A New Twist on Low-Complexity Digital Backpropagation

Stella Civelli, *Member, IEEE*, Debi Pada Jana, *Member, IEEE*, Enrico Forestieri, *Senior Member, IEEE*, and Marco Secondini, *Senior Member, IEEE*

***Abstract*—This work proposes a novel low-complexity digital backpropagation (DBP) method, with the goal of optimizing the trade-off between backpropagation accuracy and complexity. The method combines a split step Fourier method (SSFM)-like structure with a simplified logarithmic perturbation method to obtain a high accuracy with a small number of DBP steps. Subband processing and asymmetric steps with optimized splitting ratio are also employed to further reduce the number of steps required to achieve a prescribed performance. The first part of the manuscript is dedicated to the derivation of a simplified logarithmic-perturbation model for the propagation of a dual-polarization multiband signal in an optical fiber, which serves as a theoretical background for the development of the proposed coupled-band enhanced split step Fourier method (CB-ESSFM) and for the analytical calculation of the model coefficients. Next, the manuscript presents a digital signal processing algorithm for the implementation of DBP based on a discrete-time version of the model and an overlap-and-save processing strategy. Practical approaches for the optimization of the coefficients used in the algorithm and of the splitting ratio of the asymmetric steps are also discussed. A detailed analysis of the computational complexity of the algorithm is also presented. Finally, the performance and complexity of the proposed DBP method are investigated through numerical simulations and compared to those of other methods. In a five-channel 100 GHz-spaced wavelength division multiplexing system over a 15×80 km single-mode-fiber link, the proposed CB-ESSFM achieves a gain of about 1 dB over simple dispersion compensation with only 15 steps (corresponding to 681 real multiplications per 2D symbol), with an improvement of 0.9 dB over conventional SSFM and almost 0.4 dB over our previously proposed ESSFM. Significant gains and improvements are obtained also at lower complexity. For instance, the gain reduces to a still significant value of 0.34 dB when a single DBP step is employed, requiring just 75 real multiplications per 2D symbol. A similar analysis is performed also for longer links, confirming the good performance of the proposed method.**

***Index Terms*—Digital backpropagation, nonlinear fiber channel, optical fiber communication, perturbation methods.**

Received 19 November 2024; revised 7 February 2025; accepted 11 February 2025. Date of publication 13 February 2025; date of current version 16 May 2025. This work was supported by the European Union - Next Generation EU under the Italian National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.3, CUP J53C22003120001, CUP B53C22003970001, partnership on "Telecommunications of the Future" (PE00000001 - program "RESTART"). An earlier version of this paper was presented at the European Conference on Optical Communication 2024 in Frankfurt [1]. *(Corresponding author: Stella Civelli.)*

Stella Civelli is with the Cnr-Istituto di Elettronica e di Ingegneria dell'Informazione e delle Telecomunicazioni (CNR-IEIIT), 56122 Pisa, Italy, and also with the Telecommunications, Computer Engineering, and Photonics (TeCIP) Institute, Scuola Superiore Sant'Anna, 56127 Pisa, Italy (e-mail: stella.civelli@cnr.it).

Debi Pada Jana was with the Telecommunications, Computer Engineering, and Photonics (TeCIP) Institute, Scuola Superiore Sant'Anna, 56127 Pisa, Italy. He is now with the College of Optics and Photonics (CREOL), University of Central Florida, Orlando, FL 32816 USA.

Enrico Forestieri and Marco Secondini are with the Telecommunications, Computer Engineering, and Photonics (TeCIP) Institute, Scuola Superiore Sant'Anna, 56127 Pisa, Italy, and also with the National Laboratory of Photonic Networks, CNIT, 56124 Pisa, Italy.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JLT.2025.3542166.

Digital Object Identifier 10.1109/JLT.2025.3542166

I. Introduction

THE continuous growth of global data traffic is pushing optical networks to their limits [2], [3], driving the investigation of several potential solutions [4], [5]. Among these, improving the performance of existing long-haul links by modifying the digital signal processing (DSP) of the transmitter and receiver is particularly attractive due its low cost, feasibility and immediate applicability [5, Ch. 10] [6].

A major limitation in current fiber-optic coherent systems is Kerr nonlinearity [7], which degrades system performance as optical power increases [8]. To mitigate nonlinear impairments in optical fibers, various DSP techniques have been proposed, including digital back-propagation (DBP) [9], [10], Volterra series equalization [11], and maximum-likelihood sequence detection [12], [13]. In particular, DBP aims to reverse the effects of channel propagation by digitally emulating the propagation of the signal through a fictitious fiber link, equal to the actual transmission link but reversed in space and with opposite propagation parameters. The most common implementation of DBP is based on the split-step Fourier method (SSFM) [14], [7, Sec. 2.4.1], which decomposes the propagation process into a series of simple independent linear and nonlinear steps. However, to accurately implement DBP, a large number of SSFM steps may be required, resulting in significant computational complexity, which limits its practical deployment.

Several SSFM-based methods have been proposed to achieve a good trade-off between performance and complexity. Filtered DBP accounts for correlations among neighboring symbols in the generation of the nonlinear phase rotation in each step [15], [16], [17]. The enhanced SSFM (ESSFM) reduces the number of required steps by improving the accuracy of the nonlinear step, accounting for the interaction between dispersion and intrachannel nonlinearity, practically generalizing filtered DBP through the use of a finite impulse response (FIR) filter with optimized coefficients [18], [19]. The SSFM and its more efficient variants can be directly used also for wavelength-division multiplexing (WDM) signals, as long as the channels are jointly

represented as a single optical field. In this case, however, it might be computationally more efficient to represent the WDM channels separately and account for their interaction through a systems of coupled NLSEs [20], [21]. This idea is exploited in the coupled-channel ESSFM (CC-ESSFM), which improves on ESSFM by accounting also for interchannel nonlinear effects in WDM systems [22], [23]. A similar approach can be used also for the propagation of a single channel by digitally dividing it into subbands [24]. Recently, machine learning (ML) techniques have also been explored for nonlinearity compensation, such as learned DBP [25], [26], subband-processing learned DBP [27], fully learned perturbation-based nonlinearity compensation [28], and nonlinearity mitigation based on carrier phase recovery [29].

In this work, we propose a novel single-channel DBP technique, referred to as coupled-band ESSFM (CB-ESSFM), which combines the SSFM structure with a logarithmic-perturbation (LP) model [30], [31], [32] and subband processing to increase the accuracy of nonlinear steps while keeping their complexity comparable to that of conventional SSFM. The proposed propagation model constitutes a common theoretical basis for the derivation of the ESSFM and CC-ESSFM methods, as well as the CB-ESSFM method proposed in this work. The proposed method depends on a set of coefficients that can be either obtained analytically from the LP model or optimized numerically to maximize the performance. Asymmetric steps with an optimized splitting ratio are also employed to further improve the performance of the method. Numerical simulations demonstrate that the proposed CB-ESSFM significantly enhances the performance of state-of-the-art DBP methods while maintaining low computational complexity.

The manuscript is organized as follows. Section II presents the theoretical background of the proposed CB-ESSFM and an analytical expression for its coefficients. Section III covers the practical implementation of the CB-ESSFM, detailing the digital signal processing algorithm, discussing the optimization of both the splitting ratio and the coefficients, and analyzing the computational complexity. In Section IV, we verify the accuracy of the model and demonstrate the performance of the proposed DBP approach across various scenarios using numerical simulations. Finally, Section V summarizes the conclusions.

II. Theoretical Model

In this Section, we present the theoretical background for the derivation of the proposed DBP technique. In particular, we derive a fiber propagation model based on the frequency resolved logarithmic perturbation (FRLP) [31], [32], starting from a single-polarization single-band scenario, and then extending it to the dual-polarization and multi-band cases. This model establishes a general theoretical framework for deriving the enhanced SSFM (ESSFM) and its variants [18], [19], [22], [23, Ch. 2]—including the DBP method that will be described in Section III-A—and for evaluating the coefficients of the method.

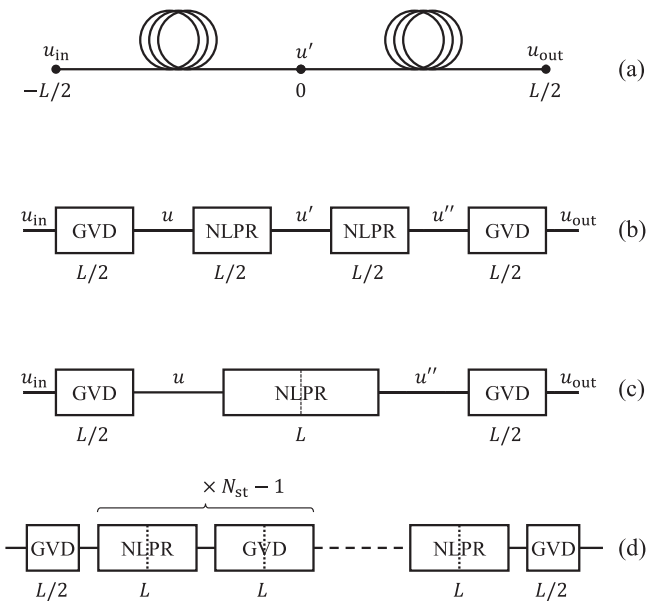

Fig. 1. Derivation of the ESSFM model: (a) each propagation step of length L is divided into two halves; (b) the approximated FRLP model is applied to each half (in reverse order in the second half); (c) the two adjacent NLPR blocks are combined into a single NLPR block; (d) the overall link of length $N_{\text{st}}L$ is divided into N_{st} steps, each modeled as above, and pairs of adjacent GVD blocks are combined into single GVD blocks.

A. Single-Polarization Single-Band Model

In the single-polarization single-band scenario, the propagation of the optical signal is governed by the nonlinear Schrödinger equation (NLSE) [3], [7, Ch. 2]

$$\frac{\partial u(z,t)}{\partial z} = j\frac{\beta_2}{2}\frac{\partial^2 u(z,t)}{\partial t^2} - j\gamma P g(z)|u(z,t)|^2 u(z,t) \quad (1)$$

where $u(z,t)$ is the lowpass equivalent representation of the propagating signal, normalized to have unit power at any distance z in the link; β_2 the fiber dispersion parameter; γ the nonlinear coefficient; $g(z)$ the power profile along the link (due to attenuation and amplification) normalized to the reference power P (typically taken as the launch power), so that $\sqrt{Pg(z)}u(z,t)$ is the actual lowpass equivalent representation of the optical signal at distance z, with power $Pg(z)$.

As shown in Fig. 1(a), we consider the propagation through a step of length L and, without loss of generality, set the origin $z = 0$ in the middle of the step, dividing the propagation into two half steps. The step represents a generic portion of the link that possibly includes one or more spans of fibers (or fractions of them) and optical amplifiers. We approximate the propagation from $u_{\text{in}}(t) \triangleq u(-L/2,t)$ to $u'(t) \triangleq u(0,t)$ by applying an approximated version of the FRLP model [31], [32], which is derived in Appendix A and consists of a group-velocity dispersion (GVD) block followed by a nonlinear phase rotation (NLPR). On the other hand, using the same expedient employed in the symmetric version of the classical SSFM [33, Sec. 2.4.1], we approximate the propagation from $u'(t)$ to $u_{\text{out}}(t) \triangleq u(L/2,t)$ by applying the same model of the first half but in reverse order, obtaining the block diagram in Fig. 1(b). The cascade of the

direct and reverse FRLP models ensures that error terms with an odd symmetry around $z = 0$ cancel out, resulting in higher overall accuracy, and yields the symmetric configuration in Fig. 1(c), where the two half-length NLPR blocks are combined into a single NLPR block.[1] After some calculation reported in Appendix A, we obtain the following model[2]

$$u(t) = \int_{-\infty}^{+\infty} U_{\text{in}}(f) H(L/2, f) e^{j2\pi f t} df \tag{2}$$

$$u''(t) = u(t) e^{-j\theta(t)} \tag{3}$$

$$u_{\text{out}}(t) = \int_{-\infty}^{+\infty} U''(f) H(L/2, f) e^{j2\pi f t} df \tag{4}$$

where

$$H(z, f) = \exp\left(-j2\pi^2 \beta_2 f^2 z\right) \tag{5}$$

is the transfer function due to the accumulated GVD after a propagation distance z;

$$\theta(t) = P \iint_{\mathbb{R}^2} K(\mu, \nu) U(\mu) U^*(\nu) e^{j2\pi(\mu-\nu)t} d\mu d\nu \tag{6}$$

is the NLPR due to Kerr nonlinearity (and its interaction with GVD); and

$$K(\mu, \nu) = \int_{-L/2}^{L/2} \gamma g(z) H(z, \mu) H^*(z, \nu) H(-z, \mu - \nu) dz \tag{7}$$

is a frequency-domain second-order Volterra kernel that accounts for the efficiency with which pairs of frequency components of the propagating signal contribute to the NLPR. A closed-form expression for (7) is given in Section II-D.

As in the classical SSFM, the propagation through a long link is usually handled by dividing it into N_{st} steps, where N_{st} is selected to obtain the desired trade-off between complexity and accuracy. When modeling each step as in Fig. 1(c), pairs of adjacent half-length GVD blocks combine into single full-length GVD blocks, yielding the overall block diagram in Fig. 1(d).

For a practical implementation of the model for the development of DBP methods, which will be discussed in detail in Section III, it is convenient to consider a discrete-time representation of the signals. As shown in Appendix B, by assuming a sufficiently large sampling rate R,[3] we obtain a discrete-time version of (6), in which the NLPR samples $\theta[k] \triangleq \theta(k/R)$ are related to the time-domain samples of the linearly propagated signal $u[k] \triangleq u(k/R)$ through a discrete-time second-order Volterra kernel. The corresponding expression is calculated in Appendix B and, as shown in (46), still involves the computation of an infinite number of terms. In practice, however, the discrete-time second-order Volterra kernel, whose coefficients $d[m, n]$ are defined in (47), has a limited duration that depends mainly on the GVD accumulated over the step length L and (in minor part) on the pulse duration. This means that it can be practically truncated to a finite number of terms. Moreover, to further simplify the computation, we consider only the diagonal terms of the kernel, represented by the coefficients $c[m] \triangleq d[m, m]$ and neglect the (usually smaller) off-diagonal terms $d[m, n]$, with $m \neq n$. In this way, letting N_c be the number of relevant pre- and post-cursor coefficients, (46) simplifies to

$$\theta[k] = \sum_{m=-N_c}^{N_c} c[m] |u[k-m]|^2 \tag{8}$$

where

$$c[m] = \frac{P}{R^2} \int_{-R/2}^{R/2} \int_{-R/2}^{R/2} K(\mu, \nu) e^{j2\pi(\mu-\nu)m/R} d\mu d\nu \tag{9}$$

can be seen as the coefficients of the discrete-time impulse response that relates the signal intensity to the NLPR. Equations (2)–(4), (8), and (9) form the theoretical foundation for deriving the ESSFM and for analytically computing its coefficients. The necessary model extensions to derive the CC-ESSFM [22], [23] and the CB-ESSFM are detailed in the upcoming sections.

B. Single-Polarization Multiband Model

The accuracy of the ESSFM model in Fig. 1(d) depends on the amount of dispersion accumulated in each step, hence on the number of steps N_{st} in which the link is divided. The accumulated dispersion increases with the signal bandwidth, so that wider bandwidth signals typically requires more steps to achieve a prescribed accuracy. At the same time, more steps entails a higher computational complexity for DBP implementation, as detailed in Section III-D. Therefore, a possible solution to reduce computational complexity while maintaining high accuracy is to divide wide-band signals into subbands.

The calculations carried out for the single-band case can be readily extended to the multi-band scenario. In this context, the propagating signal $u(z, t)$, with bandwidth B, is partitioned into N_{sb} subbands and expressed as

$$u(z, t) = \sum_{i=1}^{N_{\text{sb}}} u_i(z, t) \tag{10}$$

where $u_i(z, t)$ is the signal portion contained in the ith subband, centered at frequency f_i and with a bandwidth B/N_{sb}.[4]

[1] Strictly speaking, such a configuration is actually symmetric and ensures error cancellation only when also the power profile $g(z)$ is symmetric around $z = 0$, which is not true in general (e.g., for long steps with relevant attenuation). In more general cases, a higher accuracy is achieved by splitting the step asymmetrically, as discussed in greater detail in Section III-C.

[2] The Fourier transform of a generic function $x(t)$ is denoted with the corresponding capital letter and defined as $X(f) = \int_{-\infty}^{+\infty} x(t) \exp(-j2\pi f t) \mathrm{d}t$. Also, note that, with an abuse of notation, we use the same symbol u to denote the propagating signal $u(z, t)$, the linearly propagated signal after the first half step $u(t)$, and (later) its samples $u[k]$. A similar notation will be used in the multi-band case.

[3] In principle, the equation holds when the sampling rate is at least equal to the Nyquist rate for $u(t)$, i.e., twice its bandwidth. However, the NLPR induces some spectral broadening, meaning that an accurate representation of the output signal $u_{\text{out}}(t)$ may require a slightly larger sampling rate. This is important when considering the cascade of many steps to represent the propagation through a long link, as will be discussed in detail in Section III.

[4] An alternative representation is often adopted, replacing each bandpass component u_i in (10) with its equivalent lowpass representation multiplied by the corresponding carrier frequency term $\exp(j2\pi f_i t)$. This choice, however, entails a slightly more complex formalism, as it requires the inclusion of some additional terms in the set of coupled differential equations derived below to account for phase shifts and walk-offs between subbands.

Since the impact of dispersion within each subband is significantly less pronounced than across the entire signal bandwidth, we can increase the step size L of the model (thereby using fewer steps for the entire link) without sacrificing accuracy. At the same time, when dividing the signal into subbands, we need to decide how to handle the nonlinear interaction between them. The simplest option is to neglect it, applying the ESSFM separately to each subband as if it were the only propagating signal. In this way, we reduce the overall complexity (thanks to the increased step size), but we also lose accuracy (due to the neglected interband nonlinearity). An alternative option is to replace (10) in (1), hence expanding the nonlinear term in the latter into several sub-terms, typically categorized as self-phase modulation (SPM), cross-phase modulation (XPM), and four-wave mixing (FWM) terms [7]. Taking into account all sub-terms in the model would significantly amplify the complexity of each step, potentially nullifying the effort made to reduce it. However, the various terms have varying degrees of significance in contributing to the overall nonlinearity, particularly in the presence of dispersion. The strategy, therefore, is to discard less relevant terms, such as FWM ones, to strike the optimal balance between accuracy and complexity.

By replacing (10) in (1), and neglecting (non-degenerate) FWM terms, we obtain a set of N_{sb} coupled NLSEs

$$\frac{\partial u_i}{\partial z} = j\frac{\beta_2}{2}\frac{\partial^2 u_i}{\partial t^2} - j\gamma P g(z)\left(|u_i|^2 + 2\sum_{\ell\neq i}|u_\ell|^2\right)u_i \qquad i = 1,\ldots,N_{\text{sb}} \tag{11}$$

each including an SPM term and $N_{\text{sb}} - 1$ XPM terms.

As in the single-channel case, we approximate the propagation through the step of length L depicted in Fig. 1(a) by using the FRLP model in a split-step symmetric configuration, including this time also the XPM terms in the NLPR [32, eq. (9)]. The solution for the signal u_i in the ith subband, whose derivation is omitted due to its similarity with the single-band case, can still be expressed as in (2)–(4) and visualized by the block diagram in Fig. 1(c) (adding the subscript i to all involved signals), but with a different NLPR that includes both SPM and XPM terms

$$\theta_i(t) = P\left(\iint_{\mathbb{R}^2} K(\mu,\nu)U_i(\mu)U_i^*(\nu)e^{2\pi j(\mu-\nu)t}d\mu d\nu + \sum_{\ell\neq i}\iint_{\mathbb{R}^2} 2K(\mu,\nu)U_\ell(\mu)U_\ell^*(\nu)e^{2\pi j(\mu-\nu)t}d\mu d\nu\right) \tag{12}$$

Considering a discrete-time representation of the signals; letting $u_i[k] \triangleq u_i(k/R')$ and $\theta_i[k] \triangleq \theta_i(k/R')$ be the samples of the signal and NLPR, respectively, taken at rate $R' = R/N_{\text{sb}}$; truncating the discrete-time Volterra kernel; and neglecting its off-diagonal terms as in the single-band case, we eventually obtain

$$\theta_i[k] = \sum_{m=-N_c}^{N_c} c_{ii}[m]|u_i[k-m]|^2 + 2\sum_{\ell\neq i}\sum_{m=-N_c}^{N_c} c_{i\ell}[m]|u_\ell[k-m]|^2 \tag{13}$$

where

$$c_{i\ell}[m] = \frac{P}{R'^2}\int_{f_{i\ell}-R'/2}^{f_{i\ell}+R'/2}\int_{f_{i\ell}-R'/2}^{f_{i\ell}+R'/2} K(\mu,\nu)e^{j2\pi(\mu-\nu)m/R'}d\mu d\nu \tag{14}$$

is the m-th coefficient of the impulse response that accounts for the impact of the intensity of the ℓth subband on the NLPR of the ith subband, and $f_{i\ell} = f_\ell - f_i$ is the frequency separation between the two subbands. For $i = \ell$, $f_{i\ell} = 0$ and (14) reduces to (9).

C. Dual-Polarization Multiband Model

The previous model can be eventually extended to the case of a dual-polarization signal, whose propagation is governed by the Manakov equation [34]—formally equal to (1) but considering $u = (x, y)$ as the Jones vector collecting the normalized complex envelopes of the two polarization components, x and y, and $|u|^2 = |x|^2 + |y|^2$ as the squared-norm of the vector. As in (10), we divide the signal into N_{sb} subbands, where each subband contains two polarization components $u_i = (x_i, y_i)$, and replace them in (1). In this case, besides SPM, XPM, and FWM, we obtain also cross-polarization modulation (XPolM) terms [35]. Though XPolM can typically be as relevant as XPM, its effect cannot be simply represented as an NLPR, and its inclusion in the model significantly increases the complexity. In this case, we consider two possible solutions to obtain a good trade-off between accuracy and complexity. The first solution is obtained by omitting both FWM and XPolM terms, obtaining a set of N_{sb} coupled Manakov equations

$$\frac{\partial u_i}{\partial z} = j\frac{\beta_2}{2}\frac{\partial^2 u_i}{\partial t^2} - j\gamma P g(z)\left(|u_i|^2 + \frac{3}{2}\sum_{\ell\neq i}|u_\ell|^2\right)u_i \qquad i = 1,\ldots,N_{\text{sb}} \tag{15}$$

formally equivalent to (11) but for the 3/2 degeneracy factor in front of the XPM terms (instead of 2) and for the vector interpretation as explained above. Consequently, the propagation of the ith subband over the step of length L depicted in Fig. 1(a) can still be expressed as in (2)–(4) and visualized by the block diagram in Fig. 1(c), provided that the signals are interpreted as two-component vectors, and the GVD and NLPR are applied to each polarization component. In this case, considering a discrete-time representation of the signals, the NLPR in the ith subband can be expressed as

$$\theta_i[k] = \sum_{m=-N_c}^{N_c} c_{ii}[m]|u_i[k-m]|^2 + \frac{3}{2}\sum_{\ell\neq i}\sum_{m=-N_c}^{N_c} c_{i\ell}[m]|u_\ell[k-m]|^2 \tag{16}$$

i.e., as in the single-polarization case (13), but with the squared norm replacing the squared modulus, and a 3/2 degeneracy factor

in front of the XPM terms. The coefficients are the same as in (14). This solution provides the theoretical foundation for the reduced-complexity CC-ESSFM in [23] and serves as a basis for developing the CB-ESSFM in Section III-A.

On the other hand, the second solution is obtained by neglecting FWM terms and incorporating only certain XPolM terms in the multi-band expansion of the vector form of (1). Specifically, to avoid a significant increase in computational complexity, we include only those terms that can still be expressed as an NLPR, albeit of a different entity for the two polarizations. In this case, it is necessary to represent the evolution of the two polarization components of each subband with two coupled equations, obtaining a set of $2N_{\text{sb}}$ coupled differential equations [6]

$$\frac{\partial x_i}{\partial z} = j\beta_2 \frac{\partial^2 x_i}{\partial t^2} - j\gamma P g(z) + \left(|x_i|^2 + |y_i|^2 + \sum_{\ell \neq i} 2|x_\ell|^2 + |y_\ell|^2 \right) x_i \tag{17}$$

$$\frac{\partial y_i}{\partial z} = j\beta_2 \frac{\partial^2 y_i}{\partial t^2} - j\gamma P g(z) + \left(|x_i|^2 + |y_i|^2 + \sum_{\ell \neq i} |x_\ell|^2 + 2|y_\ell|^2 \right) y_i \qquad i = 1, \ldots, N_{\text{sb}} \tag{18}$$

The equations above imply that the XPM induced by co-polarized subbands is twice as effective as the XPM resulting from orthogonally-polarized subbands. This stands in contrast to the simpler model in (15), where the same XPM affects both polarizations, corresponding to the average XPM affecting the two polarizations in (17) and (18). For the rest, the equations are similar to those in (11), and the application of the FRLP method, whose development is omitted due to its similarity with the previous cases, yields a similar model. The propagation of the ith subband over the step of length L depicted in Fig. 1(a) can still be expressed as in (2)–(4), but replacing the NLPR in (3) with different NLPRs $\theta_i^x(t)$ and $\theta_i^y(t)$ for the two polarization components

$$x_i''(t) = x_i(t) \exp\big(-j\theta_i^x(t)\big) \tag{19}$$

$$y_i''(t) = y_i(t) \exp\big(-j\theta_i^y(t)\big) \tag{20}$$

Eventually, considering a discrete-time representation of the signals and letting $x_i[k] = x_i(k/R'), y_i[k] = y_i(k/R'), \theta_i^x[k] = \theta_i^x(k/R')$, and $\theta_i^y[k] = \theta_i^y(k/R')$, the NLPRs can be expressed as

$$\theta_i^x[k] = \sum_{m=-N_c}^{N_c} c_{ii}[m]\big(x_i^2[k+m] + y_i^2[k+m]\big) + \sum_{\ell \neq i} \sum_{m=-N_c}^{N_c} c_{i\ell}[m]\big(2x_\ell^2[k+m] + y_\ell^2[k+m]\big) \tag{21}$$

$$\theta_i^y[k] = \sum_{m=-N_c}^{N_c} c_{ii}[m]\big(x_i^2[k+m] + y_i^2[k+m]\big) + \sum_{\ell \neq i} \sum_{m=-N_c}^{N_c} c_{i\ell}[m]\big(x_\ell^2[k+m] + 2y_\ell^2[k+m]\big) \tag{22}$$

The coefficients $c_{i\ell}[m]$ are as in (14) and do not depend on the particular polarization. This solution provides the theoretical foundation for the full-complexity CC-ESSFM [22], [23] and could be used also to develop a slightly different version of the CB-ESSFM, though this is not explored in this work.

D. Evaluation of the CB-ESSFM Coefficients

In this Section, we provide an analytical expression for the kernel function (7) and a numerical procedure for the evaluation of the CB-ESSFM coefficients defined in (14). For simplicity, we assume that the step in Fig. 1(a) consists of N_{sp} identical spans of fiber of length L/N_{sp}, with attenuation coefficient α, dispersion parameter β_2, nonlinear coefficient γ, and periodic amplification at the end of each span which exactly compensates for the span loss. In this case, letting $a = \alpha/2$ and $b \triangleq 2\pi^2\beta_2\nu(\mu - \nu)$, after a few calculations reported in Appendix C, the kernel function can be expressed as

$$K(\mu, \nu) = \gamma e^{-aL/N_{\text{sp}}} \frac{\sinh\big((a+jb)L/N_{\text{sp}}\big)\sin(bL)}{(a+jb)\sin(bL/N_{\text{sp}})} \tag{23}$$

The expression (23) can be used also when the step is just a fraction of a span. In this case, it is sufficient to set $N_{\text{sp}} = 1$, L to the actual length of the step (not the whole span), and the reference power P in (14) to the power at the input of the step. For instance, if a span of length L_{sp} is divided into three equal steps, the kernel function is the same for all the steps and is obtained by setting $L = L_{\text{sp}}/3$ in (23), while the power P to be used in (14) is different: it equals the launch power in the first step, is attenuated by $\exp(-\alpha L)$ in the second step, and by $\exp(-2\alpha L)$ in the third step. The approach can easily be extended to derive more general closed-form expressions for steps that include several pieces of fiber with different length and propagation parameters.

Given the kernel function (23), the coefficients $c_{i\ell}[m]$ can be evaluated numerically from (14), e.g., as the diagonal terms of the two-dimensional FFT of the kernel function (23), evaluated over a grid of $(2N_c+1) \times (2N_c+1)$ frequency values in the range $\mu, \nu \in [f_{i\ell} - R'/2, f_{i\ell} + R'/2]$. Alternatively, as discussed in Section III, the CB-ESSFM coefficients can also be obtained through a numerical optimization procedure aimed at minimizing/maximizing some loss/performance metric (e.g., the mean square error). For a larger frequency separation $f_{i\ell}$, the walk-off between subbands induced by dispersion increases. Numerically, this effect results in a slower decay of the magnitude of the coefficients $c_{i\ell}[m]$ with m. This implies that, in general, the impulse response in (13) can be truncated to a number of coefficients N_c that depends on $\ell - i$. This dependence, which we omitted in (13) to keep the notation simple, will be analyzed and exploited in Section III-B.

Here we report some general properties of the coefficients $c_{i\ell}[m]$, which can be inferred from (14) and (23) and can be used to simplify their evaluation and optimization, as discussed in greater detail in Section III:

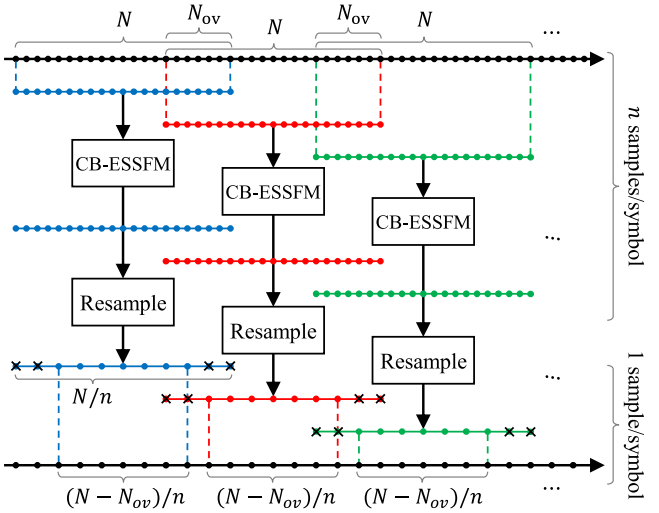

Fig. 2. The received signal is processed by the CB-ESSFM algorithm by using oversampling and overlap and save.

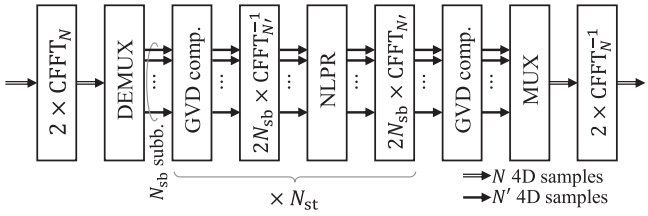

Fig. 3. CB-ESSFM algorithm with N_{st} steps and N_{sb} subbands.

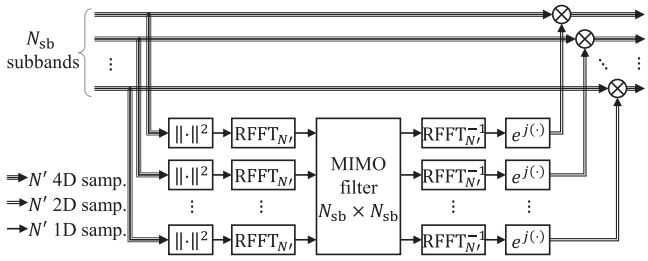

Fig. 4. Nonlinear phase rotation (NLPR) of the CB-ESSFM algorithm.

- they are linearly proportional to the product γP;
- they are independent of the modulation format (e.g., they are the same for 16-QAM and 64-QAM, with and without shaping);
- they depend on the frequency distance $f_\ell - f_i$ between the interfering channel and the observed channel, not on their absolute position, meaning that, for equally spaced channels, they depend on i and ℓ only through the difference $\ell - i$, so that

$$c_{i,\ell}[m] = c_{i+k,\ell+k}[m], \quad \forall k \in \mathbb{Z} \tag{24}$$

- they satisfy the symmetry condition

$$c_{i,\ell}[m] = c_{\ell,i}[-m] \tag{25}$$

  implying that the SPM coefficients have always an even symmetry

$$c_{i,i}[m] = c_{i,i}[-m]. \tag{26}$$

III. Implementation

This Section discusses the practical implementation of the proposed DBP method and is divided into four parts. The first part describes the signal processing algorithm. The second part explains the procedure for the numerical optimization of the CB-ESSFM coefficients; the third part introduces an alternative implementation based on asymmetric steps and discusses the optimization of their splitting ratio. Finally, the last part provides information about the computational complexity of the method.

A. The Coupled-Band ESSFM (CB-ESSFM)

The proposed DBP algorithm is based on the theoretical model described in the previous section and on the overlap-and-save technique [36], as schematically depicted in Fig. 2. First, the received signal is sampled at n samples per symbol, where the oversampling factor $n \geq 1 + r$ is selected to account for the roll-off factor r of the modulation pulses and the bandwidth expansion induced by nonlinearity. The sequence of 4D samples (two polarizations and two quadratures per sample) is then processed block-wise by using the overlap-and-save technique [36]. In practice, the sequence is divided into partially overlapping blocks of N samples, with an overlap of N_{ov} samples. The overlap should at least equal the duration of the overall channel response, which is mainly determined by the accumulated GVD. Each block is processed by the CB-ESSFM algorithm, obtaining the corresponding N output samples, and then resampled at one sample per symbol, obtaining N/n samples. Eventually, the overlapping samples are discarded—$N_{\mathrm{ov}}/2n$ on each side of each block—while the remaining $(N - N_{\mathrm{ov}})/n$ per block are saved and recombined to form the whole output sequence.

The processing performed by the CB-ESSFM algorithm on each block of N 4D samples follows the scheme in Fig. 1(d), with the dual-polarization multiband processing described in Section II-C and the NLPR in (16).[5] The algorithm is further detailed in Fig. 3 and 4. At the input, two complex FFTs (CFFTs)—one for each polarization—are performed on the whole block of N samples, followed by a demultiplexer (DEMUX) that divides the signal into N_{sb} subbands (at no cost in the frequency domain), each represented by $N' = N/N_{\mathrm{sb}}$ samples. Then, GVD compensation and nonlinear phase rotation (NLPR) are iteratively performed N_{st} times, each followed, respectively, by $2N_{\mathrm{sb}}$ (one per each subband and polarization) inverse and direct CFFTs of size N' to perform GVD compensation in the frequency domain and nonlinear phase rotation in the time domain. Finally, one additional GVD compensation is performed, followed by

[5] We found that a full-complexity CB-ESSFM based on (19)–(22) only provides a minimal performance improvement in the considered scenario. This improvement comes at a slightly higher computational cost, without any practical benefits. Therefore, we decided not to consider this implementation for the remainder of the paper.

a multiplexer (MUX) that recombines the subbands and two inverse CFFTs of size N.

The GVD compensation block consists in the multiplication of both polarizations of each signal component (at frequency f_k) by the corresponding value of the fiber transfer function $h_k = \exp(j2\pi^2\beta_2\Delta z f_k^2)$, where Δz is the fiber length for which GVD compensation is applied. In the configuration shown in Fig. 1(d), the link is divided into N_{st} steps of length L, and each step is symmetrically split into two halves of length $L/2$, so that $\Delta z = L/2$ in the first and last GVD compensation blocks, while $\Delta z = L$ in all the others. If necessary, the method could be easily adjusted to accommodate a variable step size and a different splitting ratio, with no significant changes to the implementation scheme and complexity. The benefits of an asymmetric configuration with an optimized splitting ratio are discussed in Section III-C.

The NLPR block follows the process shown in Fig. 4. First, the intensity (squared norm) of the N' 4D samples on each subband is computed. Then, MIMO filtering of the N_{sb} real intensity signals is implemented in frequency domain to compute the N_{sb} NLPRs. This operation, described in (16) in the time domain, is equivalently but more efficiently implemented in the frequency domain by performing N_{sb} real FFTs (RFFTs) of size N', multiplying the resulting vector of N_{sb} samples at each frequency component by the corresponding $N_{\text{sb}} \times N_{\text{sb}}$ MIMO transfer matrix, and then performing N_{sb} inverse RFFTs of size N' to go back to the time domain. The MIMO transfer matrix is obtained offline by transforming (through FFT) the MIMO impulse response matrix, whose elements correspond to the CB-ESSFM coefficients in (14). Finally, each 4D output sample is obtained by multiplying each polarization component by the corresponding complex-exponential term (the same for both polarizations).

B. Optimization of the CB-ESSFM Coefficients

The CB-ESSFM coefficients (or their FFT) needed for implementing the MIMO filter in Fig. 4 can be obtained analytically as explained in Section II-D, with minor modifications required to account for non-uniform span lengths or asymmetric splitting ratios.

Alternatively, they can also be found by numerical (offline) optimization, in order to maximize the system performance and/or consider unknown or complex system configurations. In the remainder of the paper, unless otherwise stated, we will use the latter approach, selecting the coefficients that minimize the mean square error between the transmitted and received symbols, using different data sets for optimization and performance evaluation. To reduce the complexity of the optimization (which is, however, done offline and does not affect the complexity of the DSP) we

i) assume that the coefficients are the same in each step, but for rescaling them proportionally to the signal power at the input of each step when more than one step per span is used;
ii) assume the symmetries in (24)–(26);
iii) consider a finite number of coefficients, which depends on the channel memory: the number of coefficients for the interaction of the subbands ℓ and $\ell + h$ is $2N_c(h) + 1 \approx \pi L\beta_2(nR/N_{\text{sb}})^2(h+1)$.

Given the above assumptions, letting $\mathbf{c}_h$ be the vector that collects the $2N_c(h) + 1$ coefficients that determine the interaction between subbands ℓ and $\ell + h$ (the same for any ℓ), the optimization problem consists in finding the vectors $\mathbf{c}_0, \mathbf{c}_1, \ldots, \mathbf{c}_{N_{\text{sb}}-1}$ that minimize the mean square error between the received symbols (after DBP, matched filtering, resampling, and mean phase rotation removal) and the transmitted symbols. To simplify the problem, we optimize the N_{sb} vectors separately, in N_{sb} iterations: in the ith iteration, we optimize $\mathbf{c}_{i-1}$, keeping $\mathbf{c}_0, \ldots, \mathbf{c}_{i-2}$ fixed at the values obtained at the previous iterations, and setting $\mathbf{c}_i, \ldots, \mathbf{c}_{N_{\text{sb}}-1-2}$ to zero. The optimization of $\mathbf{c}_{i-1}$ is performed by using Matlab's solver for nonlinear least squares problems based on the trust-region-reflective algorithm, approximating the gradient through finite differences. In our simulations, we did not observe substantial convergence issues nor a critical dependence of the solution on the initial values, which were simply set to the values corresponding to a standard SSFM (all zeros, but for the central coefficient of $\mathbf{c}_0$ equal to the average nonlinear phase rotation over the step).

The assumptions made above could be relaxed to improve the performance of the DBP algorithm. For instance, one could use a different set of CB-ESSFM coefficients in each nonlinear step. However, this change may lead to a more complex training phase. In this case, the use of automatic differentiation tools for the optimization of the coefficients might be beneficial, as they can efficiently compute gradients and potentially enhance the convergence of the training process. This possibility has not been investigated in this work and is left for future study.

C. Optimization of the Splitting Ratio

In the symmetric configuration shown in Fig. 1(a)–(d), the NLPR is positioned in the middle of the step, sandwiched between two equal GVD steps of length $L/2$. For a short piece of fiber with negligible attenuation, this configuration ensures high accuracy thanks to the cancellation of error terms with an odd symmetry around $z = 0$. This happens if the amount of nonlinearity and dispersion accumulated in the two half steps is comparable. However, the ESSFM is designed to reduce the number of propagation steps compared to the SSFM, with each step possibly encompassing one or more fiber spans, so that the power profile $g(z)$ (hence the accumulated nonlinearity) may be highly asymmetric with respect to the middle of the step. In this situation, an asymmetric configuration that divides the span in two portions with comparable accumulated nonlinearity is expected to yield better results. This even distribution of accumulated nonlinearity is analogous to the criterion used in [37] to derive the logarithmic step-size distribution for the SSFM.

Therefore, we consider a more general configuration in which the NLPR is preceded by a GVD block of length $(1-\rho)L$ and followed by a GVD block of length ρL, where ρ is referred to as splitting ratio. As an example, Fig. 5 shows the power profile $g(z)$ in a five-span (backward) link and, below, three different DBP configurations: (a) a single step for the whole link; (b) one step per span; and (c) two steps per span. The vertical arrows

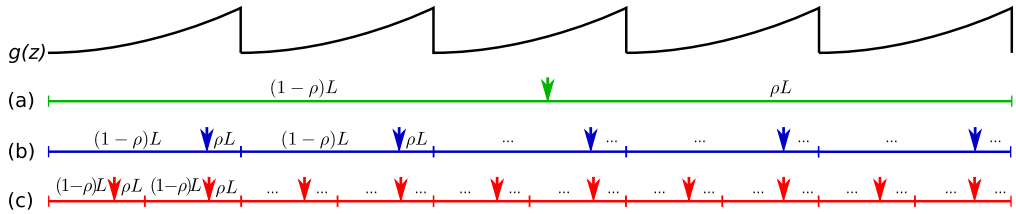

Fig. 5. Power profile $g(z)$ in the DBP link and three different step configurations: (a) 1 step/5 spans; (b) 1 step/span; (c) 2 steps/span. The optimal positions of the NLPRs steps are denoted by vertical arrows.

indicate the NLPR position within each step. With a single step, a balanced division of nonlinear effects is approximately obtained with a symmetric configuration ($\rho = 0.5$)—the approximation becoming more accurate for a higher number of spans. For instance, a symmetric configuration was shown to be optimal for the single-step DBP algorithm employed in [16]. On the other hand, with one step per span, nonlinear effects take place mostly in the last portion of each step, so that an asymmetric configuration with $\rho < 0.5$ results in a more balanced distribution of nonlinear effects in the two portions of the step. Finally, with two steps per span, a balanced distribution is still obtained with splitting ratio $\rho < 0.5$, but higher (approximately doubled) than in the previous case. Further increasing the number of steps makes the attenuation in each step less and less relevant, requiring the optimal ρ to approach again the symmetric configuration (as typically done in the conventional SSFM). These heuristic considerations will be verified numerically in Section IV-C.

With asymmetric steps, the overall DBP implementation changes only slightly with respect to the symmetric case shown in Fig. 1(d). In fact, couple of GVD blocks from adjacent steps can still be combined to form a single GVD block of length L, and only the length of the first and last GVD blocks must be modified to $(1-\rho)L$ and ρL, respectively.

D. Complexity

The computational complexity of the algorithm is assessed based on the number of real multiplications (RMs) and real additions (RAs) required for each 2D symbol. In our analysis, we assume that each complex addition is carried out with two RAs, while each complex multiplication (CM) requires three RMs and five RAs [38]. If one of the multipliers of a CM is known in advance, only three out of the five RAs need to be computed in real-time, as two of them can be precomputed offline. Analogously, if a couple of CMs involve the same complex multiplier, two RAs can be computed only once and used for both the CMs, so that each CM requires, on average, three RMs and four RAs. Finally, if one of the multipliers is real, the CM reduces to just two RMs. Moreover, we assume that the CFFT of N complex samples is implemented by the split-radix algorithm [39], which requires $N\log_2 N - 3N + 4$ RMs and $3N\log_2 N - 3N + 4$ RAs, and that the RFFT requires approximately half the RMs and RAs of the CFFT [39].[6]

At the output of the CB-ESSFM, each 4D output sample corresponds to a couple of 2D symbols—one per polarization. Thus, the overall computational complexity, is given by the following equations

$$C_M = \frac{n \cdot N_{\mathrm{RM}}(N, N_{\mathrm{st}}, N_{\mathrm{sb}})}{2 \cdot (N - N_{\mathrm{ov}})} \qquad \text{RM/2D symb.} \qquad (27)$$

$$C_A = \frac{n \cdot N_{\mathrm{RA}}(N, N_{\mathrm{st}}, N_{\mathrm{sb}})}{2 \cdot (N - N_{\mathrm{ov}})} \qquad \text{RA/2D symb.} \qquad (28)$$

where N_{RM} and N_{RA} denote, respectively, the number of RMs and RAs required by the CB-ESSFM to process a block of N 4D samples, and we have made explicit their dependence on N and on the number of steps N_{st} and of subbands N_{sb} employed by the algorithm.

The expression for N_{RM} and N_{RA} is derived here. The CB-ESSFM in Fig. 3 applied to a block of N samples, with N_{sb} bands, and N_{st} steps uses 4 (direct or inverse) CFFTs of size N, $4N_{\mathrm{sb}}N_{\mathrm{st}}$ CFFTs of size $N' = N/N_{\mathrm{sb}}$, $N_{\mathrm{st}} + 1$ GVD compensations, and N_{st} NLPRs. Concerning GVD compensation, it requires only $2N$ CMs—two per each polarization of each 4D sample—with a reduced complexity due to the fixed multiplier, and because the factor h_k is fixed and can be precomputed offline. This corresponds to $6N$ RMs and $6N$ RAs for each GVD compensation step.

Regarding the cost of each NLPR, the computation of the intensity of the N' 4D samples uses $4N$ RMs and $3N$ RAs. For what concerns the MIMO filtering, by exploiting the Hermitian symmetry of the N' frequency components (both the signal intensity and the CB-ESSFM coefficients are real in time domain), only half ($N'/2$) matrix–vector multiplications need to be actually implemented. Considering that the elements of the MIMO transfer matrix can be precomputed offline, and that the diagonal elements are real due to the even symmetry of the SPM coefficients in (26), each matrix–vector multiplication requires $N_{\mathrm{sb}}(N_{\mathrm{sb}} - 1)$ CMs with one fixed multiplier, N_{sb} CMs with one real multiplier, and $N_{\mathrm{sb}}(N_{\mathrm{sb}} - 1)$ complex additions. Overall, MIMO filtering requires

$$N\log_2 N' + \frac{N}{2}(3N_{\mathrm{sb}} - 7) + 4N_{\mathrm{sb}} \qquad \text{RM}$$

$$3N\log_2 N' + \frac{N}{2}(5N_{\mathrm{sb}} - 11) + 4N_{\mathrm{sb}} \qquad \text{RA}$$

[6]These assumptions differ from those in [23], where we considered a naive CM implementation with four RMs and two RAs, and the classical Cooley–Tukey FFT algorithm [40]. The present choice is slightly more efficient, particularly when RMs are considered more expensive than RAs.

For the computation of the complex exponential terms, we assume a look-up-table implementation or similar approach, so we neglect its complexity. The multiplication by the complex-exponential term requires a couple of CMs with a common multiplier, resulting in $6N$ RMs and $8N$ RAs to obtain the N 4D samples (N' on each subband). Overall, by combining the cost of intensity computation, MIMO filtering, and output CMs, the NLPR step in Fig. 4 requires

$$N\log_2 N' + \frac{N}{2}(3N_{\mathrm{sb}} + 13) + 4N_{\mathrm{sb}} \qquad \text{RM}$$

$$3N\log_2 N' + \frac{N}{2}(5N_{\mathrm{sb}} + 11) + 4N_{\mathrm{sb}} \qquad \text{RA}$$

In total, adding up the number of RMs and RAs required by all the steps of the CB-ESSFM algorithm in Fig. 2, and replacing them in (27) and (28), we obtain the following equations for the computational complexity

$$C_M = \frac{n}{2}\frac{N}{N-N_{\mathrm{ov}}}\left((5N_{\mathrm{st}}+4)\log_2\frac{N}{N_{\mathrm{sb}}} + N_{\mathrm{st}}\frac{3N_{\mathrm{sb}}+1}{2} + 4\log_2 N_{\mathrm{sb}} - 6 + \frac{20N_{\mathrm{sb}}N_{\mathrm{st}}+16}{N}\right) \quad \text{RM/2D symb.} \tag{29}$$

$$C_A = \frac{n}{2}\frac{N}{N-N_{\mathrm{ov}}}\left((15N_{\mathrm{st}}+12)\log_2\frac{N}{N_{\mathrm{sb}}} + N_{\mathrm{st}}\frac{5N_{\mathrm{sb}}-1}{2} + 12\log_2 N_{\mathrm{sb}} - 6 + \frac{20N_{\mathrm{sb}}N_{\mathrm{st}}+16}{N}\right) \quad \text{RA/2D symb.} \tag{30}$$

For the sake of comparison, we report below the complexity of the other DBP methods discussed in this work. The (single-band) ESSFM can be obtained simply by considering a CB-ESSFM with a single band (frequency-domain NLPR), or can be alternatively implemented with a time-domain NLPR, with a complexity that depends on the number of real symmetric CB-ESSFM coefficients $2N_c + 1$ as

$$C_M = \frac{n}{2}\frac{N}{N-N_{\mathrm{ov}}}\left((N_{\mathrm{st}}+1)\left(4\log_2 N - 6 + \frac{16}{N}\right) + N_{\mathrm{st}}(11+N_c)\right) \quad \text{RM/2D symb.} \tag{31}$$

$$C_A = \frac{n}{2}\frac{N}{N-N_{\mathrm{ov}}}\left((N_{\mathrm{st}}+1)\left(12\log_2 N - 6 + \frac{16}{N}\right) + N_{\mathrm{st}}(11+2N_c)\right) \quad \text{RA/2D symb.} \tag{32}$$

The complexity of the SSFM and the optimized SSFM (OSSFM), corresponding to a standard SSFM with optimized nonlinear coefficient, is obtained by setting $N_c = 0$ in (31) and (32), while the complexity of GVD compensation is obtained by setting $N_{\mathrm{st}} = 0$.

IV. Numerical Results

A. System Description

The link consists of 15 (unless otherwise stated) spans of 80 km single-mode fiber (SMF), with attenuation $\alpha_{\mathrm{dB}} =$ 0.2 dB/km, dispersion $D = 17\,\mathrm{ps/nm/km}$, and Kerr parameter $\gamma = 1.27\,\mathrm{W^{-1}km^{-1}}$. After each span, an erbium-doped fiber amplifier (EDFA) with a noise figure of 4.5 dB compensates for the span loss. The transmitted signal is made of five identical WDM channels with 100 GHz spacing. Each channel carries a uniform dual-polarization 64-QAM signal, with baud rate $R = 93$ GBd and (frequency-domain) root-raised-cosine pulses with rolloff $r = 0.05$. At the receiver side, the central channel is demultiplexed and processed by either electronic dispersion compensation (EDC) or digital backpropagation (DBP) with different possible algorithms: the CB-ESSFM described in this paper; the ESSFM proposed in [18], [19], equivalent to CB-ESSFM with $N_{\mathrm{sb}} = 1$ and $\rho = 0.5$; the OSSFM, corresponding to a standard SSFM with optimized nonlinear coefficient and equivalent to an ESSFM with a single coefficient ($N_c = 0$); and ideal DBP, practically obtained from a standard SSFM with a very large number of steps N_{st} ($N_{\mathrm{st}} \to \infty$), sufficient to achieve performance saturation. In the overlap-and-save algorithm, the number of overlapping samples is set to match the overall channel memory ($N_{\mathrm{ov}} \approx 1800$ for the 15×80 km link), while the blocklength is selected as the power of two that minimizes computational complexity ($N = 16384$ for the 15×80 km link). All DBP algorithms use a uniform step size (N_{st} steps of the same length) and, unless otherwise stated, $n = 1.125$ samples/symbol. Next, the signal undergoes matched filtering, resampling at 1 sample/symbol, and mean phase rotation removal. Finally, the performance is evaluated in terms of received signal-to-noise ratio (SNR), where noise is defined as the difference between the output samples and the transmitted symbols. The results shown in the following are all obtained at the optimal launch power (approximately between 3 and 4 dBm per channel).

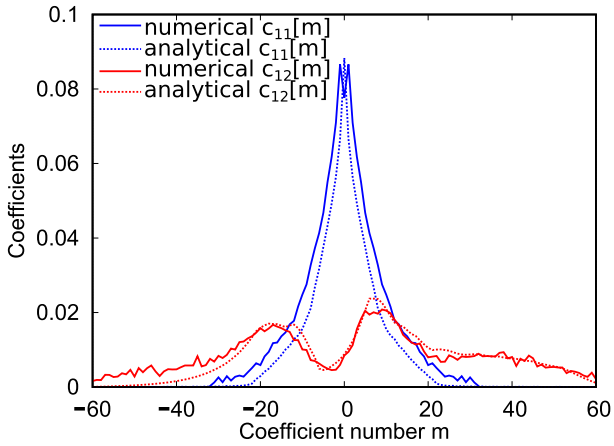

Fig. 6. Comparison of intraband coefficients $c_{11}[m]$ and interband coefficients $c_{12}[m]$ obtained analytically or by numerical optimization for the case of $N_{\mathrm{st}} = 5$ steps.

B. CB-ESSFM Coefficients: Analytical Evaluation vs Numerical Optimization

Fig. 6 compares the CB-ESSFM coefficients obtained analytically as explained in Section II-D (dashed) with those obtained by the numerical optimization procedure described

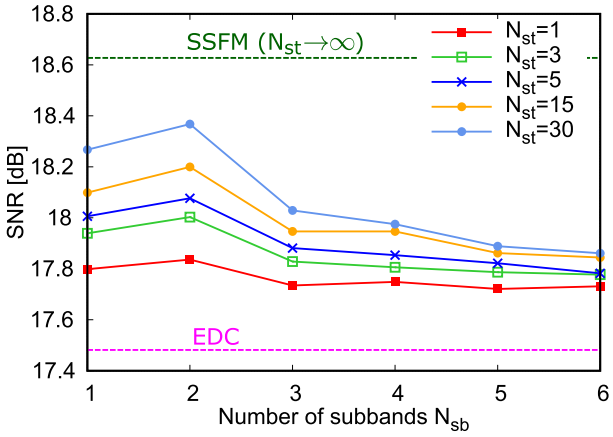

Fig. 7. SNR versus number of subbands N_{sb} for the symmetric CB-ESSFM ($\rho = 0.5$) and various numbers of steps N_{st}.

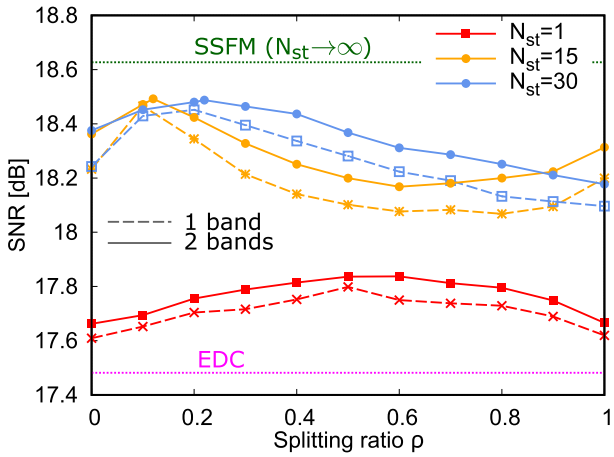

Fig. 8. SNR versus splitting ratio ρ for the CB-ESSFM with $N_{\mathrm{sb}} = 1$ band (dashed) and $N_{\mathrm{sb}} = 2$ bands (solid) and various numbers of steps N_{st}.

in Section III-B (solid). The coefficients are evaluated for the 15×80 km link, considering $N_{\mathrm{st}} = 5$ identical DBP steps (each made of three fiber spans) and $N_{\mathrm{sb}} = 2$ bands, and their values are normalized to the average nonlinear phase rotation over the step. The impulse response determined by the intraband coefficients $c_{11}[m]$ (blue) is symmetric and limited to approximately 50 samples. In contrast, the impulse determined by the interband coefficients $c_{12}[m]$ (red) extends to about 120 samples due to walk-off between the two bands, and exhibits an irregular asymmetric shape caused by the combined effect of walk-off and periodic attenuation/amplification. In both cases, there is good agreement between the analytical and numerical curves, validating the accuracy of the theoretical model developed in Section II and supporting the use of the analytical procedure for setting the algorithm coefficients or as an initialization for numerical optimization. In the next sections, we will always use the coefficients obtained by numerical optimization.

C. System Performance

We start by investigating how the division in subbands and the optimization of the splitting ratio affect the performance of the proposed algorithm. Fig. 7 shows the SNR obtained with the symmetric CB-ESSFM ($\rho = 0.5$) as a function of the number of subbands N_{sb} for different numbers of steps N_{st}. As a reference, the performance of EDC and ideal DBP are also shown. As expected, the SNR improves when N_{st} increases. On the other hand, as discussed in Section II-B, increasing N_{sb} causes two opposite effects: the reduction of dispersion within each subband, which improves model accuracy; and the appearance of additional interband nonlinear interactions not included in the model (FWM and XPolM), which degrades accuracy. In the current scenario, regardless of the number of steps, the best tradeoff between the two effects is obtained with $N_{\mathrm{sb}} = 2$ subbands, with a gain of 0.1 dB with respect to the single-band case for $N_{\mathrm{st}} = 15$, and a total gain of 0.72 dB with respect to EDC. On the other hand, we expect that the optimal number of subbands depends on the signal bandwidth and accumulated dispersion, meaning that more subbands may be needed for higher baud rates, while a single band might be optimal for low-dispersion links (e.g., in the O-band). The investigation of different scenarios is left for future study.

Next, Fig. 8 shows the SNR as a function of the splitting ratio ρ for different numbers of steps N_{st} and considering either a single band or two subbands. First, the figure confirm that the two-subband implementation always outperforms the single-band one, even when considering asymmetric steps with different splitting ratios. Moreover, it shows that optimizing the splitting ratio may be even more important in some cases. As conjectured in Section III-C, the best performance is obtained when the step is split into two parts with comparable nonlinear effects, resulting in more accurate error cancellation. In fact, the symmetric configuration ($\rho = 0.5$) is nearly optimal when a single step is employed, but highly suboptimal for $N_{\mathrm{st}} = 15$ (one step/span). In the latter case, the optimal splitting ratio is $\rho \approx 0.12$ (70 km/10 km splitting of the 80 km step) and yields a gain of almost 0.3 dB with respect to the symmetric configuration, with a total gain of about 1 dB compared to EDC. In this case, $\rho = 0$ and $\rho = 1$ are equivalent, as both place the nonlinear step at the beginning of the span, resulting in identical performance. For $N_{\mathrm{st}} = 30$ (two steps/span), steps are shorter (40 km). However, since nonlinear effects occur mostly toward the end of the (backward) step, the optimal length ρL of the second part of the step is only slightly reduced compared to the 80 km step, resulting in an almost doubled optimal splitting ratio $\rho \approx 0.22$ (31 km/9 km splitting) and a lower gain of 0.12 dB compared to the symmetric configuration. In this case, $\rho = 0$ and $\rho = 1$ result in different nonlinear step configurations, with the former placing the nonlinear step where the signal power is higher, thus providing better performance. Finally, we note that with an optimized splitting ratio, increasing the number of steps from $N_{\mathrm{st}} = 15$ to 30 does not provide any improvement. We believe that this is due to two main reasons: the performance approaching that of ideal DBP, and the negligible amount of nonlinearity handled by odd steps compared to even steps when using two equal steps per span. We expect that the performance for $N_{\mathrm{st}} = 30$ could be slightly improved by increasing the length of odd steps compared to that of even steps. Indeed, optimizing the step-size distribution may be beneficial in some scenarios

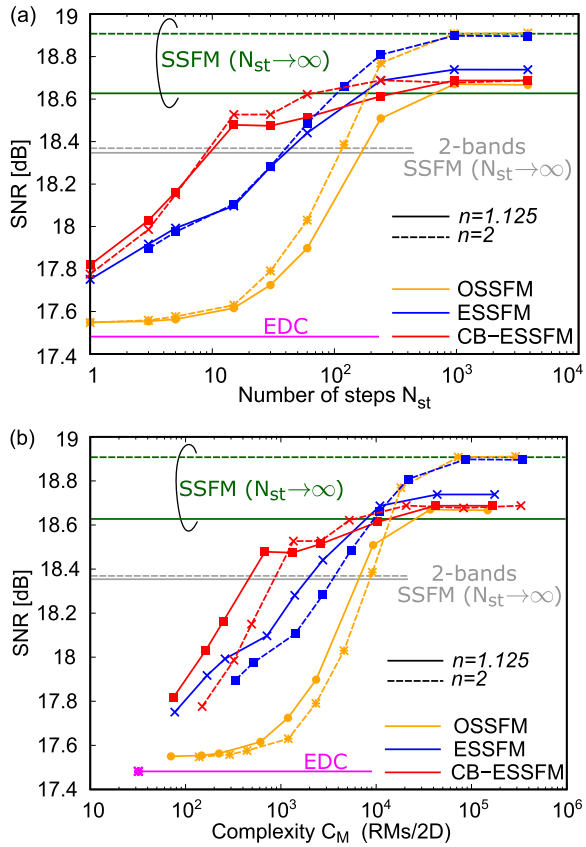

Fig. 9. SNR for different DBP techniques and oversampling factors versus (a) number of steps and (b) computational complexity.

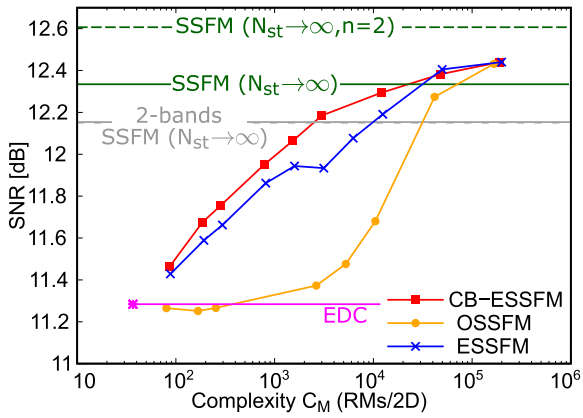

Fig. 10. SNR versus computational complexity for the $60 \times 80\,km$ link and different DBP techniques.

(e.g., when multiple steps per span are used), particularly when combined with the use of different CB-ESSFM coefficients in each step. This additional optimization is left for future study.

Next, we investigate the relation between performance and complexity of the proposed algorithm, and how it compares to other DBP methods. Based on the analysis above, all the CB-ESSFM results shown in the following are obtained for $N_{\mathrm{sb}} = 2$ and an optimized splitting ratio. Fig. 9(a) shows the performance of different DBP techniques as a function of the number of steps N_{st}. The solid curves are obtained using $n = 1.125$ samples/symbol, while the dashed ones are obtained with $n = 2$ samples/symbol. Horizontal lines indicating the performance of EDC and ideal DBP (with the two oversampling factors, and for 1 and 2 bands) are also shown as a reference. For a small number of steps, CB-ESSFM outperforms all the other algorithms, approaching its best performance already with $N_{\mathrm{st}} = 15$ steps. On the other hand, OSSFM gives almost no gain up to 15 steps, needs hundreds of steps to achieve the performance of CB-ESSFM, and saturates to its ultimate best performance for about a thousand steps. Such large N_{st} values, though not feasible for a practical implementation, give some additional insight on the behavior of the algorithms. For instance, we see that OSSFM and ESSFM saturate to a higher performance than CB-ESSFM, but only if a significant oversampling ($n = 2$) is employed. This is expected, since the use of subband processing in CB-ESSFM entails neglecting some interband nonlinear interactions, whose impact becomes relevant only when all the other effects have been fully compensated by the algorithm. In this case, also oversampling becomes important for OSSFM, ESSFM, and the ideal single-band DBP curves when N_{st} is large, but not for CB-ESSFM and two-bands DBP curves, which are limited by interband nonlinearities. On the other hand, CB-ESSFM performs significantly better than an ideal single-band DBP applied independently to two subbands (shown with grey lines), confirming that the interband XPM terms (neglected by the former) are accurately accounted for by the coupled-band mechanism of CB-ESSFM.

Next, Fig. 9(b) reports the results in Fig. 9(a) as a function of the computational complexity, defined as the number of required real multiplications per 2D symbol and computed according to (29) and (31), as discussed in Section III-D. EDC is performed in a single step and has a fixed complexity, represented by a single purple square symbol in the figure. Additionally, a horizontal purple line at the same SNR value is shown as a reference. Similarly, horizontal lines indicate the SNR achievable by ideal DBP algorithms; their complexity is considered infinite and, therefore, is not represented with symbols. Fig. 9(b) confirms and quantifies the observations from Fig. 9(a), showing that, for reasonable levels of complexity ($< 10^3$ RM/2D), the CB-ESSFM with $n = 1.125$ outperforms the other DBP techniques. Specifically, the CB-ESSFM achieves a 0.34 dB gain at 75 RM/2D, and a 1 dB gain at 681 RM/2D compared to EDC, which requires 32 RM/2D. Fig. 10 shows the performance versus complexity for the same configuration as in Fig. 9(b), but with 60 fiber spans, resulting in an overall link length of 4800 km. The observed behavior is similar to that of the 1200 km link, though the required complexity is generally higher; for instance, 10^6 RM/2D are not sufficient to reach saturation for both ESSFM and OSSFM. The largest gain of CB-ESSFM over ESSFM is again achieved with one step per span, which in this case corresponds to $N_{\mathrm{st}} = 60$ or approximately 3000 RM/2D. We expect similar results even when a larger number of WDM channels is used, possibly with a slightly smaller gain due to increased interchannel nonlinearity. Conversely, when the bandwidth of each channel is larger, the impact of intrachannel nonlinearity would be greater, allowing higher gains, albeit with

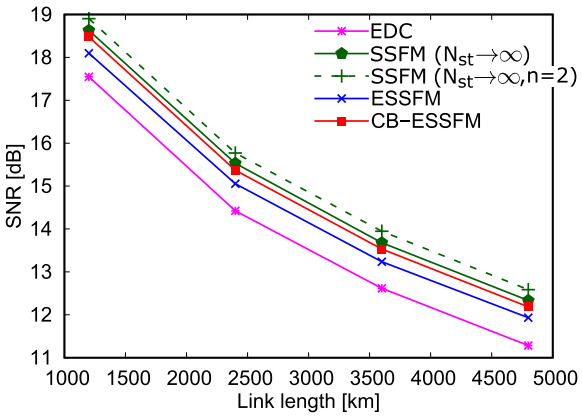

Fig. 11. SNR versus link length for EDC and various DBP techniques: ESSFM with 1 step/span; CB-ESSFM with 1 step/span, $N_{\text{sb}} = 2$ bands, and optimized ρ; ideal DBP (SSFM with $N_{\text{st}} \to \infty$) with different oversampling factors ($n = 1.125$ except otherwise specified).

possibly increased complexity. In this case, the optimal number of subbands may be larger.

Finally, Fig. 11 compares the performance of different DBP techniques as a function of link length, considering 15, 30, 45, and 60 spans of 80 km each. The figure shows the SNR achieved with EDC, ideal DBP with $n = 1.125$ or $n = 2$ samples/symbol, the symmetric ESSFM (i.e., with $\rho = 0.5$), and the CB-ESSFM with $N_{\text{sb}} = 2$ subbands and an optimized splitting ratio. Both ESSFM and CB-ESSFM are implemented with 1 step per span (i.e., 15, 30, 45, and 60 steps for the respective lengths) and $n = 1.125$. The results indicate that the relative performance of the methods and the gain provided by CB-ESSFM are consistently maintained across different link lengths when using one step per span.

V. Conclusion

The proposed coupled-band enhanced split-step Fourier method (CB-ESSFM) offers a promising approach to achieving a low-complexity, high-accuracy trade-off in digital backpropagation (DBP) for fiber nonlinearity mitigation. The approach combines an SSFM-like structure with a simplified logarithmic perturbation to achieve a good accuracy with a reduced number of steps. The use of subband processing and asymmetric steps with an optimized splitting ratio enables a further reduction in the required number of steps to achieve a desired performance.

The theoretical foundation for CB-ESSFM was developed by deriving a simplified logarithmic-perturbation model for dual-polarization multiband signals in optical fibers. This model provides a basis for both the CB-ESSFM design and for the analytical computation of its coefficients. Subsequently, a digital signal processing algorithm was devised, based on a discrete-time version of the model and an overlap-and-save strategy. Practical methods for optimizing the algorithm' s coefficients and the asymmetric step splitting ratio were introduced, with a comprehensive analysis of the computational complexity.

Through numerical simulations, the performance and complexity of CB-ESSFM were shown to be superior to existing DBP methods. In a 5×93 GBd WDM system over a 15-span (80 km each) single-mode fiber link, a 15-step CB-ESSFM, requiring 681 real multiplications per 2D symbol (RM/2D), demonstrated a 1 dB gain over simple dispersion compensation, with an improvement (for the same complexity) of 0.4 dB compared to the previously proposed ESSFM method, and 0.9 dB compared to a conventional optimized SSFM. Notable gains were also achieved at lower complexity levels; for example, gains of 0.34 dB, 0.55 dB, and 0.7 dB were obtained with 75, 162, and 248 RM/2D, respectively.

The performance trends observed for shorter links were consistent in longer links, though with a nearly proportional increase of complexity, confirming the robustness of CB-ESSFM across varying transmission distances. These results highlight the potential of CB-ESSFM as an efficient DBP approach for fiber nonlinearity compensation, making it an attractive candidate for practical deployment in long-haul optical networks where minimizing complexity without compromising performance is essential. Future work to further highlight and enhance the potential of CB-ESSFM include (i) the investigation of different scenarios (e.g., higher baud rates, single-span links, O-band); (ii) the joint optimization of the step-size distribution, splitting ratio, and of different CB-ESSFM coefficients for each step; and (iii) the use of automatic differentiation tools to enhance the convergence of the training process for such more complex optimization problems.

Appendix A

According to the FRLP model, the propagation through the first half of the step in Fig. 1(a) can be approximated as [32, eq. (7)],

$$u'(t) \approx \int_{-\infty}^{+\infty} U_{\text{in}}(f) H(L/2, f) e^{-j\theta'(f,t)} e^{j2\pi f t} df \tag{33}$$

where $U_{\text{in}}(f)$ is the Fourier transform of $u_{\text{in}}(t)$, $H(z, f)$ is given in (5), and $\theta'(f, t)$ is a frequency- and time-dependent nonlinear phase rotation (NLPR) that accounts for SPM and its interaction with dispersion and attenuation [32, eq. (8)]. To further simplify the model, we neglect the frequency dependence of the NLPR and replace it with its value in the middle of the signal bandwidth, $\theta'(t) \triangleq \theta'(0, t)$.[7] By letting $f = 0$ in [32, eq. (8)], representing the input signal $U_{\text{in}}(f)$ as a function of the linearly propagated signal $U(f) = U_{\text{in}}(f)H(L/2, f)$, the NLPR can be expressed as

$$\theta'(t) = P \iint_{\mathbb{R}^2} K'(\mu, \nu) U(\mu) U^*(\nu) e^{j2\pi(\mu-\nu)t} \mathrm{d}\mu \mathrm{d}\nu \tag{34}$$

where

$$K'(\mu, \nu) = \int_{-L/2}^{0} \gamma g(z) H(z, \mu) H^*(z, \nu) H^*(z, \mu - \nu) dz \tag{35}$$

is the kernel function, properly redefined with respect to [32, eq. (10)] to account for the additional approximation ($f = 0$) and different normalizations used here. With this approximation,

[7] A more accurate approximation could be obtained by replacing $\theta'(f, t)$ with its average or effective value over the signal bandwidth.

(33) can be rewritten as

$$u(t) = \int_{-\infty}^{+\infty} U_{\text{in}}(f) H(L/2, f) e^{j2\pi f t} df \tag{36}$$

$$u'(t) = e^{-j\theta'(t)} u(t) \tag{37}$$

which can be interpreted as the cascade of the GVD and NLPR accumulated over the first half of the step, and represented by the first two blocks in Fig. 1(b). The propagation through the second half of the step in Fig. 1(a) is then approximated by a reverse-order FRLP model (with the NLPR preceding the GVD). This is obtained by using the same approximated FRLP model derived above to describe the backward propagation from $L/2$ to 0, obtaining

$$u''(t) = \int_{-\infty}^{+\infty} U_{\text{out}}(f) H(-L/2, f) e^{j2\pi f t} df \tag{38}$$

$$u'(t) = e^{-j\theta''(t)} u''(t) \tag{39}$$

where the NLPR is

$$\begin{aligned}\theta''(t) &= P \iint_{\mathbb{R}^2} K''(\mu,\nu) U''(\mu) U''^{*}(\nu) e^{j2\pi(\mu-\nu)t} \mathrm{d}\mu \mathrm{d}\nu \\ &\approx P \iint_{\mathbb{R}^2} K''(\mu,\nu) U(\mu) U^{*}(\nu) e^{j2\pi(\mu-\nu)t} \mathrm{d}\mu \mathrm{d}\nu \end{aligned} \tag{40}$$

with kernel function

$$K''(\mu,\nu) = \int_{L/2}^{0} \gamma g(z) H(z,\mu) H^{*}(z,\nu) H^{*}(z,\mu-\nu) dz \tag{41}$$

The approximation in (40) consists in replacing the potential $|u|^2$ that appears in the nonlinear term of the NLSE (1) and drives the generation of the NLPR with the intensity of the linearly propagated signal. This is the same approximation employed in [31], [32] to derive the original FRLP model in (33), so it is not expected to further reduce the accuracy of the final model. The forward propagation model for the second half of the step is obtained by inverting (38) and (39), obtaining

$$u''(t) = e^{j\theta''(t)} u'(t) \tag{42}$$

$$u_{\text{out}}(t) = \int_{-\infty}^{+\infty} U''(f) H(L/2, f) e^{j2\pi f t} df \tag{43}$$

corresponding, respectively, to the second NLPR and GVD blocks in Fig. 1(b). The model can be further simplified by combining the two adjacent NLPR blocks in Fig. 1(b) into a single block with overall NLPR $\theta(t) = \theta'(t) - \theta''(t)$, as shown in Fig. 1(c), resulting in the overallpropagation model (2)–(4), with kernel function $K(\mu,\nu) = K'(\mu,\nu) - K''(\mu,\nu)$ given by (7).

Appendix B

We assume that $u(t)$ is band-limited, so that it can be represented as

$$u(t) = \sum_{k=-\infty}^{\infty} u[k] \text{sinc}\left(Rt - k\right) \tag{44}$$

where $u[k] = u(k/R)$ are its samples taken at sufficiently high rate $R = 1/T$. By taking the Fourier transform of (44)

$$U(f) = \frac{1}{R} \sum_{k=-\infty}^{\infty} u[k] \text{rect}\left(f/R\right) \exp(-j2\pi k f/R) \tag{45}$$

and replacing it in (6), we can express the NLPR samples as

$$\begin{aligned}\theta[k] &= \theta(k/R) \\ &= (P/R^2) \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} u[m] u^{*}[n] \\ &\quad \times \int_{-R/2}^{R/2} \int_{R/2}^{R/2} K(\mu,\nu) e^{j2\pi[(k-m)\mu-(k-n)\nu)/R} d\mu d\nu \\ &= \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} d[k-m, k-n] u[m] u^{*}[n] \\ &= \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} d[m,n] u[k-m] u^{*}[k-n] \end{aligned} \tag{46}$$

where

$$d[m,n] = \frac{P}{R^2} \int_{-R/2}^{R/2} \int_{-R/2}^{R/2} K(\mu,\nu) e^{j2\pi(m\mu-n\nu)/R} d\mu d\nu \tag{47}$$

are the coefficients of the discrete-time second-order Volterra kernel, corresponding to the samples of the inverse two-dimensional Fourier transform of the frequency-domain Volterra kernel (7) over a bandwidth equal to the sampling rate. After truncating the discrete-time Volterra kernel and neglecting its off-diagonal terms as discussed in Section II-A, we finally obtain the NLPR expression in (8) with the CB-ESSFM coefficients in (9).

In the multiband case, the derivation of the discrete-time model proceeds analogously. After dividing the signal $u(t)$ into N_{sb} subbands as in (10), the generic ith subband is represented as

$$u_i(t) = \sum_{k=-\infty}^{\infty} u_i[k] \text{sinc}\left(R't - k\right) \exp\left(j2\pi f_i t\right) \tag{48}$$

where $u_i[k] = u(k/R') \exp(-j2\pi f_i k/R')$ is the kth sample of the lowpass equivalent representation of the signal in the ith subband, taken at rate $R' = R/N_{\text{sb}}$. By replacing the Fourier transform of (48) in (12), truncating the discrete-time Volterra kernel and neglecting its off-diagonal terms as in the single-band case, we finally obtain the NLPR expression in (13) with the CB-ESSFM coefficients in (14). The derivations of the dual-polarization discrete-time models (16) or (21)–(22) are very similar and are omitted.

Appendix C

By replacing (5) in (7), and defining

$$b \triangleq 2\pi^2 \beta_2 \nu(\mu - \nu) \tag{49}$$

the kernel function can be rewritten as

$$K(\mu,\nu) = \int_{-L/2}^{L/2} \gamma g(z) \exp(-j2bz) dz \tag{50}$$

For N_{sp} even, letting $\zeta = z - \lfloor z/L_{\text{sp}} \rfloor \cdot L_{\text{sp}}$, so that $g(z) = e^{-\alpha\zeta}$, we can rewrite (50) as

$$\sum_{n=-N_{\text{sp}}/2}^{N_{\text{sp}}/2-1} \exp(-j2bnL_{\text{sp}}) \int_0^{L_{\text{sp}}} \gamma \exp[-(\alpha + j2b)\zeta] d\zeta$$

$$= \gamma \frac{1 - e^{-(\alpha+j2b)L_{\text{sp}}}}{\alpha + j2b} \sum_{n=-N_{\text{sp}}/2}^{N_{\text{sp}}/2-1} \exp(-j2bnL_{\text{sp}})$$

$$= \gamma \frac{1 - e^{-(\alpha+j2b)L_{\text{sp}}}}{\alpha + j2b} \exp(jbN_{\text{sp}}L_{\text{sp}}) \frac{1 - \exp(-j2bN_{\text{sp}}L_{\text{sp}})}{1 - \exp(-j2bL_{\text{sp}})} \tag{51}$$

Eventually, after a few passages, we obtain the expression in (23) for the kernel function. The same result is obtained, with similar calculations, when N_{sp} is odd.

References

[1] S. Civelli, D. P. Jana, E. Forestieri, and M. Secondini, "Coupled-band ESSFM for low-complexity DBP," in *Proc. Eur. Conf. Opt. Commun.*, 2024.

[2] R. -J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity limits of optical fiber networks," *J. Lightw. Technol.*, vol. 28, no. 4, pp. 662–701, Feb. 2010.

[3] M. Secondini, "Chapter 20 - Information capacity of optical channels," in *Optical Fiber Telecommunications VII*, A. E. Willner, Ed. Norwell, MA, USA: Academic Press, 2020, pp. 867–920.

[4] P. J. Winzer, "Scaling optical fiber networks: Challenges and solutions," *Opt. Photon. News*, vol. 26, no. 3, pp. 28–35, 2015.

[5] E. Agrell et al., "Roadmap of optical communications," *J. Opt.*, vol. 18, no. 6, 2016, Art. no. 063002.

[6] M. Secondini, E. Agrell, E. Forestieri, D. Marsella, and M. R. Camara, "Nonlinearity mitigation in WDM systems: Models, strategies, and achievable rates," *J. Lightw. Technol.*, vol. 37, no. 10, pp. 2270–2283, May 2019.

[7] G. P. Agrawal, *Nonlinear Fiber Optics*, 5th ed. Norwell, MA, USA: Academic Press, 2013.

[8] A. Bononi, R. Dar, M. Secondini, P. Serena, and P. Poggiolini, "Fiber nonlinearity and optical system performance," in *Springer Handbook of Optical Networks*, Cham, Switzerland: Springer, 2020, pp. 287–351.

[9] R. -J. Essiambre and P. J. Winzer, "Fibre nonlinearities in electronically pre-distorted transmission," in *Proc. 31st Eur. Conf. Opt. Commun.*, 2005, vol. 2, pp. 191–192.

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

[11] F. P. Guiomar and A. N. Pinto, "Simplified Volterra series nonlinear equalizer for polarization-multiplexed coherent optical systems," *J. Lightw. Technol.*, vol. 31, no. 23, pp. 3879–3891, Dec. 2013.

[12] T. Koike-Akino et al., "High-order statistical equalizer for nonlinearity compensation in dispersion-managed coherent optical communications," *Opt. Exp.*, vol. 20, no. 14, pp. 15769–15780, 2012.

[13] D. Marsella, M. Secondini, and E. Forestieri, "Maximum likelihood sequence detection for mitigating nonlinear effects," *J. Lightw. Technol.*, vol. 32, no. 5, pp. 908–916, Mar. 2014.

[14] A. Hasegawa and F. Tappert, "Transmission of stationary nonlinear optical pulses in dispersive dielectric fibers. I. anomalous dispersion," *Appl. Phys. Lett.*, vol. 23, pp. 142–144, Aug. 1973.

[15] L. B. Du and A. J. Lowery, "Improved single channel backpropagation for intra-channel fiber nonlinearity compensation in long-haul optical communication systems," *Opt. Exp.*, vol. 18, no. 16, pp. 17075–17088, Jun. 2010.

[16] D. Rafique, M. Mussolin, M. Forzati, J. Mårtensson, M. N. Chugtai, and A. D. Ellis, "Compensation of intra-channel nonlinear fibre impairments using simplified digital back-propagation algorithm," *Opt. Exp.*, vol. 19, no. 10, pp. 9453–9460, 2011.

[17] L. Li et al., "Implementation efficient nonlinear equalizer based on correlated digital backpropagation," presented at Opt. Fiber Commun. Conf./Nat. Fiber Optic Engineers Conf., Los Angeles, CA, USA, Mar. 6–10, 2011, Paper OWW3.

[18] M. Secondini, D. Marsella, and E. Forestieri, "Enhanced split-step Fourier method for digital backpropagation," in *Proc. Eur. Conf. Opt. Commun.*, 2014, pp. 1–3.

[19] M. Secondini, S. Rommel, G. Meloni, F. Fresi, E. Forestieri, and L. Poti, "Single-step digital backpropagation for nonlinearity mitigation," *Photonic Netw. Commun.*, vol. 31, pp. 493–502, Jun. 2016.

[20] J. Leibrich and W. Rosenkranz, "Efficient numerical simulation of multichannel WDM transmission systems limited by XPM," *IEEE Photon. Technol. Lett.*, vol. 15, no. 3, pp. 395–397, Mar. 2003.

[21] E. F. Mateo, F. Yaman, and G. Li, "Efficient compensation of inter-channel nonlinear effects via digital backward propagation in WDM optical transmission," *Opt. Exp.*, vol. 18, no. 14, pp. 15144–15154, 2010.

[22] S. Civelli, E. Forestieri, A. Lotsmanov, D. Razdoburdin, and M. Secondini, "Coupled-channel enhanced SSFM for digital backpropagation in WDM systems," in *Proc. Opt. Fiber Commun. Conf.*, 2021, pp. 1–3.

[23] S. Civelli, E. Forestieri, A. Lotsmanov, D. Razdoburdin, and M. Secondini, "Multichannel digital backpropagation with XPM-aware ESSFM," in *Proc. 17th Int. Symp. Wireless Commun. Syst.*, 2021, pp. 1–6.

[24] E. Ip, N. Bai, and T. Wang, "Complexity versus performance tradeoff for fiber nonlinearity compensation using frequency-shaped, multi-subband backpropagation," in *Proc. Opt. Fiber Commun. Conf.*, 2011, pp. 1–3.

[25] C. Häger and H. D. Pfister, "Physics-based deep learning for fiber-optic communication systems," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 1, pp. 280–294, Jan. 2021.

[26] V. Oliari et al., "Revisiting efficient multi-step nonlinearity compensation with machine learning: An experimental demonstration," *J. Lightw. Technol.*, vol. 38, no. 12, pp. 3114–3124, Jun. 2020.

[27] C. Häger and H. D. Pfister, "Wideband time-domain digital backpropagation via subband processing and deep learning," in *Proc. IEEE European Conf. Opt. Commun.*, 2018, pp. 1–3.

[28] S. Luo, S. K. O. Soman, L. Lampe, and J. Mitra, "Deep learning-aided perturbation model-based fiber nonlinearity compensation," *J. Lightw. Technol.*, vol. 41, no. 12, pp. 3976–3985, Jun. 2023.

[29] M. S. Neves, A. Lorences-Riesgo, P. P. Monteiro, and F. P. Guiomar, "Machine learning-aided nonlinearity-tailored carrier phase recovery for subcarrier multiplexing systems," in *Proc. Opt. Fiber Commun. Conf.*, 2024, pp. 1–3.

[30] E. Forestieri and M. Secondini, "Solving the nonlinear schrödinger equation," in *Optical Communication Theory and Techniques*. Boston, MA, USA: Springer, 2005, pp. 3–11.

[31] M. Secondini and E. Forestieri, "Analytical fiber-optic channel model in the presence of cross-phase modulation," *IEEE Photon. Technol. Lett.*, vol. 24, no. 22, pp. 2016–2019, Nov. 2012.

[32] M. Secondini, E. Forestieri, and G. Prati, "Achievable information rate in nonlinear WDM fiber-optic systems with arbitrary modulation formats and dispersion maps," *J. Lightw. Technol.*, vol. 31, no. 23, pp. 3839–3852, Dec. 2013.

[33] G. P. Agrawal, *Nonlinear Fiber Optics*. 3rd ed. San Diego, CA, USA: Academic Press, 2001.

[34] C. R. Menyuk and B. S. Marks, "Interaction of polarization mode dispersion and nonlinearity in optical fiber transmission systems," *J. Lightw. Technol.*, vol. 24, no. 7, pp. 2806–2826, 2006.

[35] M. Winter, C. A. Bunge, D. Setti, and K. Petermann, "A statistical treatment of cross-polarization modulation in DWDM systems," *J. Lightw. Technol.*, vol. 27, no. 17, pp. 3739–3751, Sep. 2009.

[36] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*. Englewood Cliffs, NJ, USA: Prentice Hall, 1999.

[37] 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 Photon. Technol. Lett.*, vol. 12, no. 5, pp. 489–491, May 2000.

[38] A. Wenzler and E. Luder, "New structures for complex multipliers and their noise analysis," in *Proc. 1995 IEEE Int. Symp. Circuits Syst.*, 1995, vol. 2, pp. 1432–1435.

[39] R. Yavne, "An economical method for calculating the discrete Fourier transform," in *Proc. Dec. 9-11, 1968, Fall Joint Comput. Conf., Part I*, 1968, pp. 115–125.

[40] J. W. Cooley and J. W. Tukey, "An algorithm for the machine calculation of complex Fourier series," *Math. Computation*, vol. 19, no. 90, pp. 297–301, 1965.