On Numerical Simulations of Ultra-Wideband Long-Haul Optical Communication Systems

Paolo Serena, *Member, IEEE*, Chiara Lasagni, Simone Musetti, and Alberto Bononi, *Senior Member, IEEE*

(Invited Paper)

***Abstract*—This paper discusses the key issues in reliable modeling and numerical simulation of ultra-wideband (UWB) long-haul fiber-optic systems, such as those spanning the C+L band. We analyze the complexity of the split-step Fourier method (SSFM), including modern scenarios based on graphical processing units (GPU), and its numerical setup for accurate simulations. We explain why using the Manakov equation in the nonlinear step of the SSFM remains a good solution also in the UWB scenario, even though the underlying assumptions may be violated. We next propose a technique, called virtual channel grouping, to speed up SSFM simulations in the UWB regime. Importance sampling-based techniques are proposed to speed up the computation of the enhanced Gaussian noise (EGN) model, with computational times of the order of seconds. With such novel solutions, accurate simulations are feasible even for UWB systems.**

***Index Terms*—EGN model, NLIN, split-step Fourier method, Raman effect, ultra-wideband systems.**

I. Introduction

EXPLOITING the bandwidth of optical systems beyond the conventional C-band spanned by erbium doped fiber amplifiers (EDFAs) is probably the most prominent short-term solution to increase the capacity of fiber-optic communication systems. Although increasing capacity through bandwidth expansion is much less power-efficient than increasing it through space division multiplexing [1], still the technology for bandwidth expansion is mature and several vendors are developing ultra-wideband (UWB) systems. The numerical simulation of such UWB systems may be prohibitive, hence a proper channel model should be used, depending on the required accuracy in emulating the propagation over the optical fibers. Aim of this work is to discuss the challenges of such UWB numerical simulations, providing physical insight and solutions to get accurate results in reasonable times.

On a local scale, the reference channel model is the nonlinear Schrödinger equation (NLSE) [2]–[4], whose Kerr nonlinear effect can be described by the coupled-nonlinear Schrödinger equation (CNLSE) or its averaged version, the Manakov-PMD equation [3].

The Manakov-PMD equation, and especially its version without PMD usually called the Manakov equation, is the reference propagation model for most of the literature on nonlinear optical communications. However, it is worth noting that the assumptions that lead to the Manakov equation may be violated in the regime of UWB systems. In this work we explain why the Manakov equation is indeed reliable in estimating the nonlinear interference (NLI) variance even in the UWB-regime.

The novelty introduced by very large bandwidths is in the non-negligible stimulated Raman scattering among channels [5]–[8], and the frequency dependence of the fiber parameters, whose most significant effect is summarized by the presence of the third-order dispersion.

Moving from a local scale to a global scale, i.e., relating the received signal to the transmitted one, is extremely challenging and approximations are mandatory. The most reliable channel model on a global scale is the iterative solution implemented by the split-step Fourier method (SSFM) [2], [9]–[11]. The SSFM is a convergent algorithm, hence it gets the exact solution for step sizes approaching zero. This reassuring property is suggesting us the most reliable way to keep control of the accuracy of a simulation: running SSFM simulations for smaller step-sizes until observing convergence. Such a wise choice becomes unpractical because of the huge required computational time in UWB systems, such that the typical approach is to check it for one reference link and to apply the numerical setup in other scenarios, hoping for a comparable accuracy.

However, in UWB scenarios, this approach may lead to significant errors unless the step size is scaled inversely with the square of the bandwidth of the signal [12], [13]. The simulation time of a reliable simulation should thus scale roughly with the fourth power of the bandwidth, as we will show in Section III. This means that, for instance, doubling the bandwidth calls for 16 times more operations at the same accuracy. The complexity is even more severe if the Raman effect is included since it calls for further filtering operations in the numerical solution.

Running an SSFM simulation over ultra-large bandwidths might thus be unfeasible, hence novel solutions are necessary.

A first solution is proposed in Section IV, where we first show that the curse of SSFM is due to the use of small step sizes to

Manuscript received May 28, 2019; revised July 27, 2019; accepted August 26, 2019. Date of publication September 2, 2019; date of current version March 11, 2020. This work was supported by Nokia Bell-labs, under grant VX/BL/LEG/E/18/00564 V, Villarceaux, France. *(Corresponding author: Paolo Serena.)*

The authors are with the Department of Ingegneria e Architettura, Università di Parma, Parma 43124, Italy (e-mail: paolo.serena@unipr.it; chiara.lasagni@unipr.it; simone.musetti@studenti.unipr.it; alberto.bononi@unipr.it).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JLT.2019.2938580

correctly reproduce four-wave mixing (FWM). Curiously, while in many optical links FWM is small, especially in modern links where the trend is to increase symbol rates, we are forced to run simulations with extremely small step sizes to avoid artificially increasing FWM. As a solution, in Section IV we show that it is better to neglect FWM, or part of it, rather than running simulations with too-large steps that incorrectly inflate FWM. To discriminate FWM we need to evaluate signal propagation at the channel level, hence to handle a large set of coupled NLSE [2] in UWB systems. This approach increases the complexity of the nonlinear step but allows an increase of the step size. We propose to split the simulation bandwidth into sub-bands, so as to reduce the number of coupled NLSEs and to capture the most relevant Kerr effects. The technique is called virtual channel grouping (VCG) and is described in Section IV, where the computational savings of VCG due to neglecting FWM at a given accuracy are quantified. Besides the advantage of capturing the dominant nonlinear effects, another advantage of VCG is that the Raman effect may be easily implemented as a flat gain per group, thus saving extra fast Fourier transforms (FFT).

An alternative to SSFM-based simulations is provided by perturbative solutions of the NLSE [14]. Such channel models are able to close integrations along distance, thus avoiding the problem of the SSFM step selection, but are valid only in the limit of weak nonlinearity. Luckily, this is the case of most communication systems of interest, hence the predictions of perturbative models are quite relevant. Another advantage of perturbative solutions is that it is possible to calculate the variance of the perturbation in closed form. Among existing models, of particular interest is the enhanced Gaussian noise (EGN) model [15], which is able to provide an expression of each nonlinear contribution to the NLI variance. Such a model, once expressed in the time domain and generally referred to as nonlinear interference noise (NLIN) model [14], [16], is able to highlight further properties of the NLI, such as the correlation functions and the statistics of the phase noise [17]. A particularly efficient numerical implementation of the EGN is obtained by using Monte Carlo integrations [18].

The EGN model and novel importance-sampling techniques are discussed in Section V.

In summary, the paper is organized as follows: in Section II we introduce the main models describing the optical fiber and comment on their reliability; in Section III we analyze the challenges of SSFM simulations in the ultra-wideband scenario; in Section IV we introduce the VCG technique to speed-up simulations at a given accuracy; in Section V we discuss the EGN model and propose importance-sampling to speed up its numerical computation.

II. Channel Models

The propagation of the electric field $\vec{A} = [A_x, A_y]^{\intercal}$ along distance z of an optical fiber is described by the NLSE:

$$\frac{\partial \vec{A}}{\partial z} = -\frac{\alpha}{2}\vec{A} - j\mathbf{B}_0\vec{A} - \mathbf{B}_1\frac{\partial \vec{A}}{\partial t} + j\frac{\beta_2}{2}\frac{\partial^2 \vec{A}}{\partial t^2} + \frac{\beta_3}{6}\frac{\partial^3 \vec{A}}{\partial t^3} - j\mathcal{N}(\vec{A}) \tag{1}$$

where α accounts for fiber attenuation, $\mathbf{B}_{0,1}$ are 2×2 matrices accounting for fiber birefringence and polarization mode dispersion (PMD), respectively, $\beta_{2,3}$ are the group velocity and third-order dispersion coefficients, respectively. For bandwidths exceeding the C+L, even the frequency response of the attenuation should be accounted for [7]. The operator $\mathcal{N}$ summarizes the Kerr nonlinear effect. Depending on the behavior of the fiber birefringence, the operator $\mathcal{N}$ can be described at different levels of approximations leading to the CNLSE or the Manakov-PMD equation [3].

Many investigations in the literature adopt the Manakov-PMD equation [3]. Such an equation is generally faster to be solved numerically compared to the CNLSE by at least a factor 3. Moreover, in the absence of PMD, the Manakov equation is a deterministic equation and can be simulated with a single Monte Carlo run.

However, in the novel scenario of UWB simulations, the assumptions of the Manakov equation may be questionable. This observation was first made by Marcuse *et al.* [3] and came back in the context of large capacity systems in [4], [19], [20]. The reason is that the fundamental averaging of the Manakov equation requires the length scale of mode-coupling, known as the fiber correlation length L_{corr}, to be much smaller than the typical length scales of the linear and nonlinear effects, such as the dispersion length L_{D} and the nonlinear length L_{NL} [2]. In particular, the dispersion length scales inversely with the square of the bandwidth, thus in a UWB system it may be comparable, or even smaller, than L_{corr}. The fiber correlation length can be further related to two characteristic lengths of birefringence: the length L_{w} over which the birefringence axes change, and the fiber beat length L_{B}, i.e., the length over which the birefringence eigenvalues periodically beat. In the limit of $L_{\text{B}} \ll L_{\text{w}}$ we have $L_{\text{corr}} \simeq L_{\text{w}}$ since two neighboring waveplates are described by statistically uncorrelated random matrices. In the limit $L_{\text{B}} \gg L_{\text{w}}$, the transfer matrix of a waveplate is close to the identity matrix, hence the decorrelation is eventually reached after crossing several waveplates. In such a regime it has been proved that $L_{\text{corr}} = \frac{L_{\text{B}}^2}{2\pi^2 L_{\text{w}}}$. Typical values of L_{corr} range from 1 m to 100 m, while L_{B} is typically about 10 m [21].

In Appendix A we investigate bounds on the Manakov equation accuracy compared to the CNLSE, confirming its reliability in predicting the NLI variance even for UWB simulations. For this reason, in this paper we concentrate on using the Manakov equation.

The Manakov equation lends itself to a perturbative solution. In the special, most important case of first-order regular perturbation of the NLSE, the input/output field at angular frequency ω under the Manakov formalism is:

$$\vec{A}(z,\omega) = \vec{A}(0,\omega) - j\gamma \iint_{-\infty}^{\infty} \eta(\omega,\omega_1,\omega_2) \times \vec{A}^{\dagger}(0,\omega+\omega_1+\omega_2)\vec{A}(0,\omega+\omega_1)\vec{A}(0,\omega+\omega_1)\frac{\mathrm{d}\omega_1}{2\pi}\frac{\mathrm{d}\omega_2}{2\pi} \tag{2}$$

where $\dagger$ indicates transpose-conjugate while $\eta(\omega,\omega_1,\omega_2)$ is the fiber kernel [22]–[24], whose expression is commented in

Appendix B. For many optical fibers the integral along z inherent in the kernel η can be evaluated in closed form, thus avoiding the numerical problems related to the step size choice of the SSFM. However, eq. (2) does not yield any particular advantage for numerical computations compared to the SSFM because of the presence of the double frequency integrals. The main advantage of (2) is that it is the starting point to derive GN and EGN models, which can provide the variance of the NLI in computational times of the order of seconds, as we will show in Section V.

III. SSFM Based Simulations

The SSFM algorithm aims at finding an approximate solution of the NLSE, by accounting in the nonlinear step either for the complete Kerr effect when using the CNLSE, or for an averaged version when using the Manakov equation. Since the Manakov equation is reliable even for UWB simulations, as discussed in Section II, we concentrate on it because it is simpler and faster to solve.

The complexity of an SSFM simulation depends on the number of FFT points and the number of steps. The number of FFT points N_{FFT} is related to the number of data symbols N_{symb} and the number of samples per symbol N_{t} by the product $N_{\text{FFT}} = N_{\text{symb}} N_{\text{t}}$. In the frequency domain normalized to the symbol rate R of the channel under test (CUT), $1/N_{\text{symb}}$ is the minimum frequency resolution while $N_{\text{t}}/2$ is the maximum resolved frequency (N_{t} is the Nyquist frequency). In a reliable simulation, N_{symb} should scale with the maximum walk-off among channels to avoid aliasing in the time domain, while N_{t} should scale with the wavelength division multiplexing (WDM) bandwidth to avoid aliasing in the frequency domain. For a WDM of bandwidth B_{WDM} the walk-off per unit-length is proportional to $|D|\, B_{\text{WDM}}$, $|D|$ being the largest value of the fiber dispersion over the WDM bandwidth, hence, in the relevant case of dispersion uncompensated links, we have:

$$N_{\text{symb}} = \mathcal{O}(|D|\, B_{\text{WDM}} R L)$$

$$N_{\text{t}} = \mathcal{O}\left(\frac{B_{\text{WDM}}}{R}\right)$$

after an optical link of length L. The symbol $\mathcal{O}(g(x))$ indicates a term whose magnitude is upper bounded by the function $Mg(x)$, with M a positive constant. The number of FFT points thus scales with $N_{\text{FFT}} = \mathcal{O}(|D|\, B_{\text{WDM}}^2 L)$, independently of the granularity of the bandwidth.

The second ingredient to estimate the SSFM complexity is the number of steps N_{h}. Such a number depends on the length of the first step h_1 and the step updating rule. The first step should scale with the inverse of the square of the signal bandwidth to correctly capture the impact of FWM [12]:

$$h_1 = \mathcal{O}\left(\frac{1}{|D| B_{\text{WDM}}^2}\right). \tag{3}$$

For longer steps, the FWM is overestimated since the phase matching coefficient (18) is under-sampled along space and thus FWM accumulates more coherently [12], hence at a faster rate. Some popular step-updating rules, such as the constant local error (CLE) [25], [26], the nonlinear phase criterion (NLP) [10], and the logarithmic step (LS) size [11], update the k-th step as[1]

$$h_{k+1} \simeq h_k e^{\frac{\alpha}{q} h_k} \tag{4}$$

in the limit of small step-size, as usual in UWB simulations. The parameter q is 1 for the LS and the NLP, 2 and 3 for the CLE with asymmetric/symmetric step, respectively [12], [25], [26].

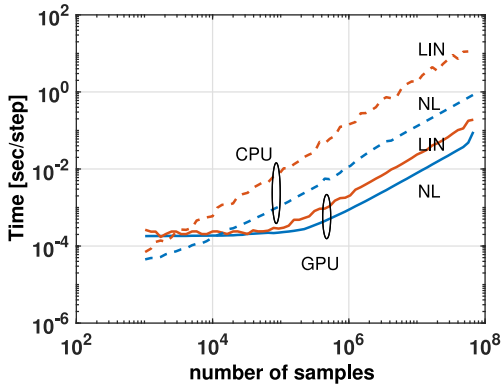

Fig. 1. Computational time to evaluate a linear (LIN) and a nonlinear (NL) step of the SSFM vs number of samples, both with or without GPU.

As a reference, we found that an SSFM simulation of a 10 THz WDM system over a 100 km single-mode fiber (SMF) including Raman scattering is accurate with an SNR error < 0.1 dB by using the symmetric-step CLE with the first step of 30 cm, for a total of 300000 SSFM steps. Such a number corresponds to a FWM-aware first step choice with $\Phi_{\text{FWM}} = 25$ rad [12, eq. (13)]. The same first step with the NLP requires a nonlinear phase per step of $6 \cdot 10^{-5}$ [rad] with a WDM total power of 23 dBm (0 dBm/channel).

The step is thus increasing along distance thanks to the decreasing nonlinear Kerr effect. For a lossless fiber the number of steps scales with $N_{\text{h}} = \mathcal{O}(L/h_1)$. Such scaling is well satisfied even in the limit of many steps per fiber, as usual in time-consuming UWB simulations, since with very small steps the stretching (4) is quite small over the fiber length.

The linear and nonlinear step evaluation takes time, whose cost depends on the amount of parallelization available in the computer architecture. As a reference, in Fig. 1 we estimated the elapsed time to run an elementary linear/nonlinear step for a polarization division multiplexing (PDM) signal vs N_{FFT}. In the linear step, we included only group velocity dispersion (GVD), while in the nonlinear step only memoryless self-phase modulation (SPM). The propagated signal constellation has no impact on the computational time, hence we used a Gaussian distributed complex signal.

The computational time for the linear step accounts for linear filtering made by FFTs, the evaluation of the exponential function of the GVD filter, and the filtering operation in the frequency domain. The nonlinear step needs the additional computation of absolute square values to get SPM, but it is FFT free. We

[1] Such expression does not account for the Raman effect. However, by including it in (4), we observed a negligible impact on the results since fiber attenuation is the dominant effect on the power variations.

used a cluster based on INTEL XEON E5-2683v4 2.1 GHz 32 cores central processing units (CPU) with 128 GB of RAM and NVIDIA Tesla P100 graphics processing unit (GPU). All simulations were performed with MATLAB R2017b. With GPU we measured just the pure computational time within the GPU since the transfer of data GPU/memory is generally performed only at the beginning/end of a numerical simulation.

While the original FFT algorithm has been introduced for signals having $\log_2(N_{\text{FFT}})$ an integer, mixed-radix FFT algorithms [27] relax this constraint and just require not too big prime factorizations. For this reason, we used a number of samples N_{FFT} whose factorization contained only combinations of powers of 2 or 3 or 5, e.g., $720000 = 2^7 3^2 5^4$. Such 3 numbers are enough to sample the set $\mathbb{N}$ of the integers with a sufficiently fine grid such that any value of N_{symb} or N_{t} can be reproduced without wasting too many resources as for a coarse grid made of powers of 2 only.

In Fig. 1 we observe that the GPU parallelization performs all computations at the same time up to about 10^5 samples, after which the time starts to scale as much as in CPU. The reason is related to the exhaustion of the floating point operations (Gflops) available in the GPU. In the large N_{FFT} region we observe that the linear step is almost a factor 3 slower than the nonlinear step in GPU, and a factor 18 in CPU. It is worth noting that in the presence of the Raman effect the cost of the nonlinear step is comparable to the cost of the linear step since the Raman filtering on nonlinearity calls for an additional FFT/inverse-FFT operation.

Since the cost of the nonlinear step is a fraction of the cost of the linear step for $N_{\text{FFT}} \gg 1$, we can focus on the number of multiplications C_{m} to perform the linear step to have an idea of the scaling properties of a reliable SSFM simulation. Putting together all the previous observations, we have:

$$\begin{aligned} C_{\text{m}} &= \mathcal{O}\left(N_{\text{FFT}}\left(\log_2\left(N_{\text{FFT}}\right)+1+\kappa_{\text{e}}\right)N_{\text{h}}\right) \\ &= \mathcal{O}\left(|D|\,B_{\text{WDM}}^4 L^2\left(\log_2\left(|D|\,B_{\text{WDM}}^2 L\right)+1+\kappa_{\text{e}}\right)\right). \end{aligned}$$

where κ_{e} is the cost of each exponentiation while the cost of FFT has been taken to scale as $N_{\text{FFT}}\log_2 N_{\text{FFT}}$. The complexity scales extremely fast for increasing bandwidths, with at least a fourth power in B_{WDM}. Such a scaling may be too fast for UWB systems, so that alternative solutions have to be found. In the next section we investigate in more detail the requirement of the walk-off window to set N_{symb}.

A. Walk-Off Window

We sized the number of symbols to tackle the worst case of aliasing in the time domain, i.e., such that the walk-off between the edge channels of the WDM is less than the signal duration. The reason is related to the use of circular convolutions that make the discrete time axis modulo N_{FFT}. With this choice, cross-channel nonlinear effects occurring at two different coordinates are statistically independent and the simulation does not show numerical artifacts. However, it is worth noting that, after the distance corresponding to a walk-off equal to the entire signal duration, the interfering signal may be well different than its initial shape in case of large accumulated GVD, as in UWB scenarios. Hence the aliased replica is only partially correlated with the original one, such that the artificial resonances may be small. We thus investigated how much temporal aliasing can be tolerated by testing the following setup.

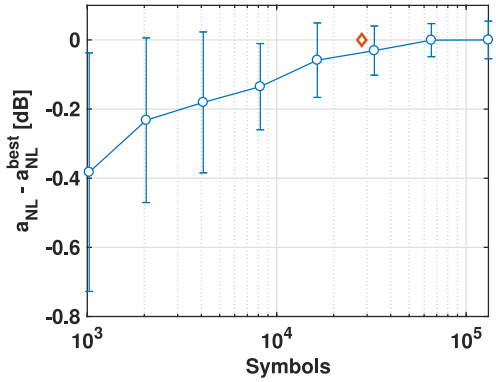

Fig. 2. Error on the unit-power NLI variance a_{NL} versus sequence length N_{symb}. The error bars are one standard deviation from the average. The diamond highlights the number of symbols required by the walk-off length between the edge channels, eq. (6).

We estimated the unit-power nonlinear interference variance a_{NL}, which is related to the NLI variance per channel σ^2_{NLI} by the definition:

$$\sigma^2_{\text{NLI}} \triangleq a_{\text{NL}} P^3 \tag{5}$$

with P channel power, equal to 0 dBm in our setup. In the figures we express a_{NL} in a dB-scale by using $10\log_{10}(\frac{a_{\text{NL}}}{1\text{mW}^{-2}})$.

We tested different sequence lengths N_{symb} in the range 2^{10} to 2^{17} with different random seeds and then computed the average a_{NL} and its standard deviation with respect to the most accurate available value, i.e., the mean value with the longest sequence, called $a_{\text{NL}}^{\text{best}}$. The mean of a_{NL} yields information about the bias introduced by using a given number of symbols, while its standard deviation yields information about the Monte Carlo randomness.

We analyzed a transmission of 41 channels spaced 50 GHz and modulated at $R = 49$ Gbaud by PDM quadrature phase shift keying (QPSK) signals with 0.01 pulse roll-off. The CUT was centered at $\lambda = 1550$ nm. The link was made of 20×100 km spans of SMF with dispersion $D = 17$ ps/nm/km, third-order dispersion coefficient 0.057 ps/nm^2/km, attenuation 0.2 dB/km and nonlinear coefficient $\gamma = 1.26$ 1/W/km. We used a matched filter based receiver where we recovered the average phase and polarization mismatch by a trained least-squares estimation with 1 tap [28].

Fig. 2 shows the corresponding results. The error bars indicate one standard deviation bar while circles indicate the average values. The diamond is a label to indicate the number of symbols N_{wo} called by the walk-off length between the edge channels, which is [12]:

$$N_{\text{wo}} = |D_{\text{cum}}|\,B_{\text{WDM}}\frac{\lambda_{\text{c}}^2}{c}R\cdot 10^{-3} \quad [\text{symbols}] \tag{6}$$

where B_{WDM} is in GHz, the symbol rate R in Gbaud, the WDM central-channel wavelength λ_{c} in nm, the speed of light c in

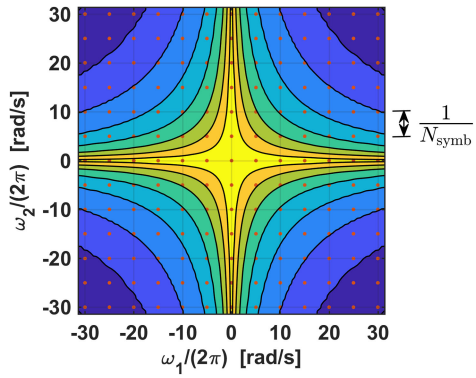

Fig. 3. Example of contour levels of the fiber kernel $|\eta|^2$and corresponding normalized frequency grid by using sequences of N_{symb} symbols.

m/s and the largest, channel-dependent, accumulated dispersion along the link D_{cum} in ps/nm. We observe that below N_{wo} the average a_{NL} is appreciably smaller than the most accurate one, thus showing a negative bias that is not canceled by iterating over Monte Carlo runs.

The negative bias can be understood from the following reasoning. Since we are using circular convolutions, our digital signals are implicitly periodic signals, thus showing a spectrum concentrated at discrete frequencies multiple of $1/N_{\text{symb}}$ (see Section III). With reference to eq. (2) without loss of generality, we are thus sampling both frequency axes (ω_1, ω_2) over such a grid. Now, the fiber kernel η in eq. (2) has typical contour levels which take the form of a cross concentrated over the main (ω_1, ω_2) axes. Fig. 3 sketches the idea. On the axes the kernel is real, which implies a nonlinear distortion in quadrature with the symbol of interest, i.e., a constant phase shift. However, at a given value ω_1, the kernel is a decreasing function of $|\omega_2|$ (see Appendix B) at a faster rate when increasing $|\omega_1|$. Hence, when decreasing the data length N_{symb}, the frequency grid becomes coarser such that, because of the bad frequency sampling, the dominant contribution is more and more provided by the samples on the main axes, corresponding to the frequencies of the interfering channel's carriers. This way, the cross-channel nonlinearity due to far away channels gets closer to a constant phase shift for decreasing N_{symb}. But a constant phase shift is not distortion since it is fully recovered by the carrier phase estimator, hence the negative bias observed in Fig. 2.

Fig. 2 also highlights the importance of the Monte Carlo estimation noise. This is particularly evident in the large error bars at small values of N_{symb}, while it disappears at a large number of symbols. We recommend using large symbol windows to mitigate both the Monte Carlo uncertainty and the time aliasing problems in a single simulation.

IV. Virtual Channel Grouping

The previous discussion provided computational cost estimates to explain why UWB simulations, e.g., encompassing the C+L bands, are technically barely feasible with today's simulation technology. In this section we introduce a technique, which we call virtual channel grouping (VCG), to speed-up UWB simulations. The technique consists of partitioning the WDM channels into N_{g} contiguous groups (akin to super-channels), and separately simulate the propagation of the groups along the fiber accounting for just self- and cross-group nonlinear effects. Namely, the solution of the NLSE equation presented in Section I is approximated by the superposition of the solutions of N_{g} coupled-NLSEs, one for each group. Therefore, VCG represents a group-wise version of the well-known separate fields approach [2], [29] which we next describe, where the FWM among groups is neglected.

The total propagating field, composed of N_{ch} WDM channels, can be written as a multiplex of $N_{\text{g}} \leq N_{\text{ch}}$ neighboring groups as:

$$\vec{A}(z,t) = \sum_{n=1}^{N_{\text{g}}} \vec{A}_n(z,t)\exp(j\Omega_n t)$$

with $\vec{A}_n$ the n-th group of channels with central frequency Ω_n. The corresponding set of coupled-Manakov equations for $n = 1, \ldots, N_{\text{g}}$ is:

$$\begin{aligned}\frac{\partial \vec{A}_n}{\partial z} = &-\frac{\alpha}{2}\vec{A}_n - j\mathbf{B}_{0n}\vec{A}_n - \mathbf{B}_{1n}\frac{\partial \vec{A}_n}{\partial t} + j\frac{\beta_{2n}}{2}\frac{\partial^2 \vec{A}_n}{\partial t^2} \\ &+ \frac{\beta_{3n}}{6}\frac{\partial^3 \vec{A}_n}{\partial t^3} - j\frac{8}{9}\gamma \sum_{k,\ell,m\in\mathcal{S}_n} \vec{A}_m^{\dagger}\vec{A}_k\vec{A}_\ell \end{aligned} \tag{7}$$

where the subscript n indicates that the parameters of the NLSE in (1) are evaluated at frequency Ω_n, and

$$\mathcal{S}_n \triangleq \{(k,\ell,m) \,:\, \Omega_k + \Omega_\ell - \Omega_m - \Omega_n = 0\} \tag{8}$$

is the set of group indexes (k,ℓ,m) whose corresponding central frequencies satisfy the law of conservation of energy [2]. In the limit of a single group composed of the whole WDM, the separate groups Manakov equations reduce to the standard eq. (1). The VCG savings come from solving (7) for each group of channels through the SSFM by accounting only for cross-channel interactions and neglecting FWM, for which a closed form expression of the nonlinear step exists [29]–[31], as also reported in Appendix C for convenience. The solution obtained in the single group scenario, that we call the *unique field* solution, represents the "true" solution since it accounts for all the nonlinear effects occurring during propagation, including FWM. On the other hand, VCG neglects part of the FWM, as we will discuss in detail in Section IV-A.

The VCG-SSFM requires to solve a set of coupled differential equations, as opposed to the single equation of the unique field. However, each group has a bandwidth $B_{\text{g}} < B_{\text{WDM}}$ and thus can be sampled at a smaller frequency, such that the overall number of samples is unchanged. The idea is thus an application of multirate signal processing [27] to the NLSE. The reduction of the complexity of the VCG-SSFM with respect to the unique field SSFM is due to the possibility to increase the step size, such that the first step (3) becomes $\mathcal{O}(1/(B_{\text{g}}B_{\text{WDM}}))$. This way, longer steps by a factor N_{g} are possible. However, while the complexity of the linear step changes little by changing N_{g} [27], the complexity of the nonlinear step significantly changes.

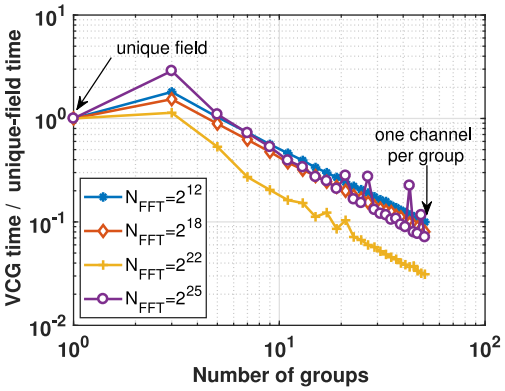

Fig. 4. Ratio of the computational time of VCG over the unique SSFM for different FFT sizes. Computations performed in GPU by vectorizing both the nonlinear and the linear steps.

To quantify the computational advantage of the VCG we measured the time to perform a single step with both methods, and normalized the VCG time by N_g to account for the smaller number of steps. Fig. 4 shows the computational gain of VCG expressed as the ratio of such computational times versus the number of groups N_g, for different signal sizes. We followed the same procedure of Fig. 1 by optimizing the numerical implementation of each step at our best. In particular, all computations have been vectorized in MATLAB. We note that for $N_g > 3$ the cost is decreasing for increasing number of groups, yielding a gain of almost a factor 10 in the separate groups setup with one channel per group. Please remember that the larger the number of groups, the larger the error introduced by VCG, hence a trade-off complexity/accuracy exists. Numerical results in Section IV-B will quantify this trade-off.

A. Physical Insight

The nonlinear Kerr effect is generally classified with different names depending on the involved channels. Historically, the classification was conceived for quasi constant-wave (CW) WDM channels spaced far apart, thus focusing only on the carriers involved in the process [2]. Referring to (7) where the indexes label the WDM channels (rather than the groups), in the standard classification the contribution in the nonlinear term summation with indexes $k = \ell = m = n$ corresponds to SPM, those with indexes $(k = \ell) \neq (m = n)$ to cross-phase modulation (XPM), and all remaining contributions to FWM [2]. However, in modern transmissions with large bandwidth efficiency, Poggiolini first showed in [32] that channels may interact by means of the Kerr effect even if their carrier frequencies do not meet (8). In this framework, with reference to eq. (2), it is more appropriate to interpret the Kerr nonlinearity as a FWM among the Fourier frequencies of the WDM spectrum satisfying $\omega = \omega_1 + \omega_2 - \omega_3$. If we set, for the sake of example, $\omega = 0$ for the low-pass frequency corresponding to the carrier frequency Ω_n of the CUT, the remaining frequencies are constrained by $\omega_3 = \omega_1 + \omega_2$.

Fig. 5a sketches an example of the frequency domains in the (ω_1, ω_2) plane involved in such FWM, i.e., the domain where the cubic term in the integrand of (2) is non-zero [32], for channels having the spectra depicted in Fig. 5e. We refer without loss of generality to a single polarization since the additional contributions coming from cross-polarization interactions can be treated similarly [33]. In Fig. 5c we note that the traditional SPM, XPM and FWM terms, which would correspond in the quasi-CW regime to dots on a regular grid having the carrier frequency spacing, show-up now as hexagons (inscribed in a square of edge equal to the channel bandwidth B_{ch}) centered at grid dots. In this paper, we will keep calling the contributions of such hexagons as SPM (purple), XPM (green), and classic FWM (red). However, due to the close spacing of the WDM channels, new important triangular domains appear, as first noticed in [32].

It is worth noting that in [15], [32] all contributions involving two channels were called cross-channel interference (XCI), while all the contributions involving three or four channels were called multi-channel interference (MCI).

In this paper, we will call the contributions of the blue triangles in Fig. 5a as XCI (this is consistent with the XCI naming in [15, Fig. 7]), and those of the orange triangles as MCI, as better specified in Fig. 5d, where we highlight by examples the position of the four spectral frequencies generating the Kerr nonlinearity (purple arrow is the carrier of the CUT), our labeling, and their k, ℓ, m, n index relations. For numerical purposes, we found important to adopt the classification in the figure instead of treating everything as FWM in order to efficiently sample the islands by importance sampling, as discussed in Section V.

The efficiency of a generic FWM process is weighted by the fiber kernel, which depends on the phase matching coefficient $\Delta\beta$ defined in Appendix B, being maximum for $\Delta\beta = 0$. This can occur only when the four frequencies are degenerate or pairwise degenerate, e.g., the carrier frequencies involved in the SPM and XPM effects. In Fig. 5d, the contributions of SPM and XPM are thus highly efficient. Among the other effects, XCI is generally efficient since it involves closer frequencies coming from neighboring channels. Classic FWM and MCI are generally much less efficient since they involve far away frequencies.

The unique field approach treats the whole WDM comb as a single super-channel, hence it implicitly accounts for all the SPM, XPM, XCI, FWM and MCI effects. On the other hand, as we observed previously, a bad SSFM setup with too long steps emphasizes XCI, FWM, and MCI as if they had a smaller phase matching coefficient, especially in UWB systems, thus inducing wrong estimations.

The Kerr contributions captured by the separate groups approach of (7) are visualized in Fig. 5b, where as an example we grouped the 9 WDM channels in groups of 3. We will call the group containing the CUT as the group under test (GUT). Fig. 5b is organized on two layers, with the channel islands at the bottom layer and the group islands at the higher layer. We note that in the higher layer all the triangles have disappeared since they correspond to group index combinations which do not satisfy (8). However, some triangles remained in the lower layer within the group-hexagons, which are able to capture all the Kerr effects within each group.

The proposed VCG method reduces the computational complexity of (7) by neglecting all classic group-FWM terms, as

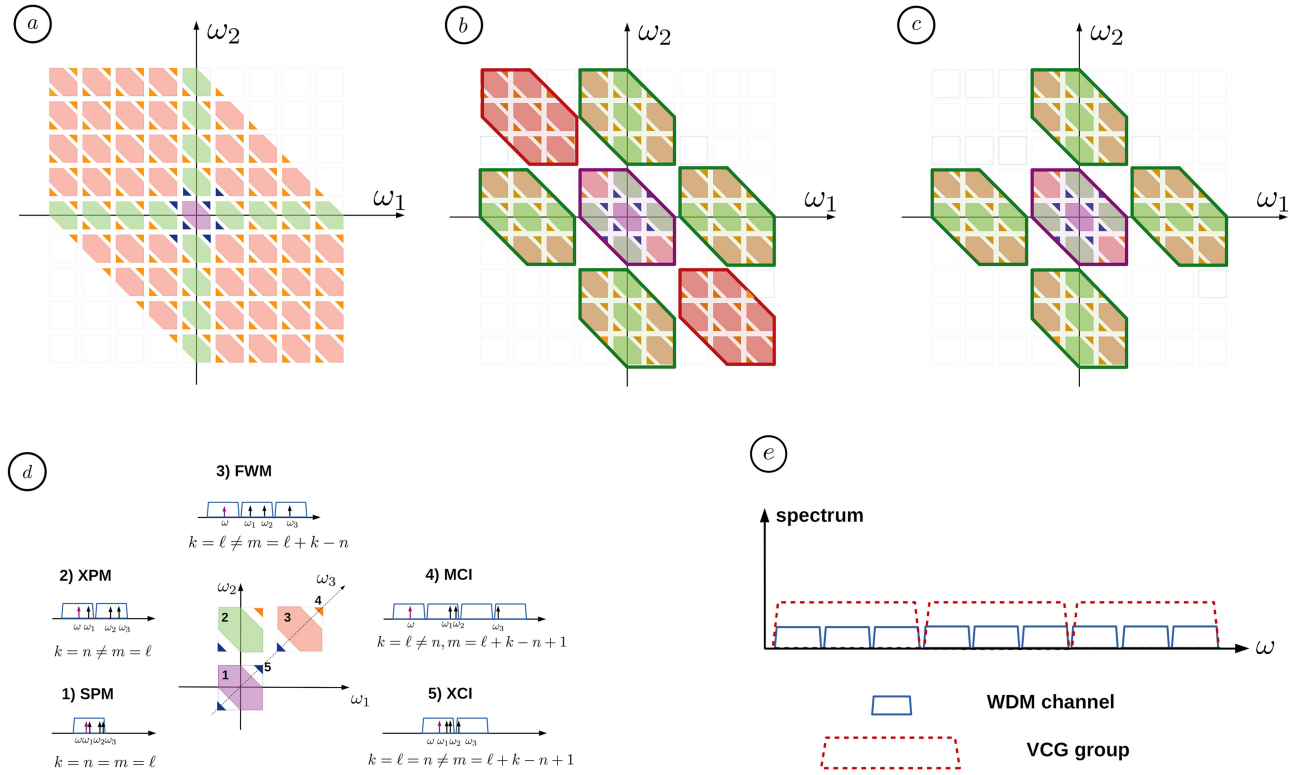

Fig. 5. ⓐ: Frequencies ω_1, ω_2 involved in the Kerr effect of (2) at $\omega = 0$, where the fourth frequency of the FWM-related process is $\omega_3 = \omega_1 + \omega_2$ for the law of energy conservation. ⓑ: Same as left but referred to (7) with superimposed the group- islands accounting for intra-group SPM, inter-group XPM and FWM. ⓒ: Islands accounted by the VCG algorithm. ⓓ: corresponding NLI classification with examples. The example is for a 9 channel WDM transmission with 3 channels-per-group VCG, as sketched in ⓔ.

exemplified in Fig. 5c. The gain comes not only from avoiding classic FWM computations, but also from the fact that the nonlinear term in the SSFM can be expressed in closed-form, as discussed in Appendix C. Unfortunately, not all the FWM contributions can be captured by VCG, as visible by comparing Figs. 5a and 5c. In fact, inter-group FWM, cross-group interference (XGI) and multi-group interference (MGI) are missed by VCG. VCG thus saves time at the expense of a bias. Such a trade-off is investigated in the next Section.

B. Numerical Results

In order to highlight the main features of the VCG-SSFM approach, in this section we focus on a scenario where FWM is non-negligible [34]. Namely, we consider a 216 WDM channel transmission of PDM-16 quadrature amplitude modulation (QAM) signals with symbol rate 4 Gbaud and channel spacing 4.7 GHz, for an overall bandwidth of ≈1 THz, with per-channel power −9 dBm. We analyzed a 20 × 100 km dispersion uncompensated SMF link. The transmitted sequence length was 4096 symbols to capture the maximum walk-off length N_{wo} in (6), while the number of points per symbol was set to have a Nyquist frequency per group a factor 1.5 larger than the group bandwidth. The SSFM was implemented with the CLE step-updating rule, with the symmetric step described in Section III.

Figure 6 shows the SNR evaluated at CUT as a function of the number of SSFM steps, both for the unique field approach

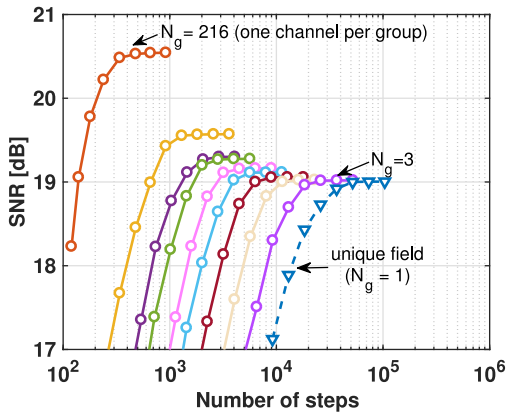

Fig. 6. SNR vs the number of SSFM steps for a PDM-16QAM transmission after 20 × 100 km. Solid lines: VCG with variable number of channels per group $N_{\mathrm{g}} = 216,73,44,23,15,11,9,5,3$. Dashed line: unique field solution.

and the VCG with a decreasing number of groups N_{g}. For each scenario, we increased the number of steps until observing convergence, considering an SNR error smaller than 0.005 dB as a convergence criterion. Thus, the point of each curve associated with the largest number of steps represents the most accurate solution for that setup.

We note that, according to the adopted stop-criterion, the most accurate SNR of VCG with a single channel per group, i.e.,

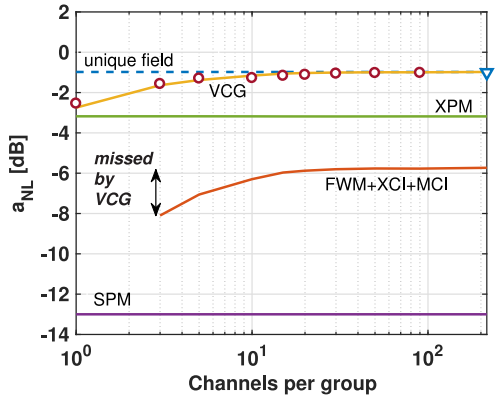

Fig. 7. a_{NL} vs the number of channels per group. Markers: SSFM simulations (triangle for the unique field, circles for the VCG). The solid lines show the EGN model of VCG, thus accounting only for the islands of Fig. 5c, and its corresponding Kerr contributions to a_{NL}.

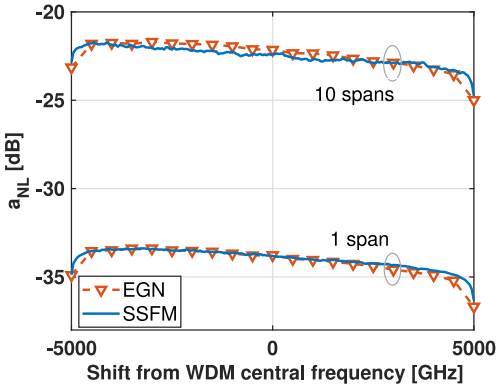

Fig. 8. a_{NL} vs the shift from the WDM central frequency for a PDM-64QAM transmission over a 10 THz bandwidth after 100 km and 10×100 km. Solid lines: SSFM simulations. Dashed lines with markers: EGN-model.

the standard separate-fields approach, is obtained after only 920 steps, thus with big step-savings compared to the 103820 steps required by the most accurate unique-field benchmark SNR, called $\mathrm{SNR}_{\mathrm{best}}$. However, the separate-fields estimate suffers from a bias of $\approx$1.5 dB with respect to $\mathrm{SNR}_{\mathrm{best}}$. Such a bias approaches zero as we increase the group size. For instance, the green curve in the figure with $N_{\mathrm{g}} = 23$ saturates with a bias of 0.28 dB with respect to $\mathrm{SNR}_{\mathrm{best}}$ after 5660 steps. The same error, in absolute value, is reached by the unique field curve after 26000 steps, thus with a reduction of steps by VCG by a factor 4.6. For the edge channels of the WDM comb we observed convergence at almost twice the number of steps. We ascribe this result to the lower NLI variance of such channels that calls for more accuracy to keep the SNR error unchanged over the WDM bandwidth, according to [12, eq. (6)]. However, we noted that the SNR of 90% of the WDM channels showed to saturate at a comparable number of steps.

We stopped all the simulations with a very small confidence error on the SNR of 0.005 dB, to clearly observe saturation and to have a reliable reference value for $\mathrm{SNR}_{\mathrm{best}}$. However, it is worth noting that using larger confidence errors might help, thanks to a cancellation of errors. In fact, a higher confidence error calls for a smaller number of steps, hence with an SNR for sure smaller than the saturation value at infinite steps [12], as visible in Fig. 6. On the other hand, VCG always overestimates SNR.

The improvement in accuracy of VCG for an increasing number of channels per group is related to the inclusion of subsets of FWM, XCI, and MCI effects. To better appreciate this point and corroborate the discussion of Section IV-A, we plot in Fig. 7 the a_{NL} estimated with VCG as a function of the group size, and the relative contributions of SPM, XPM and FWM (including MCI and XCI). The corresponding variance contributions (solid lines) were computed by means of the perturbative EGN model that will be discussed in Section V, whose accuracy is confirmed by the excellent match with SSFM simulations in Fig. 7. Please note that the red curve accounting for FWM, XCI, and MCI with one channel starts from $-\infty$ dB with one channel per group, with a significant error in the VCG prediction. However, by including the two next-neighbor channels, hence XCI, the accuracy improves by 1 dB saturating for increasing number of channels per group.

V. Sampling Methods for the EGN Model

An efficient alternative to the SSFM is provided by semi-analytical models such as the EGN model, which we next discuss in more efficient importance-sampling based versions.

The EGN model accurately evaluates the variance of the NLI under the approximation that the Kerr effect is a perturbation. Such a variance can be written as the sum of two contributions [14]–[16], [33]:

$$\sigma^2_{\mathrm{NLI}} = \sigma^2_{\mathrm{GN}} + \sigma^2_{\mathrm{HON}} \tag{9}$$

where σ^2_{GN} accounts for the second order cumulants of the constellation symbols, while σ^2_{HON} accounts for the higher-order cumulants [33]. Since only Gaussian distributed symbols have $\sigma^2_{\mathrm{HON}} = 0$, the part σ^2_{GN} is called the GN-model contribution and σ^2_{HON} the modulation-format dependent EGN correction.

Equation (9) can be easily converted to a signal-to-noise ratio (SNR), while the SNR is usually converted to a bit-error rate (BER) under the assumption of an additive white Gaussian noise channel. The latter assumption holds whenever the receiver is not able to exploit correlations in the NLI or when the NLI shows a negligible contribution of nonlinear phase and polarization rotation noise (PPRN) [35], and represents a good approximation for nowadays dispersion-uncompensated ultra-long links. For instance, in Fig. 8 we show the a_{NL} coefficient defined in (5) evaluated by the EGN together with SSFM simulations for a 201 WDM channel transmission with per-channel power 0 dBm, symbol rate $R = 49$ Gbaud and channel spacing $\Delta f = 50$ GHz, for an overall bandwidth of 10 THz, for a PDM-64QAM system after $n \times 100$ km of SMF, $n = 1{,}10$, with fiber parameters as in Section III-A. We included the stimulated Raman scattering among channels in the EGN model by means of the triangular approximation of the polarization-averaged Raman profile [36] with slope 0.028 $(\mathrm{THz} \cdot \mathrm{W} \cdot \mathrm{km})^{-1}$, while in the SSFM we used a polynomial interpolation of the experimental profile [2], with a peak value of 0.42 $(\mathrm{W} \cdot \mathrm{km})^{-1}$. The numerical setup used

$N_{\text{symb}} = 65700$, $N_{\text{t}} = 432$ with a symmetrized-SSFM with a first step of 18 cm and the CLE step-updating rule. The fit is very good, confirming the validity of the EGN as well as the triangular approximation of the Raman profile up to a bandwidth of 10 THz, i.e., close to the frequency-shift of maximum Raman gain [2].

The numerical computation of (9) aims at evaluating multi-dimensional integrals [15], [18], [33]. For instance, if all channels are modulated with sinc pulses, σ^2_{GN} after matched filtering requires the evaluation of the following:

$$\sigma^2_{\text{GN}} = \left(\frac{8}{9}\gamma\right)^2 P^3 \sum_{h,\ell,s} \iiint_{-\infty}^{\infty} |\eta(\omega,\omega_1,\omega_2)|^2 \left|\tilde{P}_n(\omega-\Omega_n)\right|^2 \times \left|\tilde{P}_h^*(\omega+\omega_1+\omega_2-\Omega_h)\right|^2 \left|\tilde{P}_s(\omega+\omega_2-\Omega_s)\right|^2 \times \left|\tilde{P}_\ell(\omega+\omega_1-\Omega_\ell)\right|^2 \frac{\mathrm{d}\omega}{2\pi}\frac{\mathrm{d}\omega_1}{2\pi}\frac{\mathrm{d}\omega_2}{2\pi} \tag{10}$$

where $\tilde{P}_k(\omega)$ is the Fourier transform, with rectangular shape, of the supporting pulse of the channel modulated at carrier frequency Ω_k. The triple integral in (10) is over the islands shown in Fig. 5a, while the triple summation is over the set of channels satisfying $|\Delta_{h\ell sn}| \le 2\pi B_{\text{ch}}$, with $\Delta_{h\ell sn} \triangleq \Omega_h - \Omega_\ell - \Omega_s + \Omega_n$. In particular, we observe that the combinations yielding $|\Delta_{h\ell sn}| = 2\pi B_{\text{ch}}$ correspond to the triangles in Fig. 5a.

In [18] Dar *et al.* proposed to evaluate the triple integral in (10) by Monte Carlo (MC) integration, which in high dimensions is more efficient than quadrature rules based on interpolating polynomials since it better tolerates the "curse of dimensionality" [37], i.e., for well-behaved functions the standard deviation of the MC error, contrary to quadrature rules, scales with $1/\sqrt{N_{\text{MC}}}$ independently of the number of dimensions, where N_{MC} is the number of samples. Besides this aspect, the simplicity of the MC method is another argument to support its application.

Here we suggest to fully exploit the MC idea to randomly sample also the triple summation $\sum_{h,\ell,s}$ over the channel indexes in (10). The approach gives some advantages in evaluating FWM contributions because it removes the loops of $\sum_{h,\ell,s}$ in favor of a single MC summation, whose vectorized implementation yields some improvements in numerical interpreted languages such as MATLAB. For FWM estimations, instead of sampling the channels independently over a square grid, we suggest to sample only in the visible islands in Fig. 5a, thus by generating one channel index at random and by conditioning the generation of the remaining one.

We also found some advantage in using importance sampling (IS) [38], [39] by using different sampling distributions than the simplest uniform distribution adopted by the MC method. A first IS strategy, also called stratified-sampling (SS) [37], [40], is to estimate independently each nonlinear effect according to the terminology of Fig. 5. We propose to estimate XPM by sampling channel $n, n \neq n_{\text{CUT}}$, by a warped probability mass function (PMF) equal to $c/|n-n_{\text{CUT}}|^x$, with c a normalization constant. Such a sub-optimal PMF pushes more samples in channels closer to the CUT, following intuition, with a probability governed by

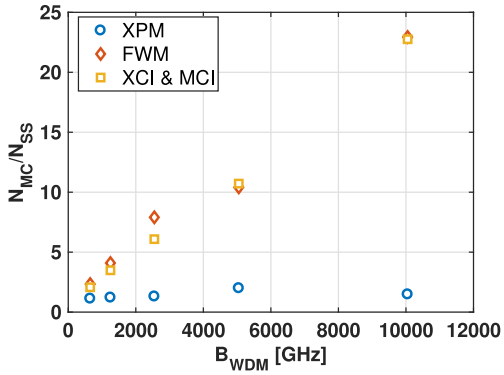

Fig. 9. Ratio of the number of samples with MC and SS to get a given relative error vs. WDM bandwidth.

x. The parameter x can be found by a least-squares fitting of the optimal SS-strategy $\sigma_n / \sum_n \sigma_n$ [37], with σ_n the standard deviation of the integrand within stratum n. The value of σ_n can be found by a short MC pre-run (e.g., a factor 100 shorter of the target number of samples) in the XPM islands. The least-squares fitting helps in smoothing the uncertainty of σ_n estimations. The FWM and XCI islands can be sampled in the same way by using the same PMF for each channel involved in the process. The warped PMF can be sampled by the inverse-transform sampling method [37].

The samples generation and the pre-run add a small overhead to the computational time. However, the proposed SS strategy allows to save samples, particularly for the most computationally intensive terms as FWM, XCI, and MCI. Fig. 9 shows the ratio of the number of samples required by MC and SS for the evaluation of each nonlinear effect at the same relative error of -30 dB. The figure refers to 10 spans including the Raman effect. We note significant savings for multi-channel effects, with increasing gains for increasing bandwidths. Even if not reported here, we observed similar gains without the Raman effect.

In the SS strategy, the variables $(\omega,\omega_1,\omega_2)$ within each island are sampled randomly by MC.[2] However, for homogeneous point-to-point links, it is possible to apply IS even within the islands of Fig. 5c, in particular for the GN-part of the EGN. The motivation comes from the observation that in such ultra-long links the phased-array term becomes a peaked function, whose random sampling is very inefficient by standard MC. The problem can be circumvented in the following way. The triple integral in (10) is weighted by the kernel function, hence by the phase matching coefficient $\Delta\beta$ between the four frequencies involved in the FWM process. If, for instance, we sample in (10) (h,ℓ,s) by a discrete uniform distribution and (ω,ω_1) by a uniform distribution as in the MC method, it is convenient to sample the remaining ω_2 over the regions where the kernel takes the largest absolute value. Such regions appear periodically at ω_2 values where the phased array (see Appendix B) has resonances, that is for $\Delta\beta = 2\pi k$ with k an integer. Fig. 10 sketches the idea, where we can observe: the kernel function (absolute squared

[2]It is important to have independent frequency samples in each island to ensure independent errors among islands.

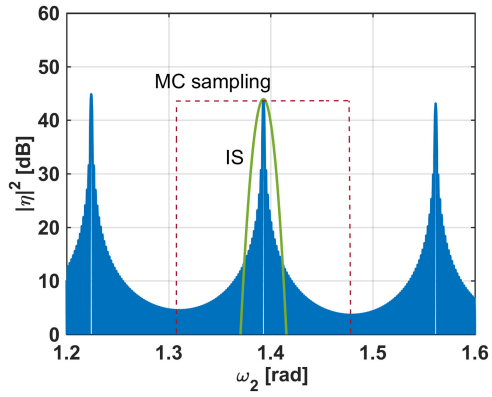

Fig. 10. Example of kernel and Monte Carlo (MC) or importance sampling (IS) distributions for sampling the frequency ω_2 normalized to the symbol rate. Both sampling strategies randomly select a lobe of the kernel.

TABLE I
COMPUTATIONAL TIME PER CHANNEL AND ACCURACY

B_{WDM} [GHz]	length [km]	std(a_{NL} [dB])		time [s]	
		MC	IS	MC	IS
50	1000	0.01	0.01	0.1	0.1
50	10000	0.04	0.04	0.1	0.1
5050	1000	0.08	0.07	1.1	1.5
5050	10000	0.24	0.04	1.1	1.5
10050	1000	0.10	0.09	1.1	1.5
10050	10000	0.37	0.06	1.1	1.5

value) versus the variable ω_2 for a given random choice of the remaining variables; the probability density function (PDF) of ω_2 with pure MC sampling (red dashed) and an example of importance sampling by using a Gaussian PDF for ω_2 (solid green). In practice, pure MC is equivalent to randomly selecting a given lobe of the phased-array, i.e., a value of k, with a discrete uniform distribution, and to randomly sample with a uniform distribution between the two neighboring lobes, as illustrated in the figure. On the other hand, IS concentrates probability around the peak. The best IS PDF replicates the kernel shape. However, it may be better to use simpler PDFs in favor of simpler algorithms, motivated by the observation that any PDF sufficiently concentrated around the peak, may give some computational advantages. In particular, we used a Gaussian PDF of variance equal to twice the variance of the absolute squared value of the kernel around a lobe.

As a reference, Table I shows some values of the uncertainty on a_{NL}, obtained by evaluating σ^2_{GN} with the proposed methods, and the corresponding computational time in different scenarios, by using MC or IS with 10^6 samples per effect. All the results are referred to the central WDM channel and are obtained with an INTEL XEON E5-2650 v4 2.20 GHz CPU based architecture. IS was applied only to XPM and FWM hexagons of Fig. 5 which experience the largest number of phased-array peaks.

The computational time is very short. If the NLI variance of each channel is required, the computation has to be repeated for each of them. However, since the a_{NL} is a smooth function over the WDM bandwidth, we suggest to estimate it just for a few channels and to use interpolation to find the missing values.

Please note that the accuracy of the proposed EGN depends on just one theoretical approximation, i.e., the perturbative solution of the Manakov equation, while the accuracy of the MC/IS is under control depending on the number of points. Hence, the proposed algorithm differs from approximations available in the literature, such as [41] where the authors were able to close integrals by further approximating them, for instance by neglecting FWM, XCI, and MCI.

VI. CONCLUSION

We investigated the challenges in numerical simulating ultra-wideband systems. We discussed three options, i.e., SSFM, VCG-SSFM, and the EGN model. We first analyzed the scaling properties of SSFM and its main bottlenecks, by investigating the implications of modern computing hardware based on GPU. We introduced VCG-SSFM as a faster algorithm able to capture just the dominant contributions of FWM. Finally, we discussed the EGN model and introduced advanced Monte Carlo techniques to speed up its numerical implementation.

As a general rule, we suggest to use SSFM when accuracy is mandatory, provided that the algorithm is set up correctly, while VCG-SSFM is the right solution to speed up simulations when a bias can be tolerated. We stress the importance of a correct SSFM setup, since in some scenarios we found more reliable to run a simulation *without* FWM rather than an SSFM accounting for FWM but with the wrong step size.

If the target of the simulation is the SNR of a basic receiver unable to exploit any correlation in the received signal, the EGN is definitely the best model to use. Thanks to the proposed Monte Carlo techniques, the EGN computational times including all nonlinear contributions can be reduced to the order of seconds, with excellent accuracy and without the need to limit the computation to just SPM and XPM as typically done in approximated closed-form formulas.

We also discussed the validity of the Manakov equation in ultra-wideband systems. We were able to show that it is a reliable equation by bounding the error of the corresponding NLI variance with respect to the most complete coupled-NLSE, showing a maximum error of 0.5 dB.

ACKNOWLEDGMENT

The authors would like to acknowledge valuable discussions with Amirhossein Ghazisaeidi about the EGN model.

APPENDIX A
CNLSE VS MANAKOV EQUATION

We want here to show that the error introduced by the Manakov equation is tolerable even in the extreme case of fibers having $L_{\text{corr}} \gg L_{\text{s}}$ with $L_{\text{s}} = \min(L_{\text{D}}, L_{\text{NL}})$. We make use of the GN model which can be extended to the CNLSE by treating the additional nonlinear term as a FWM term. The nonlinear term experienced by the generic x-polarization for the Manakov

TABLE II
MAIN CONTRIBUTIONS TO THE KERR EFFECT

	$\lvert A_x\rvert^2 A_x$	$\lvert A_y\rvert^2 A_x$	$A_y^2 A_x^*$
GN	$2P_{\mathrm{G}}$	P_{G}	$P_{\mathrm{G}}^{'} \leq P_{\mathrm{G}}$
FON	$4P_{\mathrm{F}}$	P_{F}	$P_{\mathrm{F}}^{'} \leq P_{\mathrm{F}}$

equation and the CNLSE is, respectively:

$$\mathcal{N}_x \overset{\text{(Manakov)}}{=} -j\gamma\frac{8}{9}\left(|A_x|^2+|A_y|^2\right)A_x$$
$$\mathcal{N}_x \overset{\text{(CNLSE)}}{=} -j\gamma\left[\left(|A_x|^2+\frac{2}{3}|A_y|^2\right)A_x+\frac{1}{3}A_y^2A_x^*\right] \quad (11)$$

Each nonlinear term appearing in (11) contributes to the EGN NLI variance through a GN contribution and an additive term accounting for the input-constellation higher-order moments, as for (9). In a UWB system, the higher-order noise contribution is essentially due to the fourth-order noise (FON) term [14].

Let P_{G} be the NLI power contributed by the second order moments of the input constellation (GN-model contribution) through the cross-polarization nonlinear term $|A_y|^2 A_x$ only. In absence of PMD the self-polarization nonlinear term $|A_x|^2 A_x$ yields a NLI power of $2P_{\mathrm{G}}$ because of twice combinations of coupling a symbol with its conjugate in the NLI variance evaluation [33]. The additional CNLSE term $A_y^2A_x^*$ is a FWM term that can be analytically managed as much as the cross-polarization one, but with a higher phase matching coefficient, depending on the amount of birefringence [2]. It contributes with a power $P_{\mathrm{G}}^{'}$ where $0 \leq P_{\mathrm{G}}^{'} \leq P_{\mathrm{G}}$. The two extrema occur with very small and very large beat length L_{B}, respectively. With similar arguments, the FON power P_{F} given by the term $|A_y|^2 A_x$ is 4 times smaller than the FON of $|A_x|^2 A_x$ [33], and larger than the term $A_y^2A_x^*$. By weighting such contributions with the coefficients in Table II we have the following variance of the NLI:

$$\sigma^2_{\mathrm{NLI}} = \frac{16}{81}(3P_{\mathrm{G}}+5P_{\mathrm{F}}) = \frac{16}{27}P_{\mathrm{G}}+\frac{80}{81}P_{\mathrm{F}}, \quad L_{\mathrm{corr}} \ll L_{\mathrm{s}} \quad (12)$$

$$\sigma^2_{\mathrm{NLI}} = \frac{1}{4}\left(\frac{23}{9}P_{\mathrm{G}}+\frac{40}{9}P_{\mathrm{F}}\right) = \frac{23}{36}P_{\mathrm{G}}+\frac{40}{36}P_{\mathrm{F}}, \quad L_{\mathrm{corr}}, L_{\mathrm{B}} \gg L_{\mathrm{s}} \quad (13)$$

$$\sigma^2_{\mathrm{NLI}} = \frac{1}{4}\left(\frac{22}{9}P_{\mathrm{G}}+\frac{40}{9}P_{\mathrm{F}}\right) = \frac{22}{36}P_{\mathrm{G}}+\frac{40}{36}P_{\mathrm{F}}, \quad L_{\mathrm{B}} \ll L_{\mathrm{s}} \ll L_{\mathrm{corr}} \quad (14)$$

where the GN factor $\frac{16}{27} = 2(\frac{1}{2})^3(\frac{8}{9})^2 3$ is commented in [33]. While (12) is well known, equations (13) and (14) are introduced here for the first time and represent the mentioned extreme cases of the CNLSE that we are trying to bound. We observe a minor impact of the beat length, and a maximum discrepancy between the Manakov equation and the CNLSE of $10\log_{10}(\frac{23}{36}\frac{27}{16}) = 0.32$ dB for the GN contribution and $10\log_{10}(\frac{40}{36}\frac{81}{80}) = 0.51$ dB for the FON.[3] Such numbers are quite small, with even a smaller impact at the best power maximizing the Q-factor, where any variations in σ^2_{NLI} in [dB] maps almost in a variation of $\sigma^2_{\mathrm{NLI}}/3$ [dB] on the best Q-factor [42].

The Raman effect has minor implications when moving from the CNLSE to the Manakov equation since it just weights by a factor f_{R} the memoryless nonlinear Kerr effect of (11), while it adds to the nonlinear effect a contribution with memory $-j\gamma f_{\mathrm{R}}(||\vec{A}||^2 \otimes H_{\mathrm{R}})\vec{A}$, with H_{R} the Raman filter [4], that remains identical in both equations. With the typical Raman factor of $f_{\mathrm{R}} \simeq 0.2$ [2], [4] the memoryless Kerr effect contributes with a NLI 0.2 dB higher in the CNLSE compared to the Manakov case.

Having observed minor discrepancies between CNLSE and Manakov equation, we now focus on PMD. If we push PMD at its extreme values, the Manakov nonlinearity experienced by channel $\vec{A}_1 = [A_{1x}, A_{1y}]$ due to channel $\vec{A}_2 = [A_{2x}, A_{2y}]$ is [4], [30]:

$$\mathcal{N}_x \overset{\text{(Manakov)}}{=} -j\gamma\frac{8}{9}\Big(|A_{1x}|^2A_{1x}+|A_{1y}|^2A_{1x} + 2|A_{2x}|^2A_{1x}+|A_{2y}|^2A_{1x} + A_{2x}A_{2y}^*A_{1y}\Big), \quad \text{PMD} \ll 1 \quad (15)$$

$$\mathcal{N}_x \overset{\text{(Manakov)}}{=} -j\gamma\frac{8}{9}\Big(|A_{1x}|^2A_{1x}+|A_{1y}|^2A_{1x}+\frac{3}{2}|A_{2x}|^2A_{1x} + \frac{3}{2}|A_{2y}|^2A_{1x}\Big), \quad \text{PMD} \gg 1 \quad (16)$$

If P_{GX} is the XPM variance induced by $|A_{2x}|^2 A_{1x}$, using similar arguments as in the previous case, $|A_{2y}|^2 A_{1x}$ and $A_{2x}A_{2y}^*A_{1y}$ contribute with P_{GX} and $P_{\mathrm{GX}}/2$, respectively [33]. Therefore, the large PMD case has a smaller average XPM nonlinear variance compared to the small PMD case by at most $10\log_{10}(\frac{4+1+1}{\frac{9}{4}+\frac{9}{4}}) = 1.25$ dB regarding the GN contribution [46]. Such a number represents an upper bound that is extremely difficult to reach. With practical values of PMD one should expect much smaller deviations. In any case, PMD helps to mitigate nonlinearity, as observed numerically [47] and experimentally [48] in the context of single-mode fibers, and theoretically in space-division multiplexing in [4]. Hence the worst-case analysis of the Kerr effect provided by the Manakov equation can be taken as a safely achievable bound of the nonlinearity for link design.

APPENDIX B
FIBER KERNEL

The fiber kernel weights the FWM process [14], [22], [23], [42] for any optical link, even those including Raman amplification [43]. For the sake of example, for a homogeneous N-span link with span length L_1 and lumped amplification it is $\eta(\omega,\omega_1,\omega_2) = \eta_1(\omega,\omega_1,\omega_2)\cdot\chi(\omega,\omega_1,\omega_2)$, where χ is the

[3] We assumed that cross-channel contributions have twice the weight of intra-channel terms within P_{G} and equal weight within P_{F}, which is an approximation for (13) with minor implications.

phased array function while η_1 is the kernel of a single span:

$$\chi(\omega,\omega_1,\omega_2)=\frac{1-e^{jN\Delta\beta\vartheta L_1}}{1-e^{j\Delta\beta\vartheta L_1}}$$

$$\eta_1(\omega,\omega_1,\omega_2)=\int_0^{L_1} e^{-\int_0^{\xi}\Delta\alpha\mathrm{d}\zeta}e^{j\Delta\beta\xi}\mathrm{d}\xi \tag{17}$$

where ϑ is the normalized residual dispersion per span and $\Delta\beta$ is the phase matching coefficient, equal to:

$$\begin{aligned}\Delta\beta &\triangleq \beta(\omega)-\beta(\omega+\omega_1)-\beta(\omega+\omega_2)+\beta(\omega+\omega_1+\omega_2)\\ &=\omega_1\omega_2\left[\beta_2+\frac{1}{2}\left(\omega_1+\omega_2+2\omega\right)\beta_3\right].\end{aligned} \tag{18}$$

In (17) $\Delta\alpha$ is defined as:

$$\begin{aligned}\Delta\alpha \triangleq -\frac{1}{2}\Big[&\alpha(\omega)-\alpha(\omega+\omega_1)-\alpha(\omega+\omega_2)\\ &-\alpha(\omega+\omega_1+\omega_2)\Big]\end{aligned}$$

where $\alpha(\omega)$ is a frequency and distance dependent attenuation accounting for the fiber attenuation and the possibly Raman effect [6]–[8]. When α is frequency independent it is $\Delta\alpha=\alpha$, such that the single-span kernel takes the simple expression:

$$\eta_1(\omega,\omega_1,\omega_2)\big|_{\alpha(\omega)=\text{const.}}=\frac{1-e^{-\alpha L_1}e^{j\Delta\beta L_1}}{\alpha-j\Delta\beta}. \tag{19}$$

In the other case, an analytical expression for the term $e^{-\int_0^{\xi}\Delta\alpha\mathrm{d}\zeta}$ in (17) can be found when the spectrum of $\alpha(\omega)$ is set by Raman as [36]:

$$e^{-\int_0^{\xi}\Delta\alpha\mathrm{d}\zeta}\simeq\frac{P_{\text{tot}}e^{-\alpha\xi}e^{-\frac{1}{2\pi}P_{\text{tot}}L_{\text{eff}}C_{\text{r}}(\omega+\omega_1+\omega_2)}}{\int_{-\infty}^{\infty}G_{\text{TX}}(\mu)e^{-\frac{1}{2\pi}P_{\text{tot}}L_{\text{eff}}C_{\text{r}}\mu}\frac{\mathrm{d}\mu}{2\pi}} \tag{20}$$

where P_{tot} is the total WDM launched power, $G_{\text{TX}}(\omega)$ is the power spectral density of the transmitted signal, and the Raman gain is approximated by a triangular profile of slope C_{r} [44].

Appendix C
Closed-Form Solution of the Nonlinear Step Without FWM

According to the Manakov equation in presence of the Raman effect, the nonlinear step experienced by the unique field $\vec{A}$ is:

$$\begin{aligned}&\frac{\partial\vec{A}}{\partial z}=-j\gamma'\mathcal{N}_{\text{R}}(\vec{A})\vec{A}\\ &\mathcal{N}_{\text{R}}(\vec{A})\triangleq\left(||\vec{A}||^2(1-f_{\text{R}})+f_{\text{R}}||\vec{A}||^2\otimes h_{\text{R}}(t)\right)\end{aligned} \tag{21}$$

where $\gamma'=\frac{8}{9}\gamma$, $h_{\text{R}}(t)$ is the impulse response of the Raman effect [2], $\otimes$ indicates convolution, and f_{R} (≈ 0.2) indicates the fractional contribution of Raman. The SSFM solves (21) by neglecting the signal depletion along the step:

$$\vec{A}(z)\simeq e^{-j\gamma'\mathcal{N}_{\text{R}}(\vec{A}(0))z}\vec{A}(0). \tag{22}$$

Such a result is exact in absence of the Raman effect, and an excellent approximation for very short step sizes as for UWB transmissions. For numerical purposes, it is more convenient to include the attenuation in the nonlinear step rather than in the linear step, with a minor modification of substituting z with $L_{\text{eff}}(z)$ in (22) [9].

By splitting the field $\vec{A}$ in groups, a closed form solution can be found in absence of Raman and FWM. In such a case, the generic group $\vec{A}_n$ experiences the following Kerr effect [29]–[31]:

$$\begin{aligned}\frac{\partial\vec{A}_n}{\partial z}=&-j\gamma'|\vec{A}_n|^2\vec{A}_n-j\frac{3}{2}\gamma'\sum_{k\neq n}|\vec{A}_k|^2\vec{A}_n\\ &-j\frac{1}{2}\gamma'\sum_{k\neq n}(\vec{a}_k\cdot\vec{\sigma})\vec{A}_n\end{aligned} \tag{23}$$

where $\vec{\sigma}$ is the Pauli tensor [21] and $\vec{a}_k\triangleq\vec{A}_k^{\dagger}\vec{\sigma}\vec{A}_k$ the Stokes representation of the Jones vector $\vec{A}_k$. The closed-form solution is [29]:

$$\vec{A}_n(z)=e^{-j\frac{\gamma'}{2}z\left(|\vec{A}_n(0)|^2+3\sum_{k\neq n}|\vec{A}_k(0)|^2\right)}\underline{\underline{e}}^{-\frac{j\gamma'}{2}z(\vec{s}_t(0)\cdot\vec{\sigma})}\vec{A}_n(0)$$

with $\vec{s}_t\triangleq\sum_k\vec{a}_k(0)$ the Stokes vector of the pivot, and $\underline{\underline{e}}^{(\cdot)}$ denotes matrix exponential, which can be computed for the generic real vector $\vec{s}=[s_1,s_2,s_3]$ as:

$$\underline{\underline{e}}^{-j(\vec{s}\cdot\vec{\sigma})}=\cos(|\vec{s}|)\sigma_0-j\frac{\sin(|\vec{s}|)}{|\vec{s}|}(\vec{s}\cdot\vec{\sigma})$$

where σ_0 is the identity matrix and:

$$\vec{s}\cdot\vec{\sigma}=\begin{bmatrix}s_1 & s_2-js_3\\ s_2+js_3 & -s_1\end{bmatrix}.$$

The inclusion of the Raman effect in (23) is more challenging. However, since the groups are likely to have much smaller bandwidth than $h_{\text{R}}(t)$, a polarization-averaged Raman effect can be included as a channel-dependent gain as in [5], [45].

References

[1] P. J. Winzer, "Would scaling to extreme ultra-violet or soft x-ray communications resolve the capacity crunch?," *J. Lightw. Technol.*, vol. 36, no. 24, pp. 5786–5793, Dec. 2018.

[2] G. P. Agrawal, *Nonlinear Fiber Optics*, 5th ed. New York, NY, USA: Academic, 2013.

[3] D. Marcuse, C. R. Menyuk, and P. K. A. Wai, "Application of the Manakov-PMD equation to studies of signal propagation in optical fibers with randomly varying birefringence," *J. Lightw. Technol.*, vol. 15, no. 9, pp. 1735–1746, Sep. 1997.

[4] C. Antonelli, M. Shtaif, and A. Mecozzi, "Modeling of nonlinear propagation in space-division multiplexed fiber-optic transmission," *J. Lightw. Technol.*, vol. 34, no. 1, pp. 36–54, Jan. 2016.

[5] D. Semrau, R. Killey, and P. Bayvel, "Achievable rate degradation of ultra-wideband coherent fiber communication systems due to stimulated Raman scattering," *Opt. Express*, vol. 25, no. 12, pp. 13024–13034, Jun. 2017.

[6] I. Roberts, J. M. Kahn, J. Harley, and D. Boertjes, "Channel power optimization of WDM systems following Gaussian noise nonlinearity model in presence of stimulated Raman scattering," *J. Lightw. Technol.*, vol. 35, no. 23, pp. 5237–5249, Dec. 2017.

[7] D. Semrau, R. I. Killey, and P. Bayvel, "The Gaussian noise model in the presence of inter-channel stimulated Raman scattering," *J. Lightw. Technol.*, vol. 36, no. 14, pp. 3046–3055, Jul. 2018.

[8] M. Cantono *et al.*, "On the interplay of nonlinear interference generation with stimulated Raman scattering for QoT estimation," *J. Lightw. Technol.*, vol. 36, no. 15, pp. 3131–3141, Aug. 2018.

[9] J. Shao, X. Liang, and S. Kumar, "Comparison of split-step fourier schemes for simulating fiber optic communication systems," *IEEE Photon. J.*, vol. 6, no. 4, pp. 1–15, Aug. 2014.

[10] O. V. Sinkin, R. Holzlöhner, J. Zweck, and C. R. Menyuk, "Optimization of the split-step fourier method in modeling optical fiber communications systems," *J. Lightw. Technol.*, vol. 21, no. 1, pp. 1–27, Jan. 2003.

[11] 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.

[12] S. Musetti, P. Serena, and A. Bononi, "On the accuracy of split-step fourier simulations for wideband nonlinear optical communications," *J. Lightw. Technol.*, vol. 36, no. 23, pp. 5669–5677, Dec. 2018.

[13] C. Francia, "Constant Step-size analysis in numerical simulation for correct four-wave-mixing power evaluation in optical fiber transmission systems," *IEEE Photon. Technol. Lett.*, vol. 11, no. 1, pp. 69–71, Jan. 1999.

[14] A. Mecozzi and R. Essiambre, "Nonlinear Shannon limit in pseudolinear coherent systems," *J. Lightw. Technol.*, vol. 30, no. 12, pp 2011–2024, Jun. 2012.

[15] A. Carena, G. Bosco, V. Curri, Y. Jiang, P. Poggiolini, and F. Forghieri, "EGN model of non-linear fiber propagation," *Opt. Express*, vol. 22, no. 13, pp. 16335–16362, Jun. 2014.

[16] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "Properties of nonlinear noise in long, dispersion-uncompensated fiber links," *Opt. Express*, vol. 21, no. 22, pp. 25685–25699, Oct. 2013.

[17] O. Golani, R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "Modeling the bit-error-rate performance of nonlinear fiber-optic systems," *J. Lightw. Technol.*, vol. 34, no. 15, pp. 3482–3489, Aug. 2016.

[18] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "Accumulation of nonlinear interference noise in multi-span fiber-optic systems," *Opt. Express*, vol. 22, no. 12, pp. 14199–14211, Jun. 2014.

[19] G. Saavedra *et al.*, "Experimental investigation of nonlinear signal distortions in ultra-wideband transmission systems," in *Proc. Opt. Fiber Commun.*, San Diego, CA, USA, 2017, Paper W1G.1.

[20] M. Cantono, D. Pilori, A. Ferrari, A. Carena, and V. Curri, "Observing the interaction of PMD with generation of NLI in uncompensated amplified optical links," in *Proc. Opt. Fiber Commun. (OFC)*, San Diego, CA, USA, 2018, Paper W1G.4.

[21] J. N. Damask, *Polarization Optics in Telecommunications*. New York, NY, USA: Springer, 2005.

[22] M. J. Ablowitz and T. Hirooka, "Managing nonlinearity in strongly dispersion-managed pulse transmission," *J. Opt. Soc. Amer. B*, vol. 19, no. 3, pp. 425–439, Mar. 2002.

[23] H. Louchet, A. Hodzic, K. Petermann, A. Robinson, and R. Epworth, "Simple criterion for the characterization of nonlinear impairments in dispersion-managed optical transmission systems," *IEEE Photon. Technol. Lett.*, vol. 17, no. 10, pp. 2089–2091, Oct. 2005.

[24] P. Johannisson and M. Karlsson, "Perturbation analysis of nonlinear propagation in a strongly dispersive optical communication system," *J. Lightw. Technol.*, vol. 31, no. 8, pp. 1273–1282, Apr. 2013.

[25] Q. Zhang and M. I. Hayee, "An SSF scheme to achieve comparable global simulation accuracy in WDM systems," *IEEE Photon. Technol. Lett.*, vol. 17, no. 9, pp. 1869–1871, Sep. 2005.

[26] Q. Zhang and M. I. Hayee, "Symmetrized split-step Fourier scheme to control global simulation accuracy in fiber-optic communication systems," *J. Lightw. Technol.*, vol. 26, no. 2, pp. 302–315, Jan. 2008.

[27] D. G. Manolakis and V. K. Ingle, *Applied Digital Signal Processing: Theory and Practice*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[28] S. Haykin, *Adaptive Filter Theory*, 5th ed. London, U.K.: Pearson, 2014.

[29] N. Rossi, P. Serena, and A. Bononi, "Symbol-rate dependence of dominant nonlinearity and reach in coherent WDM links," *J. Lightw. Technol.*, vol. 33, no. 14, pp. 3132–3143, Jul. 2015.

[30] 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.

[31] M. Karlsson and H. Sunnerud, "Effects of nonlinearities on PMD-induced system impairments," *J. Lightw. Technol.*, vol. 24, no. 11, pp. 4127–4137, Nov. 2006.

[32] P. Poggiolini, "The GN model of non-linear propagation in uncompensated coherent optical systems,"*J. Lightw. Technol.*, vol. 30, no. 24, pp. 3857–3879, Dec. 2012.

[33] P. Serena and A. Bononi, "A time-domain extended gaussian noise model," *J. Lightw. Technol.*, vol. 33, no. 7, pp. 1459–1472, Apr. 2015.

[34] P. Poggiolini *et al.*, "Analytical and experimental results on system maximum reach increase through symbol rate optimization," *J. Lightw. Technol.*, vol. 34, no. 8, pp. 1872–1885, Apr. 2016.

[35] R. Dar and P. J. Winzer, "Nonlinear interference mitigation: Methods and potential gain," *J. Lightw. Technol.*, vol. 35, no. 4, pp. 903–930, Feb. 2017.

[36] M. Zirngibl, "Analytical model of Raman gain effects in massive wavelength division multiplexed transmission systems," *Electron. Lett.*, vol. 34, no. 8, Apr. 1998, pp. 789–790.

[37] R. Korn, E. Korn, and G. Kroisandt, *Monte Carlo Methods and Models in Finance and Insurance*. Boca Raton, FL, USA: CRC Press, 2010.

[38] K. S. Shanmugam and P. Balaban, "A modified Monte-Carlo simulation technique for the evaluation of error rate in digital communication systems," *IEEE Trans. Commun.*, vol. 28, no. 11, pp. 1916–1924, Nov. 1980.

[39] J.-C. Chen, D. Lu, J. S. Sadowsky, and K. Yao, "On importance sampling in digital communications-Part I: Fundamentals," *IEEE J. Sel. Areas Commun.*, vol. 11, no. 3, pp. 289–299, Apr. 1993.

[40] N. Rossi, P. Serena, and A. Bononi, "Stratified-sampling estimation of PDL-induced outage probability in nonlinear coherent systems," *J. Lightw. Technol.*, vol. 32, no. 24, pp. 4303–4309, Dec. 2014

[41] D. Semrau, E. Sillekens, R. I. Killey, and P. Baylev, "A modulation format correction formula for the Gaussian noise model in the presence of inter-channel stimulated raman scattering," *J. Lightw. Technol.*, 2019, to be published, doi: 10.1109/JLT.2019.2929461

[42] P. Poggiolini, G. Bosco, A. Carena, V. Curri, Y. Jiang, and F. Forghieri, "The GN-model of fiber nonlinear propagation and its applications," *J. Lightw. Technol.*, vol. 32, no. 4, pp. 694–721, Feb. 2014.

[43] V. Curri, A. Carena, P. Poggiolini, G. Bosco, and F. Forghieri, "Extension and validation of the GN model for non-linear interference to uncompensated links using Raman amplification," *Opt. Express*, vol. 21, no. 3, pp. 3308–3317, Feb. 2013.

[44] D. Semrau, R. I. Killey, and P. Bayvel, "A closed-form approximation of the Gaussian noise model in the presence of inter-channel stimulated raman scattering,"*J. Lightw. Technol.*, vol. 37, no. 9, pp. 1924–1936, May 2019.

[45] Q. Lin and G. P. Agrawal, "Vector theory of stimulated Raman scattering and its application to fiber-based Raman amplifiers," *J. Opt. Soc. Amer. B*, vol. 20, no. 8, pp. 1616–1631, Aug. 2003.

[46] C. Antonelli, O. Golani, M. Shtaif, and A. Mecozzi, "Nonlinear interference noise in space-division multiplexed transmission through optical fibers," *Opt. Express*, vol. 25, no. 12, pp. 13055–13078, Jun. 2017.

[47] P. Serena, N. Rossi, O. Bertran-Pardo, J. Renaudier, A. Vannucci, and A. Bononi, "Intra-versus inter-channel PMD in linearly compensated coherent PDM-PSK nonlinear transmissions," *J. Lightw. Technol.*, vol. 29, no. 11, pp. 1691–1700, Apr. 2011.

[48] O. Bertran-Pardo *et al.*, "Demonstration of the benefits brought by PMD in polarization-multiplexed systems," in *Proc. ECOC*, Turin, Italy, Sep. 2010, Paper Th.10.E.4.