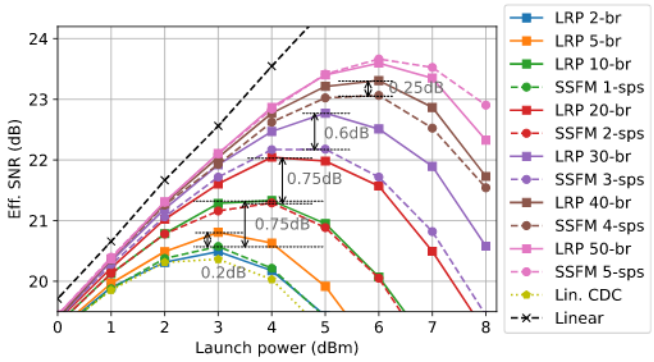

Fig. 4. Effective SNR performances of LRP versus SSFM.

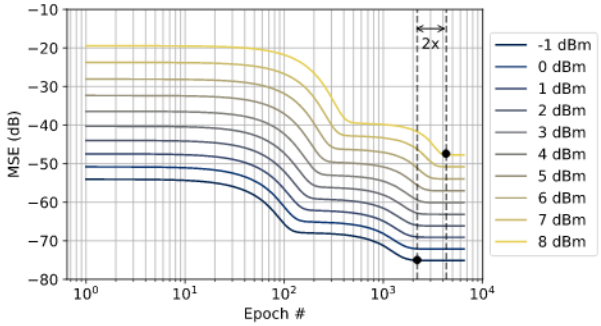

Fig. 5. MSE versus training epoch for 30-branch LRP.

The LRP shows considerable NLC gains (peak-to-peak) over the SSFM-DBP for less than 5-SPS. The 10-branch LRP has approximately 0.75 dB gain over γ-tuned 1-SPS SSFM-DBP, whereas the 20-br LRP has 0.75 dB gain over 2-SPS SSFM-DBP. The 30-br LRP is still 0.6 dB better, while 40-br LRP has only 0.25 dB gain over 4-SPS SSFM-DBP. Interestingly, the fractional models—with less than one branch per fiber span—are beneficial over 1-SPS SSFM-DBP. Specifically, the 5-br LRP model (i.e., 1/2-SPS) offers 0.2 dB gain, whereas 2-br LRP (i.e., 1/5-SPS) achieves nearly the same peak performance as 1-SPS SSFM-DBP. Note that for the fractional cases we use the value of $f(z) = 1$ to model multi-span nonlinearity per branch.

Fig. 5 shows the training trajectory for 30-br LRP. One can see that more training epochs are needed to converge for higher launch powers. Fig. 6 (left) shows the optimized z (solid lines) versus initial values for 30-branch LRP. It can be observed that the trained steps are close to log-distribution. The trained gains for 30-br LRP are shown in Fig. 6 (right). Except for the first and last branch, the trained gain corrections give a clear indication of the OA locations. Interestingly, the last branch (corresponding to the linear CDC) requires higher gain control values which are dependent on the launch power, whereas the gain contribution of the first branch is vastly reduced. In terms of phase rotation, as shown in Fig. 6 (right), we can see that the amount of rotation increases with the launch power (i.e., higher Kerr nonlinearity).

To justify the advantage of phasor correction in LRP, we fi-

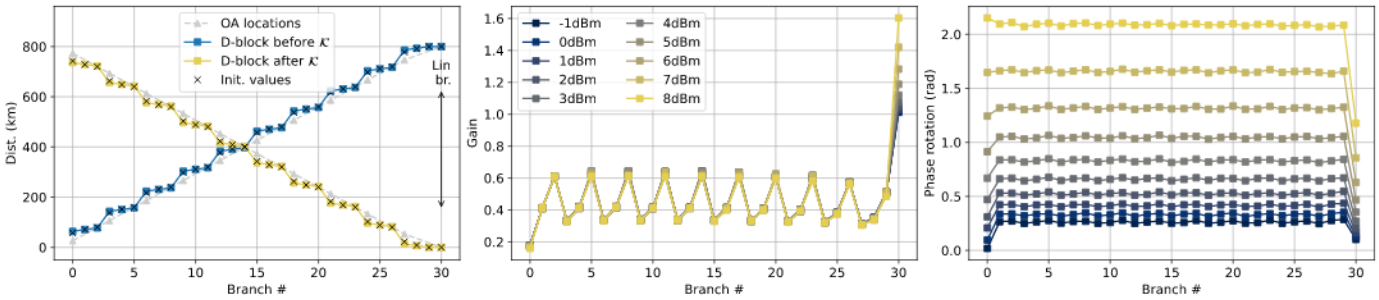

Fig. 6. Optimized steps (left), gain (center) and phase (right) values for a 10-span LRP with 30 branches.

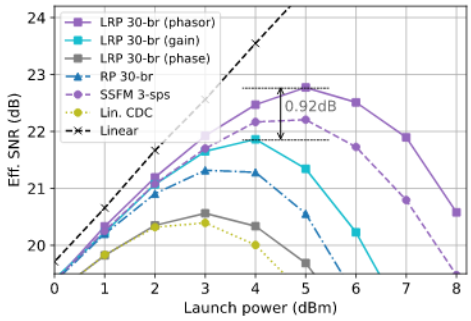

Fig. 7. LRP versus common gain and phase correction models.

nally compare the cases with common-gain (but per-branch phase) correction or common-phase (but per-branch gain) correction models for the 30-branch case in Fig. 7. The performances of the common gain and common phase corrections are significantly worse than the phasor-correcting LRP's. The peak to peak difference is 0.92 dB. Interestingly, the common phase rotation degrades the performance below an untrained RP model (dash-dotted blue), and is just 0.2 dB better than the linear CDC. For the common gain model, we observe a slight improvement over an untrained 30-br RP, whereas its performance is still worse than 3-SPS SSFM. It should be noted that there is a potential for both LRP and SSFM methods to be improved with additional filters before the nonlinear blocks, as shown in [25]. Although EGN model could be replaced with a full-field SSFM WDM simulation, the results presented herein can be viewed as a conservative estimate of the overall performance of a WDM system.

5. CONCLUSION

We demonstrated that ML-aided gain and phase rotation adjustment for the inverse RP model can significantly improve the performance of NLC in a 5-channel WDM system. The proposed LRP model achieved up to 0.75 dB gain over SSFM-based DBP by efficiently alleviating the intra–channel impairments. Fractional SPS modelling cases, where one branch represents multi-span nonlinearity, were shown to outperform 1-SPS SSFM-DBP. Moreover, the presented LRP can mitigate the gain expansion issue that is a main disadvantage of the conventional RP method [5, 6].

Disclosure

The authors declare no conflicts of interest.

REFERENCES

1. S. J. Savory, Opt. express **16**, 804 (2008).
2. M. Kuschnerov, F. N. Hauske, K. Piyawanno, B. Spinnler, M. S. Alfiad, A. Napoli, and B. Lankl, J. lightwave technology **27**, 3614 (2009).
3. P. J. Winzer, D. T. Neilson, and A. R. Chraplyvy, Opt. express **26**, 24190 (2018).
4. G. P. Agrawal, *Nonlinear Fiber Optics* (Elsevier, Amsterdam, 2013).
5. A. Vannucci, P. Serena, and A. Bononi, J. Light. Technol. **20**, 1102 (2002).
6. García-Gómez, Francisco Javier and Kramer, Gerhard, J. Light. Technol. **39**, 3390 (2021).
7. P. Johannisson and M. Karlsson, J. Light. Technol. **31**, 1273 (2013).
8. E. Ip and J. M. Kahn, J. Light. Technol. **26**, 3416 (2008).
9. J. W. Nevin, S. Nallaperuma, N. A. Shevchenko, X. Li, M. S. Faruk, and S. J. Savory, APL Photonics **6**, 121101 (2021).
10. C. Häger and H. D. Pfister, IEEE J. on Sel. Areas Commun. **39**, 280 (2021).
11. E. Sillekens, W. Yi, D. Semrau, A. Ottino, B. Karanov, D. Lavery, L. Galdino, P. Bayvel, R. I. Killey, S. Zhou *et al.*, 2020 IEEE Work. on Signal Process. Syst. (SiPS) (2020).
12. R. M. Bütler, C. Häger, H. D. Pfister, G. Liga, and A. Alvarado, J. Light. Technol. **39**, 949 (2020).
13. C. Fougstedt, C. Häger, L. Svensson, H. D. Pfister, and P. Larsson-Edefors, 2018 Eur. Conf. on Opt. Commun. (ECOC) pp. 1–3 (2018).
14. X. Lin, S. Luo, S. K. O. Soman, O. Dobre, L. Lampe, D. Chang, and C. Li, J. Light. Technol. (2021).
15. A. Redyuk, E. Averyanov, O. Sidelnikov, M. Fedoruk, and S. Turitsyn, J. Light. Technol. **38**, 1250 (2020).
16. I. Kozulin and A. Redyuk, Opt. Commun. **493**, 127026 (2021).
17. S. Zhang, F. Yaman, K. Nakamura, T. Inoue, V. Kamalov, L. Jovanovski, V. Vusirikala, E. Mateo, Y. Inada, and T. Wang, Nat. communications **10**, 1 (2019).
18. T. Xu, T. Jin, W. Jiang, J. Zhang, X. Yi, and K. Qiu, Asia Commun. Photonics Conf. pp. T4A–37 (2021).
19. O. S. Kumar, L. Lampe, S. Luo, M. Jana, J. Mitra, and C. Li, Signal Process. Photonic Commun. SpF2E (2021).
20. V. Neskorniuk, A. Carnio, V. Bajaj, D. Marsella, S. K. Turitsyn, J. E. Prilepsky, and V. Aref, 2021 ECOC (2021).
21. C. Li, Y. Wang, L. Han, S. Chen, Q. Zhang, L. Yang, and X. Xin, Opt. Commun. **507**, 127627 (2022).
22. D. Semrau, E. Sillekens, R. I. Killey, and P. Bayvel, J. Light. Technol. **37**, 5122 (2019).
23. J. Zhang, X. Li, and Z. Dong, J. lightwave technology **31**, 3546 (2013).
24. E. Ip, J. lightwave technology **28**, 939 (2010).
25. L. B. Du and A. J. Lowery, Opt. express **18**, 17075 (2010).