UNIVERSITÀ DI PARMA
ARCHIVIO DELLA RICERCA

University of Parma Research Repository

The Ergodic GN Model for Space-Division Multiplexing with Strong Mode Coupling

This is the peer reviewd version of the followng article:

Original
The Ergodic GN Model for Space-Division Multiplexing with Strong Mode Coupling / Serena, P.; Lasagni, C.; Bononi, A.; Antonelli, C.; Mecozzi, A.. - In: JOURNAL OF LIGHTWAVE TECHNOLOGY. - ISSN 0733-8724. - (2022), pp. 1-1. [10.1109/JLT.2022.3160207]

Availability:
This version is available at: 11381/2923808 since: 2022-05-18T16:39:58Z

Publisher:
Institute of Electrical and Electronics Engineers Inc.

Published
DOI:10.1109/JLT.2022.3160207

Terms of use:

Anyone can freely access the full text of works made available as "Open Access". Works made available

Publisher copyright

note finali coverpage

(Article begins on next page)

12 October 2024

The Ergodic GN Model for Space-Division Multiplexing with Strong Mode Coupling

Paolo Serena, *Member, IEEE,* Chiara Lasagni, *Student Member, IEEE,* Alberto Bononi, *Senior Member, IEEE,* Cristian Antonelli, *Fellow, OSA, Senior Member, IEEE,* and Antonio Mecozzi, *Fellow, OSA, Fellow, IEEE*

(Top-Scored Paper)

***Abstract*—The Gaussian noise (GN) model, thanks to its simplicity, is the most widely used model to describe the nonlinear interference (NLI) in fiber-optic systems. This paper gives an extension of the GN model to space-division multiplexed (SDM) systems operating in the regime of strong coupling between modes, where the understanding of the effect of spatial mode dispersion (SMD) on the accumulation of the NLI noise stands as an open problem of utmost importance. We show that SMD may significantly reduce the NLI power and that a reliable estimation can be done with the extended GN model in very fast times, of the order of seconds per channel. We derive a closed-form expression for the NLI power averaged with respect to the frequency-dependence random-mode coupling introduced by mode dispersion, and show that its random fluctuations around the average value are negligible for all practical purposes. The derived expression is proved to be very accurate in the relevant range of SMD values. Scaling properties of the NLI variance with respect to the main system parameters are also discussed to gain some physical insight.**

***Index Terms*—Space-division multiplexing, mode dispersion, GN model, nonlinear interference, perturbative models.**

I. Introduction

THE exponential growth of the fiber-optic network traffic necessitates parallelizing the information among all degrees of freedom [1]. Besides the well-known time and frequency dimensions, there is an enormous interest in exploiting the spatial dimension to meet the capacity demand, leading to space-division multiplexed (SDM) links. However, when SDM is implemented in multi-core and multi-mode fibers, the novel spatial dimension does not come for free, since the propagating modes are usually coupled to some extent, both by linear and nonlinear effects. In particular, the regime of strong mode-mixing on which we focus here implies the need for multiple-input multiple-output (MIMO) techniques to disentangle the information transmitted in the various modes, thereby increasing the receiver complexity. On the other hand, strong mixing has been shown to be beneficial in other ways, as it yields a slower accumulation of the spatial mode dispersion (SMD), proportional to the square-root of propagation distance, rather than to the propagation distance itself, as would be the case in the absence of strong-mode mixing [2], [3]. Moreover, strong mixing mitigates the nonlinear distortions of the transmitted signals [4], [5]. Unavoidably, the complexity of such high-capacity networks is becoming increasingly high, such that a design and an analysis based on the benchmark numerical algorithm, the split-step Fourier method (SSFM), may not be feasible [6]. As a reference, the complexity of accurate SSFM-based simulations scales at least with the fourth power of the wavelength division multiplexing (WDM) bandwidth [7]. Adding the spatial dimension further increases the complexity at least linearly with the number of modes, making the numerical effort likely unmanageable.

In this context, the search for analytical, yet approximated descriptions of the nonlinear distortions becomes necessary, and most research efforts focused on regular perturbation models of the nonlinear Kerr effect. This approach is particularly convenient, as it treats the nonlinear distortion in terms of additive noise, which is the most popular channel model analyzed in textbooks of digital communications. Therefore, within this framework, the noise impairing the propagated signal results from the sum of a nonlinear interference (NLI) induced by the Kerr effect and amplified spontaneous emission (ASE) coming from the optical amplifiers. The NLI can be modeled either in the time domain [8] or in the frequency domain [9], where each approach has its own strengths and weaknesses.

In this framework, several authors contributed to developing a Gaussian noise (GN) model, which by postulating Gaussian statistics of the transmitted digital signal can predict the NLI variance by either quick algorithms, e.g., [10], or by approximated closed-form formulas, e.g., [9]. The basic GN model has been extended to account for extra effects, such as the modulation format dependence of the NLI variance [8], [11], the effect of polarization mode dispersion (PMD) on digital back-propagation [12], the polarization-dependent loss (PDL) [13], and the stimulated Raman scattering (SRS) [14]. The GN model has been also extended to SDM links in [15]–[17]. However, the main limitation of such models is that they either neglect the effect of SMD on the nonlinear interaction between strongly coupled modes [15], or they account for

Manuscript received ******* **, 2021. This work was supported by the Italian PRIN 2017 project Fiber Infrastructure for Research on Space-Division Multiplexed Transmission (FIRST).

P. Serena, C. Lasagni, and A. Bononi are with the Department of Engineering and Architecture, Università degli Studi di Parma, Parma, 43124, Italy, and with the CNIT national laboratory of advanced optical fibers for photonics (FIBERS) (e-mail: paolo.serena@unipr.it; chiara.lasagni@unipr.it; alberto.bononi@unipr.it).

C. Antonelli and A. Mecozzi are with the Department of Physical and Chemical Sciences, Università degli Studi dell'Aquila, L'Aquila, 67100, Italy, and with the CNIT national laboratory of advanced optical fibers for photonics (FIBERS) (e-mail: cristian.antonelli@univaq.it; antonio.mecozzi@univaq.it).

Color versions of one or more of the figures in this paper are available online at http://xxxxxxxxxx.xxxx.xxx.

Digital Object Identifier xx.xxxx/JLT.xxxx.xxxxxxx

it only under certain simplifying assumptions [17]. Although neglecting mode dispersion in single-mode fibers (SMF) is justified by its small values in modern fibers, SMD can be orders of magnitude larger in SDM fibers [18], [19], thus playing a key role in setting the strength of the NLI.

In this work, which is an extended version of [20], we fill an important gap in the study of the interplay between SMD and NLI in SDM systems by introducing a discrete-time GN model for SDM links operating in the regime of strong mode mixing with arbitrary levels of SMD. We start from the GN model framework developed in [13] and add SMD to the fundamental theory. SMD makes the strong mixing between modes frequency-dependent, which renders the NLI power random in nature. Here we evaluate its expected value and dub the resulting GN model as the *ergodic GN model*, following a similar definition adopted for the Shannon's (ergodic) capacity [21]. As a major result, we show that SMD changes the four-wave mixing (FWM) efficiency, that can be easily exploited to numerically compute the NLI variance formulas that we show to be accurate in the relevant range of SMD values. We validate the proposed theory against SSFM results and show that the random fluctuations of the NLI power around its average value are practically negligible.

The proposed ergodic GN model can be an effective tool to design and analyze future SDM networks [22] operating in the regime of strong mixing between all spatial modes, as is the case in SDM systems based on coupled-core multi-core fibers.

The paper is organized as follows. In Section II we introduce the notation and the key ingredients of the extended GN model. In Section III we show the key results of our theory and the main numerical validations by an extensive comparison with SSFM-based simulations. In Section IV we compute the ergodic FWM kernel and the corresponding NLI variance. Section V focuses on an approximated closed-form expression of the NLI variance with SMD. In Section VI we discuss the main scaling properties of the interaction between SMD and the Kerr effect. In Section VII we finally draw our conclusions.

II. Physical model

We consider optical fibers with $2N$ strongly-coupled modes, where by N we denote the number of spatial modes and the factor of two accounts for polarization degeneracy. We adopt the bra-ket notation, hence with $|A(z,t)\rangle$ indicating a column vector of $2N$ components containing the complex envelopes of the modes at coordinate z and time t. The corresponding Fourier transform, expressed in engineering notation, is $|\tilde{A}(z,\omega)\rangle$, with ω the angular frequency. Note that $N=1$ is the case of SMF.

We assume the modes propagate in optical fibers in the strong coupling regime, thus satisfying the multi-component Manakov equation [5]:

$$\frac{\partial |A\rangle}{\partial z} = -\frac{\alpha}{2}|A\rangle - \frac{\mu \mathbf{B}}{2N}\frac{\partial |A\rangle}{\partial t} + j\frac{\beta_2}{2}\frac{\partial^2 |A\rangle}{\partial t^2} - j\gamma\kappa \langle A|A\rangle |A\rangle \tag{1}$$

where the second term on the right-hand side of the equality accounts for mode dispersion, with the scalar coefficient μ setting the SMD strength and the matrix $\mathbf{B}=\mathbf{B}(z)$ producing random mode coupling. By α we denote the fiber attenuation, β_2 the mode-averaged chromatic dispersion (CD) coefficient, γ the nonlinear coefficient, and κ the Manakov correction factor. We neglect mode-dependent loss (MDL) and ultra-wideband effects such as the Raman effect and the frequency dependence of the fiber parameters. We assume the correlation length of the random coupling process to be sufficiently small such that the matrix $\mathbf{B}\mathrm{d}z$ is a Wiener process, see Appendix A. Notice that the phenomenon of mode dispersion considered throughout this work originates from the fact that the local principle states are quasi-degenerates, namely they have similar yet different group velocities but at the same time they are strongly coupled with each other owing to the perturbations. This phenomenon differs substantially from that of modal dispersion between non-degenerate modes or mode groups that are either uncoupled between them or weakly coupled and propagate with distinctively different group velocities.

We follow the notation introduced in [13], thus we write the transmitted WDM signal $|A(0,t)\rangle$ as a comb of linearly modulated digital signals:

$$|A(0,t)\rangle = \sum_{\mathbf{n}} a_{\mathbf{n}} |G_{\mathbf{n}}(0,t)\rangle \tag{2}$$

where the index vector $\mathbf{n}$ should be read as:

$$\mathbf{n} = [\underbrace{n_1}_{\text{time}},\ \underbrace{n_2}_{\text{frequency}},\ \overbrace{n_3}^{\text{mode}}]\ . \tag{3}$$

$a_{\mathbf{n}}$ is a scalar data symbol per *(time, frequency, mode)* channel use. The summation $\sum_{\mathbf{n}}$ must be read as over all possible values of $\mathbf{n}$. The shaping functions are:

$$|G_{\mathbf{n}}(0,t)\rangle \triangleq p(t-n_1T)e^{j\Omega_{n_2}t}|n_3\rangle \tag{4}$$

with p the supporting pulse, Ω_{n_2} the carrier frequency of the n_2-th channel, $|n_3\rangle$ a unit-vector identifying the n_3-th scalar (space and polarization) polarization state, and T the symbol time, possibly a function of $n_{2,3}$.

The shaping functions are assumed to be orthonormal, such that a given symbol of $|A\rangle$ can be extracted by a simple cross-correlation detector, as per the matched filter detection:

$$\sum_{\mathbf{n}} a_{\mathbf{n}} \int_{-\infty}^{\infty} \langle \tilde{G}_{\mathbf{i}}(\omega)|\tilde{G}_{\mathbf{n}}(\omega)\rangle \frac{\mathrm{d}\omega}{2\pi} = a_{\mathbf{i}}\ . \tag{5}$$

Such signals propagate in the optical fibers experiencing Kerr effect, linear crosstalk, and SMD, as sketched in Fig. 1.

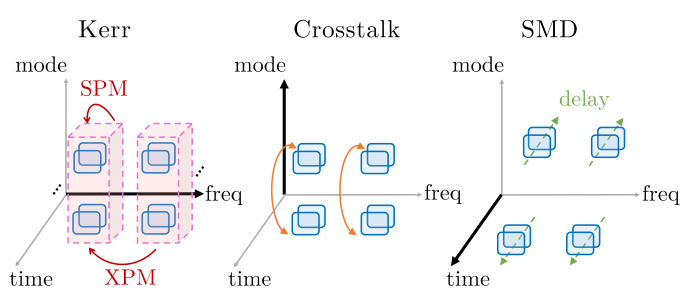

Fig. 1. Example of effects experienced by two frequency channels along all the modulated dimensions. Each blue square describes a channel use in the (time, frequency, mode) space.

Under perturbative assumptions, the discrete-time channel model of the system, including matched filter detection and average carrier phase recovery, is $y_{\mathbf{i}} = a_{\mathbf{i}} + w_{\mathbf{i}} + n_{\mathbf{i}}$, with $w_{\mathbf{i}}$ accounting for ASE noise and $n_{\mathbf{i}}$ for the NLI [8], [13]:

$$n_{\mathbf{i}} = -j \sum_{\mathbf{k},\mathbf{m},\mathbf{n}} a^*_{\mathbf{k}} a_{\mathbf{m}} a_{\mathbf{n}} \mathcal{X}_{\mathbf{kmni}} \,. \tag{6}$$

The scalar coefficient $\mathcal{X}_{\mathbf{kmni}}$ weighs the nonlinear interaction between symbols $(\mathbf{k},\mathbf{m},\mathbf{n},\mathbf{i})$, with $\mathbf{i}$ the target symbol. Its expression with strongly coupled modes is a generalization of the scalar case shown in [8]:

$$\mathcal{X}_{\mathbf{kmni}} = \gamma\kappa \int_0^{L_\mathrm{t}} f(z) \int_{-\infty}^{\infty} \langle G_{\mathbf{k}}(z,t)|G_{\mathbf{m}}(z,t)\rangle \times \langle G_{\mathbf{i}}(z,t)|G_{\mathbf{n}}(z,t)\rangle \,\mathrm{d}t\mathrm{d}z \tag{7}$$

with L_t the link length, $f(z)$ the power-loss profile of the link, and $|G_{\mathbf{n}}(z,t)\rangle$ the pulse shapes impaired by linear dispersive effects up to coordinate z. To derive (7) we exploited the observation that from scalar to vector space one needs to substitute $G_{\mathbf{k}}(z,t) \to |G_{\mathbf{k}}(z,t)\rangle$ and $G^*_{\mathbf{k}}(z,t) \to \langle G_{\mathbf{k}}(z,t)|$.

In our framework, the NLI is doubly stochastic, depending on both the random symbols and the random birefringence. In the following, we will use the term variance/covariance with reference to the random symbols only. The variance/covariance are thus random variables as well, depending on the remaining random birefringence.

If the receiver is not able to exploit any correlation in the detected signal, the NLI can be safely treated as signal-independent additive noise and its covariance can be evaluated with the GN model theory when the transmitted symbols $a_{\mathbf{i}}$ are zero-mean complex Gaussian random variables. In this framework, with P_p the power per channel in a given polarization of a spatial mode, hence $P = 2P_\mathrm{p}$ in a given mode, we have the following main result [13]:

$$\mathbb{E}_a\left[n_{\mathbf{i}} n^*_{\mathbf{j}}\right] = P_\mathrm{p}^3 \sum_{\mathbf{k},\mathbf{m},\mathbf{n}} \mathcal{X}_{\mathbf{kmni}} \left(\mathcal{X}^*_{\mathbf{kmnj}} + \mathcal{X}^*_{\mathbf{knmj}}\right) \tag{8}$$

where the expectation $\mathbb{E}_a$ is with respect to the transmitted symbols.

If we neglect nonlinear signal-ASE interaction, the signal-to-noise ratio (SNR) is therefore [23], [24]

$$\mathrm{SNR} = \frac{P}{\sigma^2_{\mathrm{ASE}} + \sigma^2_{\mathrm{NLI}}} \tag{9}$$

with $\sigma^2_{\mathrm{NLI}} \equiv 2\mathbb{E}_a\left[|n_{\mathbf{i}}|^2\right]$, the factor 2 accounting for the two polarizations, and σ^2_{ASE} the variance of ASE in the mode and bandwidth of interest.

The tensor (7) accounts for all possible interactions among symbols that we now investigate. Let us simplify the notation with $\mathbf{kmni}_j \triangleq k_j m_j n_j i_j$, with $j = 1,2,3$ according to (3). The set $\mathbf{kmni}_1$ identifies different classes of temporal pulse collisions, manifesting in phase-, hyper-polarization rotations or circular noise, with similar observations made in [25]. The main difference with [25] is that SMD modifies the interaction during the collision, adding another length scale to the process.

The set $\mathbf{kmni}_2$ classifies FWM among carrier frequencies, which can reduce to self-phase modulation (SPM), cross-phase modulation (XPM), cross-channel interference (XCI), or multi-channel interference (MCI) [7], [11]. To better investigate them, it is convenient to express the tensor in terms of $|\tilde{G}_{\mathbf{n}}(z,\omega)\rangle$. By exploiting the Rayleigh theorem, the temporal integral in (7) can be transformed into frequency integrals, yielding:

$$\mathcal{X}_{\mathbf{kmni}} = \gamma\kappa \int_0^{L_\mathrm{t}} \iiint_{-\infty}^{\infty} f(z) \langle \tilde{G}_{\mathbf{k}}(z,\omega+\omega_1+\omega_2)|\tilde{G}_{\mathbf{m}}(z,\omega+\omega_2)\rangle \times \langle \tilde{G}_{\mathbf{i}}(z,\omega)|\tilde{G}_{\mathbf{n}}(z,\omega+\omega_1)\rangle \frac{\mathrm{d}\omega}{2\pi}\frac{\mathrm{d}\omega_1}{2\pi}\frac{\mathrm{d}\omega_2}{2\pi}\mathrm{d}z \,. \tag{10}$$

The shaping functions at coordinate z are $|\tilde{G}_{\mathbf{n}}(z,\omega)\rangle = \mathbf{T}(z,\omega)\,|\tilde{G}_{\mathbf{n}}(0,\omega)\rangle$, with $\mathbf{T}(z,\omega)$ a $2N \times 2N$ matrix accounting for all linear dispersive impairments accumulated from coordinate 0 to z. Such a matrix can be separated into a scalar and a mode-dependent contribution:

$$\mathbf{T}(z,\omega) = e^{-j\int_0^z \beta(x,\omega)\mathrm{d}x}\mathbf{U}(z,\omega) \tag{11}$$

with β the mode-averaged propagation constant. The matrix $\mathbf{U}$ is a unitary matrix accounting for frequency-dependent random-mode coupling, that is, mode dispersion. The tensor (10) thus becomes:

$$\mathcal{X}_{\mathbf{kmni}} = \gamma\kappa \int_0^{L_\mathrm{t}} f(z) \iiint_{-\infty}^{\infty} e^{-j\int_0^z \Delta\beta(x,\omega,\omega_1,\omega_2)\mathrm{d}x} \langle \tilde{G}_{\mathbf{k}}(0,\omega+\omega_1+\omega_2)|\mathbf{P}(z,\omega+\omega_2,\omega_1)|\tilde{G}_{\mathbf{m}}(0,\omega+\omega_2)\rangle \times \langle \tilde{G}_{\mathbf{i}}(0,\omega)|\mathbf{P}^\dagger(z,\omega,\omega_1)|\tilde{G}_{\mathbf{n}}(0,\omega+\omega_1)\rangle \frac{\mathrm{d}\omega}{2\pi}\frac{\mathrm{d}\omega_1}{2\pi}\frac{\mathrm{d}\omega_2}{2\pi}\mathrm{d}z \tag{12}$$

with $\dagger$ indicating transpose-conjugate, $\Delta\beta = \beta(\omega+\omega_1) + \beta(\omega+\omega_2) - \beta(\omega+\omega_1+\omega_2) - \beta(\omega)$ the phase matching coefficient, and

$$\mathbf{P}(z,\omega,\omega_1) \triangleq \mathbf{U}^\dagger(z,\omega+\omega_1)\mathbf{U}(z,\omega) \,. \tag{13}$$

Although (12) yields the impression to involve more integrations than (7), note that, contrary to (12), the shaping functions in (7) depend on coordinate z, hence their evaluation calls for extra integrations through convolutions.

We find it convenient to introduce the FWM tiling factor $\mathcal{M}_{\mathbf{kmni}}$:

$$\mathcal{M}_{\mathbf{kmni}} \triangleq \tilde{G}^*_{k_1k_2}(\omega+\omega_1+\omega_2-\Omega_{k_2})\tilde{G}_{m_1m_2}(\omega+\omega_2-\Omega_{m_2}) \times \tilde{G}^*_{i_1i_2}(\omega-\Omega_{i_2})\tilde{G}_{n_1n_2}(\omega+\omega_1-\Omega_{n_2}) \tag{14}$$

with $\tilde{G}_{i_1i_2}$ defined implicitly by $|\tilde{G}_{\mathbf{i}}(0,\omega)\rangle \triangleq \frac{1}{\sqrt{T}}\tilde{G}_{i_1i_2}(\omega)\,|i_3\rangle$. The FWM tiling factor takes its name from the hexagonal regions that it identifies in the plane (ω_1,ω_2) at a given ω, e.g., see Fig. 2 for $\omega = 0$ and rectangular $\tilde{G}_{i_1i_2}(\omega)$ [9], [11]. It also satisfies the symmetry relation $\mathcal{M}_{\mathbf{kmni}}(\omega,\omega_1,\omega_2) = \mathcal{M}_{\mathbf{knmi}}(\omega,\omega_2,\omega_1)$.

Equation (12) can be rearranged in:

$$\mathcal{X}_{\mathbf{kmni}} = \gamma\kappa \iiint_{-\infty}^{\infty} \eta_{\mathbf{kmni}}(\omega,\omega_1,\omega_2)\mathcal{M}_{\mathbf{kmni}}(\omega,\omega_1,\omega_2)\frac{\mathrm{d}\omega}{2\pi}\frac{\mathrm{d}\omega_1}{2\pi}\frac{\mathrm{d}\omega_2}{2\pi} \tag{15}$$

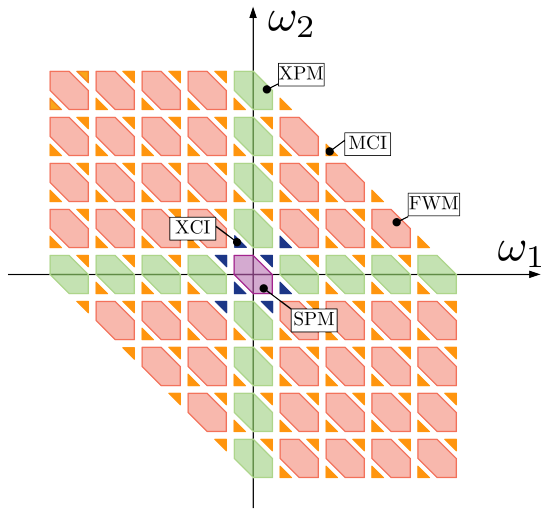

Fig. 2. Example of regions of non-zero tiling factor $\mathcal{M}_{\mathbf{kmni}}$ at $\omega = 0$.

with $\eta_{\mathbf{kmni}}$ a tensor weighting the FWM process:

$$\eta_{\mathbf{kmni}} \triangleq \int_0^{L_t} f(z) e^{-j\int_0^z \Delta\beta(x,\omega,\omega_1,\omega_2)\mathrm{d}x} \times \langle k_3|\mathbf{P}(z,\omega+\omega_2,\omega_1)|m_3\rangle \langle i_3|\mathbf{P}^\dagger(z,\omega,\omega_1)|n_3\rangle \,\mathrm{d}z\,. \tag{16}$$

In absence of SMD (15) is usually more convenient for numerical purposes than (7) because the highly oscillatory integral along z in (16) can be closed [24], [26], [27]. This is not the case with SMD, since $\mathbf{U}$, and thus $\mathbf{P}$, follow a complex and random behavior along z. It turns out that the computation of (8), as well as its interpretation, becomes extremely challenging with SMD, such that its advantage against the classical SSFM may be questionable at the present stage.

The set $\mathbf{kmni}_3$ identifies nonlinear modal crosstalk. Notice that in the absence of SMD the matrix $\mathbf{P}$ in (16) reduces to the identity matrix, thereby implying that in this case the inner products under integration differ from zero only if $k_3 = m_3$ and $n_3 = i_3$, as observed in [28]. Hence, the variance contribution $\mathcal{X}_{\mathbf{kmni}}\mathcal{X}^*_{\mathbf{knmj}}$ in (8) yields only one valid combination when the modal indexes are all identical, while $\mathcal{X}_{\mathbf{kmni}}\mathcal{X}^*_{\mathbf{kmnj}}$ is $2N$ degenerate in (k_3, m_3). Therefore, without SMD and at fixed P_p, the NLI variance of a given polarization is $\frac{2N+1}{2}\kappa^2$ higher than a scalar propagation without any polarization effect. Again, with SMD such a degeneracy is broken thus changing the mode interaction. In particular, the modes that participate to FWM interactions are not constrained to be identical in pairs.

The expression of the tensor $\mathcal{X}$ entails a dependence of the NLI variance on the specific path along z of the random-coupling matrix $\mathbf{B}(z)$, which makes it random in nature. One may be tempted to average the tensor (12) with respect to the birefringence realizations, following similar arguments adopted in the derivation of the Manakov equation, and then use it in the NLI variance (8). This is generally wrong unless the correlation length of SMD is much shorter than all the other characteristic lengths of the Manakov equation, a condition that is rarely satisfied. For instance, with $N = 2$, SMD coefficient $\eta_{\mathrm{SMD}} = 3$ ps/$\sqrt{\mathrm{km}}$, channel spacing $\Delta f = 50$ GHz, the SMD length of the next-neighbor interfering channel in frequency is $L_{\mathrm{SMD}} = 6.6$ km (see (51)), a value that is greater than the typical correlation length of a fiber, which is expected to be of the order of 10–100 meters in strongly-coupled fibers. However, averaging directly the NLI variance (8) without intermediate averages is legitimate, since the NLI variance is the final target of our theory. In what follows we adopt this approach, which proves to be quite accurate as a result of the fact that the NLI power randomly fluctuates only modestly around its average value.

III. Results

For the reader's convenience we anticipate here the major results of our work. Our main goal is evaluating the mean NLI variance. From (8), after normalizing to P_p^3 and the number of polarizations, the problem can be split in two contributions:

$$\mathbb{E}\left[\frac{\sigma^2_{\mathrm{NLI}}}{2P_\mathrm{p}^3}\right] = \mathbb{E}\Bigg[\underbrace{\sum_{\mathbf{k,m,n}} \mathcal{X}_{\mathbf{kmni}}\mathcal{X}^*_{\mathbf{kmni}}}_{\mathcal{I}_1} + \underbrace{\sum_{\mathbf{k,m,n}} \mathcal{X}_{\mathbf{kmni}}\mathcal{X}^*_{\mathbf{knmi}}}_{\mathcal{I}_2}\Bigg] \tag{17}$$

where $\mathbb{E}$ indicates expectation with respect to the random mode coupling. The two contributions are degenerate only in the absence of SMD. In the general case, see Section IV, they can be expressed in terms of two FWM efficiency kernels $|\eta^{(\ell)}|^2$:

$$\mathcal{I}_\ell = (\gamma\kappa)^2 \sum_{k_2,m_2,n_2=1}^{N_{\mathrm{ch}}} \iiint_{-\infty}^{\infty} \times |\eta^{(\ell)}(\omega,\omega_1,\omega_2)|^2 \, |\mathcal{M}_{\mathbf{kmni}}(\omega,\omega_1,\omega_2)|^2 \frac{\mathrm{d}\omega_1}{2\pi}\frac{\mathrm{d}\omega_2}{2\pi}\frac{\mathrm{d}\omega}{2\pi} \tag{18}$$

where $\ell = 1, 2$, the number of frequency channels is N_{ch}, and

$$|\eta^{(\ell)}|^2 \triangleq \begin{cases} \sum\limits_{k_3,m_3,n_3} |\eta_{\mathbf{kmni}}|^2, & \ell = 1 \\ \sum\limits_{k_3,m_3,n_3} \eta_{\mathbf{kmni}}\eta^*_{\mathbf{knmi}}, & \ell = 2\,. \end{cases} \tag{19}$$

We remind that $|\eta^{(1)}|^2 = |\eta^{(2)}|^2$ only without SMD. Equation (18) is similar to the GN-model reference formula (GNRF, e.g., [9, eq. (1)]) weighted by the matched filter frequency response. A major result of our work is an expression for $\mathbb{E}[|\eta^{(\ell)}|^2]$. In Section IV we will compute it analytically, obtaining:

$$\mathbb{E}\left[|\eta^{(\ell)}|^2\right] = m_\ell\left[(1+c_\ell)E_{\rho_1} + (1-c_\ell)E_{\rho_2}\right], \quad \ell = 1, 2 \tag{20}$$

where $m_\ell, c_\ell, \rho_\ell$ are elementary functions, while E_ρ is:

$$E_\rho = \frac{1}{\alpha}\mathrm{Re}\left[\frac{1-e^{-(\alpha-\rho-i\Delta\beta)L}}{\alpha-\rho-i\Delta\beta} + \frac{e^{-2\alpha L} - e^{-(\alpha-\rho-i\Delta\beta)L}}{\alpha+\rho+i\Delta\beta}\right] \tag{21}$$

with ρ a function of μ. The relation between μ and other SMD parameters, such as the SMD coefficient η_{SMD}, is discussed in Appendix C. For more details about the remaining parameters of (21) we refer the reader to Appendix B.

Equation (17) is the proposed ergodic GN model. Despite the presence of three integrals, its computation does not require advanced tools since it can be simply and quickly estimated by Monte Carlo integrations [10] with the desired accuracy, with computational times comparable to the SMF case.

Another key result of this work is an approximated expression of the NLI variance of XPM valid for moderate values of SMD, luckily the ones of practical interest. XPM corresponds to two non-degenerate sets $\mathbf{kmni}_2 = \mathbf{kkii}_2$ and $\mathbf{kmni}_2 = \mathbf{kiki}_2$. We get it by using an inter-channel approximation of the SMD, i.e., SMD is visible only by the channel carriers. While we will present its analytical derivation in Section V, we anticipate here the main result for two channels spaced away $\Delta\omega$:

$$\sigma^2_{\text{XPM}} = \frac{2N+1}{2N}\Bigg((2N+1)\sigma^2_{\text{XPM},1}(\alpha) + \frac{(2N-1)\left(\alpha+\frac{\Delta\omega^2\mu^2}{N}\right)}{\alpha}\sigma^2_{\text{XPM},1}\left(\alpha+\frac{\Delta\omega^2\mu^2}{N}\right)\Bigg). \quad (22)$$

Here $\sigma^2_{\text{XPM},1}$ is the reference XPM variance usually analyzed in the literature in an equivalent scalar propagation. While we invite the reader to read Section V for more details, we observe that the presence of SMD, weighted through the parameter μ, curiously acts like an extra-attenuation term to the fiber loss coefficient α, consistent with the Stratonovich to Ito conversion, see Appendix A. This is visible even in the complete model in (21) through the term ρ.

Throughout this work we assume for $\sigma^2_{\text{XPM},1}$ the closed-form expression derived in [9], adapted so as to account for the mode multiplicity. The explicit expression of $\sigma^2_{\text{XPM},1}$ is given in Section V. A MATLAB script for the numerical evaluation of (22) is provided in the supplementary material of this manuscript.

A. Numerical validation

We tested the accuracy of the key results by simulating different strongly coupled SDM transmissions through the SSFM. To test XPM we focused on a two-channel[1] system since the superposition principle works for the XPM variance. General parameters common to all simulations were the following. The interfering channel had power $P_{\text{i}} = 0$ dBm, while the channel under test (CUT) had a power $P_{\text{CUT}} = -30$ dBm to remove its SPM, 0 dBm otherwise. The data were complex Gaussian random variables, consistently with the assumptions of the GN model, for a total of 65536 random symbols. The digital signal was made of sinc pulses, transmitted at a symbol rate of $R = 49$ Gbd, with channel spacing Δf. The optical fibers worked in a strongly coupled regime with $N = 2$ or 4 modes. The fiber parameters, if not explicitly stated, were: span length 100 km; attenuation coefficient 0.2 dB/km; dispersion 17 ps/(nm·km); nonlinear coefficient $\gamma = 1.2668/N$ in $1/(\text{W}\cdot\text{km})$ [5, eq. (66)]; Manakov correction factor $\kappa = \frac{4}{3}\frac{2N}{2N+1}$ [5, eq. (65)]; effective area $A_{\text{eff}} = 80\ \mu\text{m}^2$. They were modeled by the symmetrized SSFM applied to (1), with a first step h_1 set up to have a worst-case channel walk-off of $1/10$ of the symbol time. The step was then updated by using the constant local error criterion (CLE) [29]. The SMD was implemented by means of the waveplate model as in [2], where each waveplate had length $2h_1$ with deterministic delays, as discussed in Appendix C. The number of waveplates was set to 10000. Each span was followed by an ideal noiseless optical amplifier. At the receiver we first removed all the linear effects accumulated during the link, then after matched filter detection and sampling, we removed the average phase to the detected symbols. We finally estimated the NLI variance by subtracting the transmitted symbols to the received ones. All the variances reported in the following are per polarization.

Figure 3 shows the XPM variance, normalized to $P_{\text{CUT}}P_{\text{i}}^2$, versus the SMD coefficient η_{SMD} in a single span at a channel spacing of $\Delta f = 100$ GHz. See Appendix C for more details about the definition of the SMD coefficient. In this set of simulations, we used 500 different random realizations of the waveplates. The blurred regions are two-dimensional histograms of the SSFM results of the average SNR among modes. Light regions indicate many estimations falling there, while dark regions account for rare cases.

The solid line is the ergodic GN model (17) evaluated by Monte Carlo integrations of (18) with the proposed ergodic kernel (20). We sampled the frequency integrals with 10^6 random samples as in [10]. The dashed line is the same but with only inter-channel SMD. We observe a very good match between the complete model and the SSFM, while the inter-channel approximation fails at large SMD values, showing saturation. However, (22) has an excellent match with SSFM up to SMD of ~ 5 ps/$\sqrt{\text{km}}$, which is a practical value for several deployed fibers [19], [30]. We observe a resonance around an SMD of ~ 35 ps/$\sqrt{\text{km}}$, followed by a monotonic decrease for very large values of the SMD due to the beneficial effect of SMD in mitigating the accumulation of the Kerr nonlinearity. We will shed light on this point in Section VI.

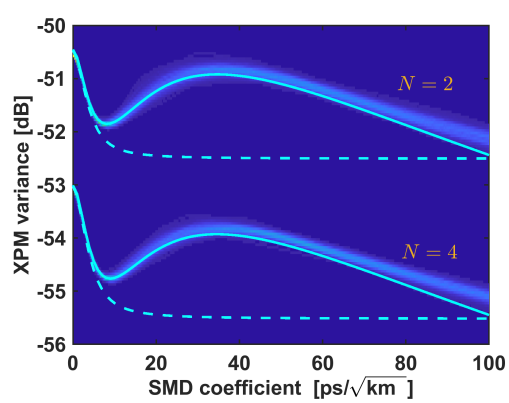

Fig. 3. XPM variance vs SMD coefficient η_{SMD}. Two-channel simulation at spacing $\Delta f = 100$ GHz in a fiber of dispersion 17 ps/(nm·km) supporting N strongly coupled modes. Blurred regions: SSFM. Lines: ergodic GN model (17). Dashed-lines: inter-channel SMD approximation (22).

We then varied the channel spacing at two different, yet

[1]Hereafter, with two-channel we mean two frequency channels, each with $2N$ strongly coupled space and polarization modes. The number of signals is thus $4N$.

relevant, values of SMD. The results are depicted in Fig. 4, including the ergodic GN model (17) and the SSFM results. We observe again an excellent match between theory and simulations. Even if difficult to see, the figure also shows max-min error bars for the SSFM results over the 1500 seeds tested in the simulation. The very small error bars are an indication that using the ergodic variance is more than enough to capture the effect of the SMD, without the need to focus on the SNR randomness. Such randomness is more evident at high SMD > 20 ps/$\sqrt{\text{km}}$, as visible in Fig. 3.

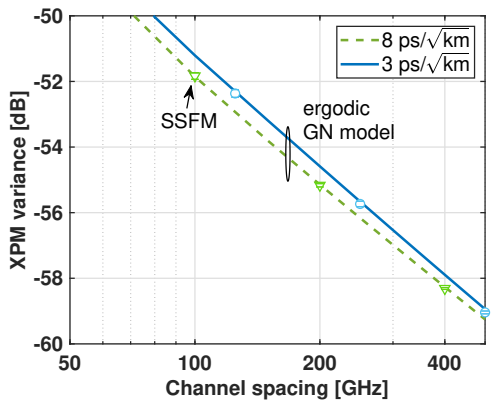

Fig. 4. XPM variance vs channel spacing of a two-channel simulation. Symbols: SSFM. Lines: ergodic GN model (17). $N = 2$ modes.

Having tested the ergodic GN model in a single-span, we moved to test its effectiveness in a multi-span scenario of 5 spans. The results are depicted in Fig. 5. Here the channel spacing was $\Delta f = 100$ GHz, and we included also the SMF case for the sake of completeness. The SSFM results are reported by error bars accounting for the 1000 random seeds of the birefringence. The match is again excellent. We also tested the same link at 10 spans, not reported here, observing a rigid worsening of the XPM variance of, almost exactly, 3 dB, an indication that the spatial disaggregation of XPM is even more true in the presence of SMD, as commented in Section V-A.

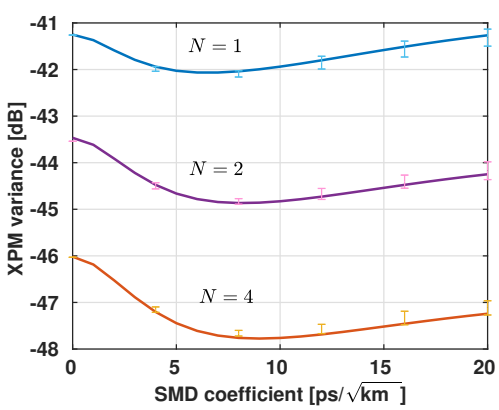

Fig. 5. XPM variance for a two-channel system propagating in a 5-span link at $\Delta f = 100$ GHz. Solid lines: ergodic GN model (17). Error bars: SSFM.

It is worth noting that while the evaluation of the closed-form approximation is almost instantaneous, the Monte Carlo integration of the GN model took a few seconds. In contrast, the SSFM simulations took, e.g., 2 days to estimate 100 seeds of the $N = 4$ curve on a server-grade computer with a graphical processing unit (GPU).

The accuracy of the ergodic GN model with respect to fiber dispersion is shown in Fig. 6 for a single span link. Here we also plotted the SPM contribution of the ergodic GN. Error bars still indicate the best/worst NLI variance from SSFM simulations. XPM is for $\Delta f = 100$ GHz. We note that dispersion is more effective in reducing XPM than SPM, as expected because of the walk-off. The slope of the variance with dispersion is SMD-dependent, approaching an almost inverse law at large dispersions.

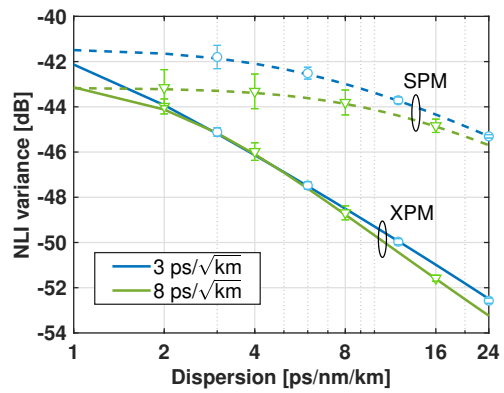

Fig. 6. SPM and XPM ($\Delta f = 100$ GHz) variance in a single span at variable fiber dispersion. Solid lines: ergodic GN model (17). Error bars: SSFM.

In Fig. 7 we investigate the dependence of the NLI variance on the main system parameters using the ergodic GN model. The figure shows the XPM variance at different dispersion values and fixed channel spacing $\Delta f = 100$ GHz (top), or at fixed dispersion 17 ps/nm/km and variable spacing (bottom). We observe that SMD is more effective in mitigating the XPM variance at low dispersive values and that an intriguing minimum of XPM variance is visible at SMD$\sim$ 8 ps/$\sqrt{\text{km}}$, whatever the channel spacing. The markers will be commented in Section VI.

In Fig. 8 we verified the proposed closed-form expressions with SMD, i.e., the XPM formula (39) used in (22), and the SPM extension (41). The link is the same of Fig. 7. The figure shows the Monte Carlo prediction of the ergodic GN model (17), which is assumed as the benchmark, and the fully closed-form expressions. We show just two instances of XPM as a reference, evaluated in a two-channel setup at spacing 500 GHz and 2 THz, respectively. The figure also shows the fully-loaded WDM case over a WDM bandwidth of 5 THz. The solid lines are the GN benchmark, the dashed lines the closed-forms. We observe a very good match. We note that SPM has a bias, which is present even in the absence of SMD, hence not related to our model. Nevertheless, the decrease with SMD is well captured. Overall, the WDM curve shows a maximum error of 0.4 dB.

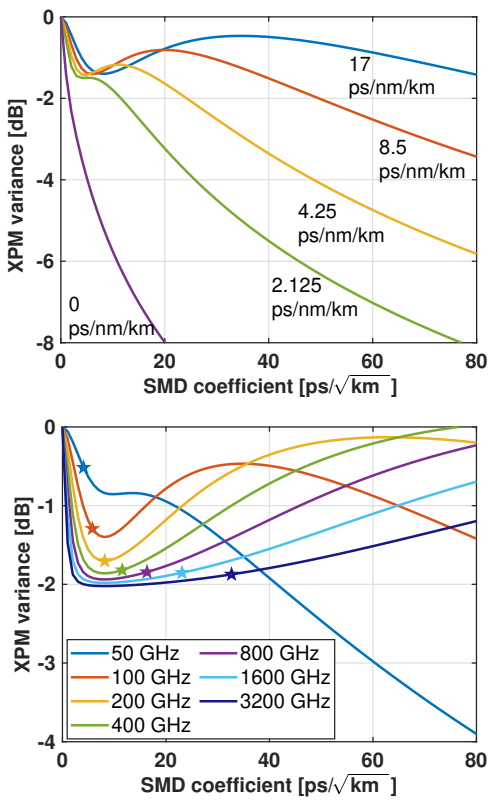

Fig. 7. XPM variance, normalized to the no SMD case, at variable dispersion and $\Delta f = 100$ GHz (top), or at variable Δf and fixed dispersion 17 ps/nm/km (bottom). 1 span, 2 modes. Markers: locus where $L_{\text{wo}} = L_{\text{SMD}}$, see section VI.

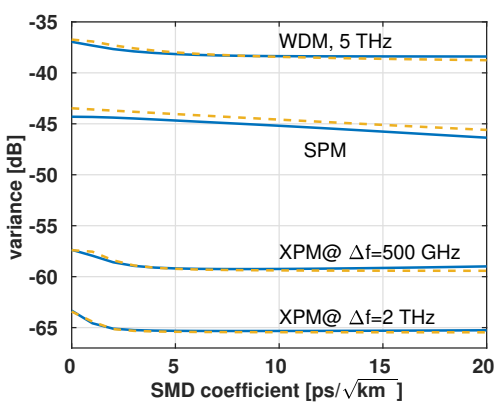

Fig. 8. Ergodic GN model (17), solid lines, vs closed-form formulas based on (39)–(41), dashed lines. XPM: example of two channels at spacing 500 GHz or 2 THz. The WDM curve refers to a fully-loaded WDM over 5 THz of bandwidth. Other parameters are identical to Fig. 7.

IV. Ergodic NLI variance

In this section we provide a detailed derivation of the ergodic GN model (17). In principle, the evaluation of $\mathcal{I}_{1,2}$ involves the computation of six frequency integrals. However, since the infinite summations over temporal indexes $(\mathbf{kmn}_1)$ can be transformed into summations involving Dirac's delta thanks to the Poisson formula, three integrals can be dropped, as discussed in [13, App. B].

We start by computing the FWM efficiency kernels $\eta^{(\ell)}$ of (19). They are special instances $\eta^{(\ell)} \equiv \eta^{(\ell)}_{i_3 i_3}$ of the more general kernel $\eta^{(\ell)}_{i_3 j_3}$ accounting for the cross-FWM between polarization i_3 and j_3. From (18) we have [2]:

$$|\eta^{(1)}_{i_3 j_3}(\omega,\omega_1,\omega_2)|^2 \triangleq \sum_{k_3,m_3,n_3} \eta_{\mathbf{kmni}}\eta^*_{\mathbf{kmnj}} = \int_0^{L_t}\int_0^{L_t} f(z)e^{j\Delta\beta z} f(s)e^{-j\Delta\beta s}\Lambda_{i_3 j_3}(z,s,\omega,\omega_1,\omega_2)\mathrm{d}z\mathrm{d}s \quad (23)$$

$$|\eta^{(2)}_{i_3 j_3}(\omega,\omega_1,\omega_2)|^2 \triangleq \sum_{k_3,m_3,n_3} \eta_{\mathbf{kmni}}\eta^*_{\mathbf{knmj}} = \int_0^{L_t}\int_0^{L_t} f(z)e^{j\Delta\beta z} f(s)e^{-j\Delta\beta s}\Gamma_{i_3 j_3}(z,s,\omega,\omega_1,\omega_2)\mathrm{d}z\mathrm{d}s\,. \quad (24)$$

In simplifying (23-24) we exploited the properties $\langle k_3|\mathbf{P}|m_3\rangle = P_{k_3 m_3}$ and $\sum_{i,j} A_{ij}B^*_{ij} = \mathrm{Tr}[\mathbf{AB}^\dagger]$. The kernels thus depend on the matrices $\mathbf{\Lambda}=(\Lambda_{ij})$, $\mathbf{\Gamma}=(\Gamma_{ij})$:

$$\mathbf{\Lambda} \triangleq \mathrm{Tr}\left[\mathbf{P}(z,\omega+\omega_2,\omega_1)\mathbf{P}^\dagger(s,\omega+\omega_2,\omega_1)\right] \times \mathbf{P}^\dagger(z,\omega,\omega_1)\mathbf{P}(s,\omega,\omega_1) \quad (25)$$

$$\mathbf{\Gamma} \triangleq \mathbf{P}^\dagger(z,\omega,\omega_2)\mathbf{P}^\dagger(s,\omega+\omega_2,\omega_1) \times \mathbf{P}(z,\omega+\omega_1,\omega_2)\mathbf{P}(s,\omega,\omega_1)\,. \quad (26)$$

Several observations can be drawn. First, in the absence of SMD it is $\mathbf{\Lambda} = 2N\mathbf{I}$ and $\mathbf{\Gamma} = \mathbf{I}$, with $\mathbf{I}$ the $2N \times 2N$ identity matrix. In such a case, the integrals in (23)–(24) can be closed. In particular, in the relevant case of a N_s-span homogeneous link with span length L and lumped amplification, the kernel takes the well-known forms [24]:

$$\frac{|\eta^{(1)}_{i_3 i_3}|^2}{2N} = |\eta^{(2)}_{i_3 i_3}|^2 = |\eta_0|^2\psi, \quad \text{(no MD)} \quad (27)$$

where η_0 is the classic one-span FWM kernel and ψ a phased array term:

$$|\eta_0|^2 \triangleq \left|\frac{1-e^{-\alpha L}e^{j\Delta\beta L}}{\alpha - j\Delta\beta}\right|^2, \qquad \psi \triangleq \left|\frac{1-e^{jN_s\Delta\beta L}}{1-e^{j\Delta\beta L}}\right|^2. \quad (28)$$

In such a case, (18) reduces, except for a constant factor, to the area of the GNRF ([9, eq. (1)]) weighted by the matched filter frequency response.

Second, the summation of frequency indexes can be reduced to a single summation if one is interested in the, usually dominant, XPM contribution. However, such XPM contributions, visible in Fig. 2, with SMD are no more degenerate along the main axes.

Third, the novelty brought by SMD is fully included into the matrices $\mathbf{\Lambda}$ and $\mathbf{\Gamma}$, which account for the spatial correlations between the coordinates z and s.

As anticipated in the previous sections, to simplify the problem we are interested in substituting the FWM kernels with their expectations with respect to the birefringence distribution, i.e., substituting $\mathbf{\Lambda}$ and $\mathbf{\Gamma}$ with $\mathbb{E}[\mathbf{\Lambda}]$ and $\mathbb{E}[\mathbf{\Gamma}]$, respectively. Luckily, such expectations are a linear combination of exponential functions that allow closing the integrals in z and s in (23)–(24), as we will show now.

[2]For the sake of simplicity, we assume $\Delta\beta$ independent of z.

A. Expected value of $\mathbf{\Lambda}$ and $\mathbf{\Gamma}$

The matrix $\mathbf{U}$ evolves according to the Ito's stochastic differential equation (SDE) (55), which can be further simplified by using the change of variable:

$$\mathbf{U}(z,\omega) \triangleq e^{-\frac{\omega^2\mu^2 z}{2N}}\mathbf{R}(z,\omega) \tag{29}$$

such that the Ito's SDE for matrix $\mathbf{R}$ is:

$$\mathrm{d}\mathbf{R} = -\frac{i\omega\mu}{2N}(\mathrm{d}\mathbf{W})\mathbf{R}\,. \tag{30}$$

We find convenient introducing:

$$\mathbf{Q}_{ij}(z) \triangleq \mathbf{R}^{\dagger}(z,\nu_i)\mathbf{R}(z,\nu_j) \tag{31}$$

with:

$$\nu_1 = \omega+\omega_1+\omega_2, \qquad \nu_2 = \omega+\omega_2$$
$$\nu_3 = \omega+\omega_1, \qquad \nu_4 = \omega\,.$$

We thus have:

$$\mathbf{\Lambda} \triangleq e^{-\frac{\mu^2(z+s)}{2N}\left(\nu_1^2+\nu_2^2+\nu_3^2+\nu_4^2\right)}\mathbf{\Lambda}'$$
$$\mathbf{\Gamma} \triangleq e^{-\frac{\mu^2(z+s)}{2N}\left(\nu_1^2+\nu_2^2+\nu_3^2+\nu_4^2\right)}\mathbf{\Gamma}' \tag{32}$$

with $\mathbf{\Lambda}'$ and $\mathbf{\Gamma}'$ having a similar structure as that of $\mathbf{\Lambda}$ and $\mathbf{\Gamma}$:

$$\mathbf{\Lambda}' = \mathrm{Tr}\left[\mathbf{Q}_{12}(z)\mathbf{Q}_{12}^{\dagger}(s)\right]\mathbf{Q}_{34}^{\dagger}(z)\mathbf{Q}_{34}(s)$$
$$\mathbf{\Gamma}' = \mathbf{Q}_{42}(z)\mathbf{Q}_{12}^{\dagger}(s)\mathbf{Q}_{31}^{\dagger}(z)\mathbf{Q}_{34}(s)\,. \tag{33}$$

Our target is the expectation of such matrices. Starting from $\mathbf{\Lambda}'$, after differentiation we solve the following SDE in the variable z (s is treated as a fixed coordinate):

$$\begin{aligned}\mathrm{d}\mathbf{\Lambda}' &= \mathrm{Tr}\left[(\mathrm{d}\mathbf{Q}_{12}(z))\mathbf{Q}_{12}^{\dagger}(s)\right]\mathbf{Q}_{34}^{\dagger}(z)\mathbf{Q}_{34}(s)\\ &+\mathrm{Tr}\left[\mathbf{Q}_{12}(z)\mathbf{Q}_{12}^{\dagger}(s)\right](\mathrm{d}\mathbf{Q}_{34}^{\dagger}(z))\mathbf{Q}_{34}(s)\\ &+\mathrm{Tr}\left[(\mathrm{d}\mathbf{Q}_{12}(z))\mathbf{Q}_{12}^{\dagger}(s)\right](\mathrm{d}\mathbf{Q}_{34}^{\dagger}(z))\mathbf{Q}_{34}(s)\,.\end{aligned} \tag{34}$$

Contrary to ordinary calculus, the last term on the right-hand side is non-negligible since we are dealing with an SDE [31]. Similarly, it is:

$$\begin{aligned}\mathrm{d}\mathbf{Q}_{ij}(z) &= (\mathrm{d}\mathbf{R}^{\dagger}(z,\nu_i))\mathbf{R}(z,\nu_j)+\mathbf{R}^{\dagger}(z,\nu_i)(\mathrm{d}\mathbf{R}(z,\nu_j))\\ &+(\mathrm{d}\mathbf{R}^{\dagger}(z,\nu_i))(\mathrm{d}\mathbf{R}(z,\nu_j))\,.\end{aligned} \tag{35}$$

We can now average. By exploiting (56) of Appendix A and the properties of $\mathrm{d}\mathbf{W}$ in (54), the only non-zero average comes from:

$$\begin{aligned}\mathbb{E}\left[\mathrm{d}\mathbf{Q}_{ij}\right] &= \mathbb{E}\left[(\mathrm{d}\mathbf{R}^{\dagger}(z,\nu_i))(\mathrm{d}\mathbf{R}(z,\nu_j))\right]\\ &= 2N\frac{\nu_i\nu_j\mu^2}{2N^2}\mathbb{E}\left[\mathbf{Q}_{ij}\right]\mathrm{d}z,\end{aligned}$$

which lets us simplify the first two addends in (34). The third addend in (34) can be found by using i) the identity (57), ii) the property $\mathrm{Tr}\left[\mathbf{AB}\right] = \mathrm{Tr}\left[\mathbf{BA}\right]$, iii) the second-order product rule (35), and iv) the Hermitian property $\mathrm{d}\mathbf{W} = \mathrm{d}\mathbf{W}^{\dagger}$. We get:

$$\begin{aligned}&\mathbb{E}\left[\mathrm{Tr}\left[(\mathrm{d}\mathbf{Q}_{12}(z))\mathbf{Q}_{12}^{\dagger}(s)\right](\mathrm{d}\mathbf{Q}_{34}^{\dagger}(z))\mathbf{Q}_{34}(s)\right]\\ &\qquad = \frac{(\nu_1-\nu_2)(\nu_3-\nu_4)\mu^2}{2N^2}\mathbb{E}\left[\mathbf{\Gamma}'\right]\mathrm{d}z\,.\end{aligned}$$

Along similar lines we can find a propagation equation for $\mathbf{\Gamma}'$. In particular, by exploiting (58) of Appendix A we get:

$$\begin{aligned}&\mathbb{E}\left[(\mathrm{d}\mathbf{Q}_{42}(z))\mathbf{Q}_{12}^{\dagger}(s)(\mathrm{d}\mathbf{Q}_{31}^{\dagger}(z))\mathbf{Q}_{34}(s)\right]\\ &\qquad = \frac{(\nu_4-\nu_2)(\nu_3-\nu_1)\mu^2}{2N^2}\mathbb{E}\left[\mathbf{\Lambda}'\right]\mathrm{d}z\,.\end{aligned}$$

After returning back to the original matrices $\mathbf{\Lambda}$ and $\mathbf{\Gamma}$ through (33), the evolution model with respect to the spatial lag $\zeta = |z-s|$ is the following:

$$\begin{aligned}\frac{\mathrm{d}\mathbb{E}\left[\mathbf{\Lambda}\right]}{\mathrm{d}\zeta} &= \frac{\mu^2}{N}\left(-\omega_1^2\mathbb{E}\left[\mathbf{\Lambda}\right]+\frac{\omega_1^2}{2N}\mathbb{E}\left[\mathbf{\Gamma}\right]\right)\\ \frac{\mathrm{d}\mathbb{E}\left[\mathbf{\Gamma}\right]}{\mathrm{d}\zeta} &= \frac{\mu^2}{N}\left(\frac{\omega_2^2}{2N}\mathbb{E}\left[\mathbf{\Lambda}\right]-\omega_2^2\mathbb{E}\left[\mathbf{\Gamma}\right]\right)\end{aligned} \tag{36}$$

with boundary condition:

$$\mathbb{E}\left[\mathbf{\Lambda}(\zeta=0)\right] = 2N\mathbf{I}$$
$$\mathbb{E}\left[\mathbf{\Gamma}(\zeta=0)\right] = \mathbf{I}\,.$$

The linear system of differential equations (36) can be solved in closed-form, see Appendix B. The main consequence is that the ergodic NLI variance with SMD takes a similar expression to the SMF case, only with different FWM kernels.

V. The simplified formula

Some simplifications are possible in the characterization of SPM and XPM, often the dominant sources of NLIs. The idea is to express the NLI variance as a function of the benchmark NLI variance induced by a given modal combination $\mathbf{kmni}_3$ in absence of SMD. In this way, it becomes easier to generalize well-known results for SMF, as well as understanding the main scaling properties with SDM and the impact of SMD on them. For a single SMF span of attenuation α, the benchmark is thus the following:

$$\begin{aligned}\sigma_1^2(\alpha) &\triangleq (\gamma\kappa)^2\iiint_{-\infty}^{\infty}|\eta_0(\omega,\omega_1,\omega_2)|^2\\ &\qquad\times|\mathcal{M}_{\mathbf{kmni}}(\omega,\omega_1,\omega_2)|^2\,\frac{\mathrm{d}\omega_1}{2\pi}\frac{\mathrm{d}\omega_2}{2\pi}\frac{\mathrm{d}\omega}{2\pi}\end{aligned} \tag{37}$$

which we will call $\sigma^2_{\mathrm{XPM},1}$ or $\sigma^2_{\mathrm{SPM},1}$ when focusing only on XPM or SPM index combinations, respectively. For instance, with $N=1$ and without SMD we have $\sigma^2_{\mathrm{XPM}} = 6\sigma^2_{\mathrm{XPM},1}$, where the relation between the factor 6 and the individual nonlinear contributions is discussed in [32, Tab. 1].

SMD locally acts as a modal differential delay, hence by a phase in the frequency domain linearly related to ω. As a result, different modes and polarizations, besides being delayed, accumulate different carrier phase shifts. We refer to the MD-induced carrier phase shift by *inter-channel* SMD [33]. This is illustrated in Fig. 9, which shows the differential propagation constant of a local eigenstate of the SMD matrix. Accounting only for inter-channel effects means switching off intra-channel SMD within a channel bandwidth, leaving a phase shift among channels as depicted by the inter-MD curve in the figure. The staircase inter-MD profile seems thus a reasonable approximation of the true profile. Indeed it helps to simplify the ergodic efficiency.

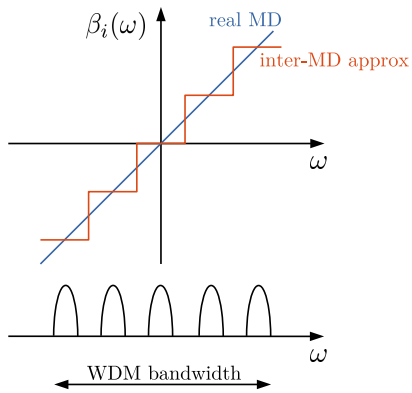

Fig. 9. Sketch of the frequency-dependent differential propagation constant induced by SMD only over the generic ith mode, and its inter-channel approximation, flat in each channel bandwidth.

Since XPM variance is additive in the number of channels, we can concentrate on a two-channel setup with channel spacing $\Delta\omega$. The inter-MD approximation suggests assuming $\omega \approx 0$ and either $(\omega_1, \omega_2) \approx (0, \Delta\omega)$ or $(\omega_1, \omega_2) \approx (\Delta\omega, 0)$ depending on whether we are analyzing an XPM hexagon centered on the ω_1 or ω_2 axis (see Fig. 2), which are non-degenerate in the presence of SMD. In the first case, the XPM variance of $\mathbf{k}$ on $\mathbf{i}$ corresponds to $\mathbf{kmni}_2 = \mathbf{kiki}_2$ in (17), while in the second case it is $\mathbf{kmni}_2 = \mathbf{kkii}_2$. In both cases, we can use the results of Appendix B with $q = p = \Delta\omega^2/2$ and thus $c_\ell = 1$ except for $\mathbf{kiki}_2$ and $\ell = 1$ where $c_1 = \frac{1}{2N^2} - 1$. As a consequence, only the term $|\mathcal{X}_{\mathbf{kiki}}|^2$ in (17) is impacted by SMD.

Let us focus on a fiber with span length $L \gg 1/\alpha$, an excellent approximation for terrestrial links, yet good for submarine links. With the above approximations the FWM kernels keep the same functional shape of the scalar FWM kernel η_0, but with different parameters. As a result, the inter-MD XPM ergodic variance takes the particularly elegant expression (22).

Some observations can be made about (22). First, the average effect of SMD is to add a contribution as if the fiber had a larger attenuation, thus justifying a beneficial effect of SMD in mitigating NLI through a smaller effective length. Second, we have the asymptotic behaviors:

$$\sigma^2_{\text{XPM}} = \begin{cases} 2(2N+1)\sigma^2_{\text{XPM},1}(\alpha) & \mu \to 0 \\ \frac{(2N+1)^2}{2N}\sigma^2_{\text{XPM},1}(\alpha) & \mu \to \infty, \end{cases} \tag{38}$$

thus matching the findings in [17]. The ratio between the no-MD and the infinite inter-channel-MD case is $\frac{4N}{2N+1}$, which is surely greater than 1. Third, $\sigma^2_{\text{XPM},1}$ per polarization can be approximated by closed-form formulas, for instance with [9, eq. (40)], here adapted to (37):

$$\sigma^2_{\text{XPM},1}(\alpha) \approx \frac{\kappa^2}{32}\frac{\gamma^2 P^3 L^2_{\text{eff}}}{\pi\beta_2 L_{\text{eff,a}} B^2_{\text{ch}}}\Big[\text{asinh}\left(\delta(k\Delta f + B_{\text{ch}}/2)\right) - \text{asinh}\left(\delta(k\Delta f - B_{\text{ch}}/2)\right)\Big] \tag{39}$$

where[3] $L_{\text{eff}} = (1 - e^{-\alpha L})/\alpha$, $L_{\text{eff,a}} = 1/\alpha$, $k\Delta f = \Delta\omega/(2\pi)$, $B_{\text{ch}} \leq \Delta f$ channel bandwidth, and $\delta = \pi^2\beta_2 L_{\text{eff,a}} B_{\text{ch}}$. The factor $\kappa^2/32$ is related to the factor $8/27$ in [9] by $\frac{\kappa^2}{32} = \frac{8}{27}\frac{\kappa^2}{(8/9)^2}\frac{1}{2\cdot 6}$, the last 2 coming from x+y power.

In the inter-MD approximation the variance of SPM is unaffected by SMD. Its value in a given polarization can thus be approximated by, e.g., [9, eq. (14)], again adapted to (37):

$$\sigma^2_{\text{SPM},1} \approx \frac{\kappa^2}{16}\frac{\gamma^2 P^3 L^2_{\text{eff}}}{\pi\beta_2 L_{\text{eff,a}} B^2_{\text{ch}}}\text{asinh}\left(\frac{\pi^2}{2}\beta_2 L_{\text{eff,a}} B^2_{\text{ch}}\right) \tag{40}$$

such that $\sigma^2_{\text{SPM}} = (2N+1)\sigma^2_{\text{SPM},1}$. An improved version of (40), which accounts heuristically for intra-channel SMD, is the expression presented in [34], reported here for convenience:

$$\sigma^2_{\substack{\text{SPM}\\ \text{improved}}} = \frac{1 - \exp\left(-\frac{T_I}{2\sqrt{|\beta_2|L}}\right)}{\frac{T_I}{2\sqrt{|\beta_2|L}}}\sigma^2_{\text{SPM}} \tag{41}$$

with T_I the standard deviation of the fiber impulse response (see Appendix C).

A. Multi-span

The previous discussion was for a single-span. For multi-span links it is customary to approximate the scaling of the NLI variance with the number of spans N_{s} by $N_{\text{s}}^{1+\epsilon}$, with $0 \leq \epsilon \leq 1$ a coherence factor strictly related to the phased array. The coherence factor is usually close to 0 in highly dispersive dispersion-uncompensated links with large WDM bandwidth [24], a situation that is satisfied by the majority of the optical links. With SMD, the coherence factor is surely smaller than without SMD, as it is evident from (69) in Appendix B. Hence, assuming a spatial disaggregation of the XPM contributions, an assumption very attractive to simplify the design of optical networks, is even better justified with SMD.

VI. Scaling Properties

Having tested the accuracy of the ergodic GN model, we here analyze its scaling properties to gain some physical insight. We analyze a two-channel field:

$$|A\rangle = |A_0\rangle + e^{i\Delta\omega t}|A_1\rangle, \tag{42}$$

in a lossless fiber in the absence of the Kerr effect, to concentrate on the dispersive effects. The propagation equations become:

$$\frac{\partial|A_0\rangle}{\partial z} = -\frac{\mu\mathbf{B}}{2N}\frac{\partial|A_0\rangle}{\partial t} + j\frac{\beta_2}{2}\frac{\partial^2|A_0\rangle}{\partial t^2} \tag{43}$$

$$\frac{\partial|A_1\rangle}{\partial z} = -\frac{\mu\mathbf{B}}{2N}\frac{\partial|A_1\rangle}{\partial t} - j\Delta\omega\frac{\mu\mathbf{B}}{2N}|A_1\rangle + j\frac{\beta_2}{2}\frac{\partial^2|A_1\rangle}{\partial t^2} - \beta_2\Delta\omega\frac{\partial|A_1\rangle}{\partial t} - j\frac{\beta_2\Delta\omega^2}{2}|A_1\rangle. \tag{44}$$

In the right-hand side of (44), from left to right, we see intra-channel SMD, inter-channel-SMD, intra-channel CD, inter-channel CD (aka walk-off), and a term inducing a phase shift.

The search for scaling laws makes it natural to normalize time and distance by $\tau = t/T$ and $\xi = z/L_r$, with L_r a reference distance that we will set later. In this framework,

[3]Note that in [9] α is the field-loss coefficient, which is half the power-loss coefficient α used here.

we can replace the coupling matrix $\mathbf{B}(L_r\xi)$, whose correlation function is proportional to $\delta(L_r(\xi-\xi^{'}))=\delta(\xi-\xi^{'})/L_r$, with the matrix $\mathbf{B}(\xi)/\sqrt{L_r}$, such that its correlation function is still proportional to $\delta(\xi-\xi^{'})/L_r$. We get for $|A_{0,1}(\xi,\tau)\rangle$:

$$\frac{\partial\,|A_0\rangle}{\partial\xi}=-\frac{\mu\sqrt{L_r}\mathbf{B}(\xi)}{T2N}\frac{\partial\,|A_0\rangle}{\partial\tau}+j\frac{\beta_2 L_r}{2T^2}\frac{\partial^2\,|A_0\rangle}{\partial\tau^2} \tag{45}$$

$$\frac{\partial\,|A_1\rangle}{\partial\xi}=-\frac{\mu\sqrt{L_r}\mathbf{B}(\xi)}{T2N}\frac{\partial\,|A_1\rangle}{\partial\tau}-j\frac{\mu\Delta\omega\sqrt{L_r}\mathbf{B}(\xi)}{2N}\,|A_1\rangle + j\frac{\beta_2 L_r}{2T^2}\frac{\partial^2\,|A_1\rangle}{\partial\tau^2}-\frac{\beta_2\Delta\omega L_r}{T}\frac{\partial\,|A_1\rangle}{\partial\tau}-j\frac{\beta_2\Delta\omega^2 L_r}{2}\,|A_1\rangle\,. \tag{46}$$

To make the intra-channel CD invariant to scalings, the natural choice for L_r is the dispersion length, $L_r=T^2/|\beta_2|$. This yields[4]:

$$\frac{\partial\,|A_0\rangle}{\partial\xi}=-\frac{\mu}{\sqrt{|\beta_2|}}\frac{\mathbf{B}(\xi)}{2N}\frac{\partial\,|A_0\rangle}{\partial\tau}+\frac{j}{2}\frac{\partial^2\,|A_0\rangle}{\partial\tau^2} \tag{47}$$

$$\frac{\partial\,|A_1\rangle}{\partial\xi}=-\frac{\mu}{\sqrt{|\beta_2|}}\frac{\mathbf{B}(\xi)}{2N}\frac{\partial\,|A_1\rangle}{\partial\tau}-\frac{j\mu}{\sqrt{|\beta_2|}}\Delta\omega T\frac{\mathbf{B}(\xi)}{2N}\,|A_1\rangle + \frac{j}{2}\frac{\partial^2\,|A_1\rangle}{\partial\tau^2}-\Delta\omega T\frac{\partial\,|A_1\rangle}{\partial\tau}-j\frac{\Delta\omega^2T^2}{2}\,|A_1\rangle\,. \tag{48}$$

The $\mu/\sqrt{|\beta_2|}$ factor underpins the heuristic expression of the SMD-dependent factor in the SPM NLI variance (41). On the other hand, XPM is expected to be mainly ruled by inter-channel effects. Hence, for XPM is more logical to make walk-off invariant to scaling, i.e., using $L_r=\frac{T}{|\beta_2|\Delta\omega}$. The propagation equations become:

$$\frac{\partial\,|A_0\rangle}{\partial\xi}=-\frac{\mu\mathbf{B}(\xi)}{2N\sqrt{|\beta_2|\Delta\omega T}}\frac{\partial\,|A_0\rangle}{\partial\tau}+j\frac{1}{2\Delta\omega T}\frac{\partial^2\,|A_0\rangle}{\partial\tau^2} \tag{49}$$

$$\frac{\partial\,|A_1\rangle}{\partial\xi}=-\frac{\mu\mathbf{B}(\xi)}{2N\sqrt{|\beta_2|\Delta\omega T}}\frac{\partial\,|A_1\rangle}{\partial\tau}-j\frac{\mu\sqrt{\frac{\Delta\omega T}{|\beta_2|}}\mathbf{B}(\xi)}{2N}\,|A_1\rangle + \frac{j}{2\Delta\omega T}\frac{\partial^2\,|A_1\rangle}{\partial\tau^2}-\frac{\partial\,|A_1\rangle}{\partial\tau}-j\frac{\Delta\omega T}{2}\,|A_1\rangle\,. \tag{50}$$

The SMD-based coefficients appearing in (49,50) can be more conveniently expressed in terms of two relevant length-scales. One is the familiar walk-off length L_{wo} [27], the other is the mode dispersion length $L_{\text{SMD}}(B)$ [3], [35]. L_{wo} is defined as the length over which an interfering pulse experiences a walk-off of one symbol time [27], and coincides with L_r; L_{SMD} is the length over which the 3dB bandwidth of the autocorrelation function of the SMD vector equals a reference bandwidth, hence providing a measure of the SMD decorrelation. They are:

$$L_{\text{wo}}=\frac{1}{|\beta_2|R2\pi\Delta f},\qquad L_{\text{SMD}}(B)=0.2^2\frac{(4N^2-1)}{(N\cdot\eta_{\text{SMD}}\cdot B)^2}\,. \tag{51}$$

We stressed the dependence of L_{SMD} on a reference bandwidth B to highlight that it can be referred to the signal bandwidth, for which $B\equiv R$, or the carrier spacing, for which $B\equiv\Delta f$. A closer look to (50) reveals that the first term on the right hand side is proportional to $L_{\text{wo}}/L_{\text{SMD}}(R)$, while the second term to $L_{\text{wo}}/L_{\text{SMD}}(\Delta f)$. Figure 3, however, suggests that inter-channel SMD is dominant in a significant range of SMD values. This implies $L_{\text{wo}}/L_{\text{SMD}}(R)\ll 1$, such that SMD is set through $L_{\text{wo}}/L_{\text{SMD}}(\Delta f)$. To test this scaling, we ran 10000 different random setups of a two-channel transmission in a single span, with attenuation 0.2 dB/km, dispersion in the range 2 to 30 ps/nm/km, R from 5 to 100 Gbd, Δf from 50 to 5000 GHz, and SMD from 0 to 30 ps/$\sqrt{\text{km}}$. We limited the random generation to values $L_{\text{wo}}/L_{\text{SMD}}(\Delta f)<50$. Figure 10 shows the XPM variance, normalized to the no-MD case, given by the ergodic GN model vs the ratio $L_{\text{wo}}/L_{\text{SMD}}(\Delta f)$. The red points are a subset of results for which $L_{\text{wo}}\geq L_{\text{SMD}}(R)$. Figure 10 thus generalizes Fig. 7 with $N=2$ to a wider scenario.

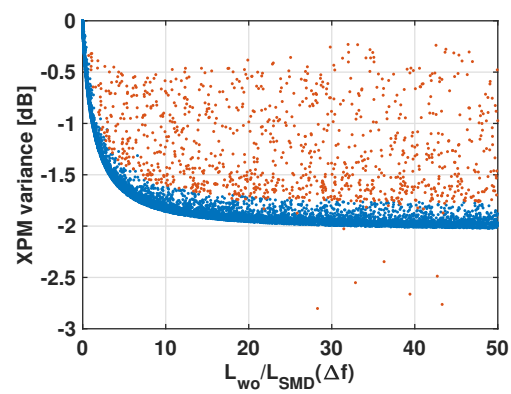

Fig. 10. XPM variance, normalized to the no SMD case, for 10000 different random setups, see text. Blue points: samples for which $L_{\text{wo}}<L_{\text{SMD}}(R)$. The narrow cloud of the blue points suggests a scaling with $L_{\text{wo}}/L_{\text{SMD}}(\Delta f)$, provided that $L_{\text{wo}}<L_{\text{SMD}}(R)$.

From the figure, we infer that the narrow cloud of blue points confirms a scaling with the ratio $L_{\text{wo}}/L_{\text{SMD}}(\Delta f)$. As a sanity check, this scaling is also visible in (22), normalized to the corresponding no-MD variance, after an asymptotic expansion in $\Delta\omega$. The scaling is broken when L_{wo} is greater than $L_{\text{SMD}}(R)$, an indication that intra-channel SMD effects are of concern in the red points of Fig. 10. However, for practical SMD values, it is usually $L_{\text{wo}}<L_{\text{SMD}}(R)$, see the stars in Fig. 7 indicating the SMD at $L_{\text{wo}}=L_{\text{SMD}}(R)$.

We also investigated the reasons for the discrepancy between the full GN model and the inter-channel one, manifesting in a MD-induced inflation of the XPM variance, e.g., see the peak in Fig. 3 at SMD$\sim$30 ps/$\sqrt{\text{km}}$. To this aim, we split the nonlinear effect into an average XPM, responsible for common phase rotation among the $2N$ polarizations, and a cross-polarization modulation (XPolM) effect accounting for rotations of the generalized Stokes vectors [3]. Hence, the Manakov equation of channel $|A_0\rangle$ is impaired by the following cross-channel Kerr effect from $|A_1\rangle$ [5, eq. (70)]:

$$\left(\underbrace{\langle A_1|A_1\rangle\,\mathbf{I}+\mathbb{E}[|A_1\rangle\langle A_1|]}_{\text{average XPM}}+\underbrace{|A_1\rangle\langle A_1|-\mathbb{E}[|A_1\rangle\langle A_1|]}_{\text{XPolM}}\right)|A_0\rangle\,. \tag{52}$$

[4]We assume $\beta_2>0$ without loosing generality.

We associated $\mathbb{E}[|A_1\rangle\langle A_1|]$ to the average XPM because it is $\mathbb{E}[|A_1\rangle\langle A_1|] = \frac{1}{2N}\langle A_1|A_1\rangle\mathbf{I}$ [5]. To shed light on the problem, we ran SSFM simulations with two channels (Δf=500 GHz) by selectively switching on/off the average XPM or XPolM, and by varying the number of waveplates in a SMF scenario for the sake of simplicity. We focused on a single span with dispersion 1.7 ps/nm/km and averaged the SNRs over 10 random seeds. Extra simulations showed that dispersion has a minor role in the following discussion, while 10 seeds were enough. The two contributions to the XPM variance are plotted in Fig. 11 vs the rms value of the differential group delay (DGD) per waveplate, normalized to the scalar walk-off time $\tau_{\text{walk-off}}$ within the waveplate. Note that the walk-off in this setup makes channel A_0 faster than channel A_1 without SMD.

We observe that the average XPM dominates XPolM in this setup. It is worth noting that both effects show inflation with a small number of waveplates. This is expected since in a waveplate the PMD accelerates one polarization and delays the other. Hence, within a waveplate, at the PMD yielding DGD=$\tau_{\text{walk-off}}$ the fast polarization of channel A_1 travels with the same speed of the slow polarization of channel A_0, thus canceling out the effect of the scalar walk-off induced by CD. As a result, the cross-channel interaction is maximized by a coherent accumulation, which is visible in the peak of the one waveplate curves in Fig. 11. By increasing the number of waveplates the peak position moves toward smaller DGD values. For instance, with two waveplates the peak is at a DGD $\sqrt{2}$ smaller. After a very large number of waveplates, approximately when the waveplate length approaches L_{wo}, the average XPM curve saturates still preserving the SMD-induced inflation, while XPolM vanishes. For very large values of DGD$/\tau_{\text{walk-off}}$ both the average XPM and XPolM are mitigated by a stronger decorrelation of the frequencies building up the channels.

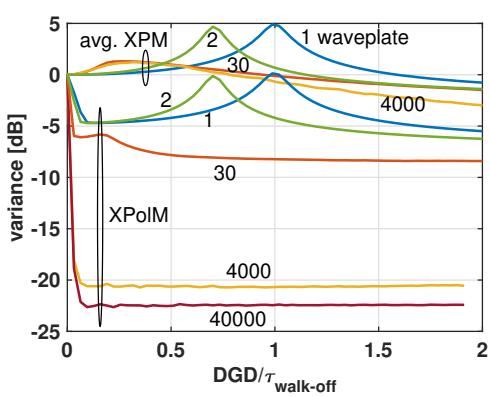

Fig. 11. Variance of average XPM and XPolM, normalized to the value without PMD, in SMF at variable number of birefringence waveplates vs the rms DGD per waveplate, normalized to the waveplate walk-off time.

Such observations are summarized in Table I.

VII. Conclusions

Computationally efficient models to predict the SNR of fiber-optic links are undoubtedly key tools to design and analyze fiber-optic networks. In this work, we extended the GN model to the study of the NLI among strongly coupled modes in SDM links with arbitrary values of SMD. We showed that SMD has a strong impact on the NLI by decreasing the variance of XPM. The first major result of our work is the ergodic GN model, i.e., the ensemble expectation of the NLI variance in the presence of SMD, that we evaluated by applying Ito's calculus to the perturbative solution of the Manakov equation with SMD. The derived NLI variance expression, given in (17) of this work, includes integrals of the kind derived in the single-mode GN model, which can be efficiently evaluated by means of Monte-Carlo integration. The effect of SMD results into a modified FWM kernel, whose derivation relies on some of the tools of Ito stochastic calculus. The second major result is the approximated closed-form formula (22) of the XPM variance, which we showed to be accurate for SMD values up to ~ 5 ps/$\sqrt{\text{km}}$. We also showed that in the relevant range of SMD values the randomness of the XPM variance is negligible, thus justifying the use of the ergodic GN model in analyzing strongly coupled SDM fiber-optic links. Finally, we discussed the main properties of the proposed ergodic GN model, showing its main scaling properties.

The investigation of general SDM networks, where not all of the modes are strongly coupled, requires addressing the nonlinear interplay between multiple groups of strongly coupled modes, each characterized by a specific amount of SMD. The starting point of this investigation is the coupled Manakov equations [5], however its execution goes beyond the scope of this work and is left for future studies.

Appendix A
Statistical Propagation Model

The transfer matrix $\mathbf{U}$ evolves along distance z as [36], [37]:

$$\mathrm{d}\mathbf{U} = -\frac{i\omega\mu}{2N}\mathbf{B}\mathbf{U}\mathrm{d}z \tag{53}$$

where μ expresses the strength of SMD, and is related to the SMD coefficient in Appendix C. The matrix $\mathbf{B}$ is a Hermitian matrix, such that the generic elements of the matrix $\mathrm{d}\mathbf{W} \equiv \mathrm{d}\mathbf{W}(z) = \mathbf{B}(z)\mathrm{d}z$ describe the differential of a complex Wiener process, i.e.[5]:

$$\begin{aligned} \mathrm{d}W_{kn} &= \mathrm{d}W_{nk}^* \\ \mathbb{E}\left[\mathrm{d}W_{kn}\right] &= 0 \\ \mathbb{E}\left[\mathrm{d}W_{ij}\mathrm{d}W_{kn}^*\right] &= 2\delta_{ik}\delta_{jn}\cdot\mathrm{d}z \end{aligned} \tag{54}$$

The propagation model (53) of the linear modal effects is a SDE in Stratonovich form since it describes a physical system [31], [37], [38]. However, for our purposes, the SDE is more useful once expressed in Ito form, because it is only in such a form that the increment $\mathrm{d}\mathbf{W}$ points "toward the future" [31, p. 92], thus allowing simplifications like $\mathbb{E}\left[\mathbf{U}(z)\mathrm{d}\mathbf{W}(z)\right] = \mathbb{E}\left[\mathbf{U}(z)\right]\mathbb{E}\left[\mathrm{d}\mathbf{W}(z)\right] = 0$. By applying the Stratonovich to Ito

[5]Note that matrix $\mathbf{W}$ includes also the common mode contribution.

TABLE I
SUMMARY OF THE SMD-XPM INTERACTION. $\overline{\text{XPM}}$ INDICATES THE AVERAGE XPM.

case	example	dominant SMD	comments	Fig. 10
$L_{\text{wo}} \ll L_{\text{SMD}}(R)$	small SMD	inter-SMD	$\sigma^2_{\text{XPM}}(\text{SMD}) < \sigma^2_{\text{XPM}}(0)$, scaling with $L_{\text{wo}}/L_{\text{SMD}}(\Delta f)$	blue points
$L_{\text{SMD}}(\Delta f) \ll L_{\text{wo}} \ll L_{\text{SMD}}(R)$	big Δf or small R	inter-SMD	channel decorrelation, max XPM mitigation	blue points
$L_{\text{wo}} > L_{\text{SMD}}(R)$	large SMD	intra+inter-SMD	vanishing of XPolM, $\overline{\text{XPM}}$ inflation due to SMD-reduced walk-off	red points
$L_{\text{wo}} \gg L_{\text{SMD}}(R)$	very large SMD	intra-SMD	vanishing of XPolM and $\overline{\text{XPM}}$	red points

conversion to each element of the matrix, (53) transforms into the following propagation model in Ito's form:

$$\mathrm{d}\mathbf{U} = -\frac{\omega^2\mu^2}{2N}\mathbf{U}\mathrm{d}z - \frac{i\omega\mu}{2N}(\mathrm{d}\mathbf{W})\mathbf{U} \tag{55}$$

where the first term on the right-hand side is the Ito's correction term.

We observe three important relationships involving matrix $\mathrm{d}\mathbf{W}$ and two generic $2N \times 2N$ matrices $\mathbf{A}, \mathbf{B}$:

$$\mathbb{E}\left[\mathbf{A}(\mathrm{d}\mathbf{W})^\dagger(\mathrm{d}\mathbf{W})\mathbf{B}\right] = 2N\mathbf{A}\mathbf{B}\mathrm{d}z \tag{56}$$

$$\mathbb{E}\left[\mathrm{Tr}\left[(\mathrm{d}\mathbf{W})\mathbf{A}\right]\mathbf{B}\mathrm{d}\mathbf{W}^\dagger\right] = 2\mathbf{B}\mathbf{A}\mathrm{d}z \tag{57}$$

$$\mathbb{E}\left[\mathbf{A}(\mathrm{d}\mathbf{W})^\dagger\mathbf{B}\mathrm{d}\mathbf{W}\right] = 2\,\mathrm{Tr}[\mathbf{B}]\mathbf{A}\mathrm{d}z\,. \tag{58}$$

The proof is straightforward by exploiting (54).

APPENDIX B
ERGODIC FWM KERNEL WITH SMD

The solution of (36) is:

$$\begin{aligned}\mathbb{E}\left[\mathbf{\Lambda}\right] &= (2NM_{11} + M_{12})\,\mathbf{I}\\ \mathbb{E}\left[\mathbf{\Gamma}\right] &= (2NM_{21} + M_{22})\,\mathbf{I}\end{aligned} \tag{59}$$

with the matrix $\mathbf{M} = (M_{ij})$:

$$\mathbf{M} \triangleq e^{\frac{\mu^2}{N}\mathbf{A}\zeta} \tag{60}$$

and:

$$\mathbf{A} = \begin{bmatrix} -\omega_1^2 & \frac{\omega_1^2}{2N} \\ \frac{\omega_2^2}{2N} & -\omega_2^2 \end{bmatrix}. \tag{61}$$

Remember that μ is the SMD strength while $\zeta = |z-s|$, with (z, s) integration variables of (23)−(24).

The matrix exponential in (60) can be simplified by the Cayley-Hamilton theorem. By introducing the two parameters:

$$\begin{aligned} p &\triangleq \frac{\omega_1^2 + \omega_2^2}{2}\\ q &\triangleq \sqrt{p^2 - \omega_1^2\omega_2^2\left(1 - \frac{1}{4N^2}\right)}\end{aligned}$$

we have:

$$\begin{aligned}\mathbf{M} = e^{-p\frac{\mu^2}{N}\zeta}\Bigg[&\left(\cosh\left(q\frac{\mu^2}{N}\zeta\right) + p\frac{\sinh\left(q\frac{\mu^2}{N}\zeta\right)}{q}\right)\mathbf{I}\\ &+ \frac{\sinh\left(q\frac{\mu^2}{N}\zeta\right)}{q}\mathbf{A}\Bigg].\end{aligned}$$

Since matrix $\mathbf{M}$ is a linear combination of exponential functions, the FWM kernel, averaged over the birefringence realizations, can be evaluated in closed-form. Two notable examples are the following.

A. Single span

After substituting (59) in (23)–(24), we found the following expression:

$$\begin{aligned}\mathbb{E}\left[|\eta^{(\ell)}(\omega,\omega_1,\omega_2)|^2\right] &= m_\ell \int_0^L\int_0^L e^{-\alpha(z+s)}e^{j\Delta\beta(z-s)}\\ \times&\left(e^{(q-p)\frac{\mu^2}{N}|z-s|}\left(1+c_\ell\right) + e^{-(q+p)\frac{\mu^2}{N}|z-s|}\left(1-c_\ell\right)\right)\mathrm{d}z\mathrm{d}s\end{aligned} \tag{62}$$

where:

$$c_\ell \triangleq \begin{cases}\frac{p}{q} - \frac{\omega_1^2}{q}\left(1 - \frac{1}{4N^2}\right) & \ell = 1\\ \frac{p}{q} & \ell = 2\end{cases}$$

and

$$m_\ell \triangleq \begin{cases} N & \ell = 1\\ \frac{1}{2} & \ell = 2\,.\end{cases} \tag{63}$$

We observe that the term in the big parentheses of (62) is equal to 2 in absence of SMD. We find convenient compacting the notation with the following real parameter:

$$\begin{aligned}E_\rho &\triangleq \int_0^L\int_0^L e^{\rho|z-s|}e^{-\alpha(z+s)}e^{i\Delta\beta(z-s)}\mathrm{d}s\mathrm{d}z\\ &= \frac{1}{2}\int_{-L}^L e^{\rho|x|}e^{i\Delta\beta x}\left(\int_{|x|}^{2L-|x|} e^{-\alpha y}\mathrm{d}y\right)\mathrm{d}x\,.\end{aligned} \tag{64}$$

The ergodic kernel is therefore the main result (20) with $\rho_1 = (q-p)\frac{\mu^2}{N}$ and $\rho_2 = -(q+p)\frac{\mu^2}{N}$. Note that $\rho_{1,2} \leq 0$. The result of the double integral in (64) is:

$$E_\rho = \frac{1}{\alpha}\mathrm{Re}\left[H_\rho\right] \tag{65}$$

with:

$$H_\rho \triangleq \frac{1 - e^{-(\alpha-\rho-i\Delta\beta)L}}{\alpha-\rho-i\Delta\beta} + \frac{e^{-2\alpha L} - e^{-(\alpha-\rho-i\Delta\beta)L}}{\alpha+\rho+i\Delta\beta}\,. \tag{66}$$

In the absence of SMD, $\rho_1 = \rho_2 = 0$ and $E_0 \equiv |\eta_0|^2$, i.e., the single-span kernel of (28).

B. Multi-span

For an homogeneous, dispersion-uncompensated, link of N_s span with lumped amplification, the ergodic FWM kernel can still be evaluated through (20) by substituting E_ρ with the generalized $E_{\rho,N}$:

$$E_{\rho,N} = \frac{1}{2}\sum_{m,n=0}^{N_\mathrm{s}-1}\int_{-L}^{L} e^{\rho|x+(m-n)L|} e^{i\Delta\beta(x+(m-n)L)} \times \int_{|x|}^{2L-|x|} e^{-\alpha y}\mathrm{d}y\mathrm{d}x\,. \tag{67}$$

The case $m=n$ has been analyzed in the single-span case. For $m>n$ we have $x+(m-n)L\geq 0$ in the domain of integration, while for $m<n$ it is $x+(m-n)L\leq 0$. Moreover, the double summation in (67) can be substituted by a single summation in the index $\ell=m-n$ by exploiting the following identity for a function $h(\ell)$ of Hermitian symmetry, i.e., $h(\ell)=h^*(-\ell)$:

$$\sum_{m=0}^{N_\mathrm{s}-1}\sum_{n=0}^{N_\mathrm{s}-1} h(m-n) = N_\mathrm{s}\left(h(0)+2\sum_{\ell=1}^{N_\mathrm{s}-1}\left(1-\frac{\ell}{N_\mathrm{s}}\right)\mathrm{Re}\left[h(\ell)\right]\right). \tag{68}$$

Hence, we have:

$$E_{\rho,N} = N_\mathrm{s}E_\rho + \frac{1}{\alpha}\sum_{\ell=1}^{N_\mathrm{s}-1}\left(1-\frac{\ell}{N_\mathrm{s}}\right)\mathrm{Re}\left[e^{(\rho+i\Delta\beta)\ell L}\left(H_\rho+H^*_{-\rho}\right)\right] \tag{69}$$

The sum in (69) can be closed by observing that:

$$\sum_{\ell=1}^{N_\mathrm{s}-1}\left(1-\frac{\ell}{N_\mathrm{s}}\right)e^{\delta\ell} = \frac{e^\delta\left(e^{N_\mathrm{s}\delta}-1+N_\mathrm{s}(1-e^\delta)\right)}{N_\mathrm{s}(e^\delta-1)^2}\,. \tag{70}$$

Appendix C
Relation between modal-dispersion parameters

To quantify SMD it is customary in the literature to focus on the mode-averaged intensity impulse response $I(t)$, i.e., the output power when a single-mode carrying white noise is excited on input. Such an impulse response is Gaussian-shaped, $I(t)=I_0\exp(-\frac{t^2}{2T_I^2})$ [35]. The variance T_I^2 is related to mean-square value of the SMD vector τ and the SMD coefficient usually adopted in the literature [39], here called η_SMD, by :

$$T_I^2 = \frac{\mathbb{E}[\tau^2]}{4N^2} = \frac{1}{2N}\sum_{n=1}^{2N}\mathbb{E}[t_n^2] \triangleq \frac{{\eta_\mathrm{SMD}}^2\cdot z}{4} \tag{71}$$

where t_n are the individual delays experienced by the principal states with respect to the mode-average delay, within the first-order picture of SMD [2]. We observe a linear growth with z of T_I^2 because of the random mode coupling. Since from [36] we have[6]:

$$\mathbb{E}\left[\tau^2\right] = \frac{4N^2-1}{N}\mu^2 z \tag{72}$$

we get the key relation:

$$\mu = \sqrt{\frac{N^3}{4N^2-1}}\eta_\mathrm{SMD}\,. \tag{73}$$

[6]The factor N in the denominator of (71) is due to the normalization factor $\sqrt{N}$ of the extended Pauli matrices Λ_i in [3].

In particular, since in SMF it is $\mu=\sqrt{\frac{\pi}{8}}\kappa_\mathrm{PMD}$ [35], [37], with κ_PMD the PMD coefficient of the SMF, we have $\eta_\mathrm{SMD}|_{N=1} = \sqrt{\frac{3\pi}{8}}\kappa_\mathrm{PMD}$.

In the SSFM simulations, we applied deterministic t_n in each waveplate [2], identical for all modes, with a differential group delay between the two polarizations of a given mode equal to T_I, evaluated by (71) with z equal to the waveplate length.

References

[1] P. J. Winzer, D. Nelson, and A. R. Chraplyvy, "Fiber-optic transmission and networking: the previous 20 and the next 20 years," *Opt. Express*, vol. 26, no. 18, pp. 24190–24239, 2018.

[2] K.-P. Ho and J. M. Kahn, "Mode coupling and its impact on spatially multiplexed systems," in *Optical Fiber Telecommunications VI* B, I. P. Kaminow, T. Li, and A. E. Willner, Eds., Amsterdam, The Netherlands: Elsevier, 2013, ch. 11.

[3] C. Antonelli, A. Mecozzi, M. Shtaif, and P. J. Winzer, "Stokes-space analysis of modal dispersion in fibers with multiple mode transmission," *Opt. Express*, vol. 20, no. 11, pp. 11718–11733, 2012.

[4] R. Ryf, J. C. Alvarado-Zacarias, S. Wittek, N. K. Fontaine, and H. Chen, "Coupled-Core Transmission over 7-Core Fiber," in *Proc. Opt. Fiber Commun. Conf.*, San Diego, CA, USA, 2019, Paper Th4B.3.

[5] 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, 2016.

[6] P. M. Krummrich, M. Brehler, G. Rademacher, and K. Petermann, "Nonlinear Impairment Scaling in Multi-Mode Fibers for Mode-Division Multiplexing," *J. Lightw. Technol.*, vol. 39, no. 4, pp. 927–932, 2021.

[7] P. Serena , C. Lasagni, S. Musetti, and A. Bononi, "On Numerical Simulations of Ultra-Wideband Long-Haul Optical Communication Systems," *J. Lightw. Technol.*, vol. 38, no. 5, pp. 1019–1031, 2020.

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

[9] P. Poggiolini, "The GN Model of Non-Linear Propagation in Uncompensated Coherent Optical Systems," *J. Lightw. Technol.*, vol. 30, no. 24, pp. 3857–3879, 2012.

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

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

[12] G. Gao, X. Chen, and W. Shieh, "Influence of PMD on fiber nonlinearity compensation using digital back propagation," *Opt. Express*, vol. 20, no. 13, pp. 14406–14418, 2012.

[13] P. Serena, C. Lasagni, and A. Bononi, "The Enhanced Gaussian Noise Model Extended to Polarization-Dependent Loss," *J. Lightw. Technol.*, vol. 38, n. 20, pp. 5685–5694, Oct. 2020.

[14] D. Semrau, R. 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, 2018.

[15] G. Rademacher and K. Petermann, "Nonlinear Gaussian Noise Model for Multimode Fibers with Space-Division Multiplexing," *J. Lightw. Technol*, vol. 34, no. 9, pp. 2280–2287, 2016.

[16] A. D. Ellis, N. Mac Suibhne, F. C. Garcia Gunning, and S. Sygletos, "Expressions for the nonlinear transmission performance of multi-mode optical fiber," *Opt. Express*, vol. 21, no. 19, p. 22834–22846, 2013.

[17] 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, p. 13055–13078, 2017.

[18] T. Hayashi, Y. Tamura, T. Hasegawa, and T. Taru, "Record-Low Spatial Mode Dispersion and Ultra-Low Loss Coupled Multi-Core Fiber for Ultra-Long-Haul Transmission," *J. Lightw. Technol.*, vol. 35, no. 3, pp. 450–457, 2017.

[19] Y. Yamada et al., "Design of High-Density Cable Parameters for Controlling Spatial-Mode Dispersion of Randomly Coupled Multi-Core Fibers," *J. Lightw. Technol.*, vol. 39, no. 4, pp. 1179–1185, 2021.

[20] P. Serena, C. Lasagni, A. Bononi, C. Antonelli, and A. Mecozzi, "A Model of the Nonlinear Interference in Space-Division Multiplexed Systems with Arbitrary Modal Dispersion," in *Proc. Eur. Conf. Opt. Commun*, Bordeaux, France, 2021, Paper Tu2D.3.

[21] A. Goldsmith, *Wireless communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
[22] B. J. Puttnam, G. Rademacher, and R. S. Luís, “Space-division multiplexing for optical fiber communications,” *Optica*, vol. 8, no. 9, pp. 1186–1203, 2021.
[23] E. Grellier and A. Bononi, "Quality parameter for coherent transmissions with Gaussian-distributed nonlinear noise,” *Opt. Express*, vol. 19, no. 13, pp. 12781–12788, 2011.
[24] P. Poggiolini, G. Bosco, A. Carena, V. Curri, Y. Jiang, and F. Forghieri, “The GN-Model of Fiber Non-Linear Propagation and its Applications,” *J. Lightw. Technol*., vol. 32, no. 4, pp. 694–721, 2014.
[25] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, “Pulse collision picture of inter-channel nonlinear interference noise in fiber-optic communications,” *J. Lightw. Technol.*, vol. 34, no. 2, pp. 593–607, 2016.
[26] A. Vannucci, P. Serena, and A. Bononi, “The RP Method: A New Tool for the Iterative Solution of the Nonlinear Schrodinger Equation,” *J. Lightw. Technol.*, vol. 20, no. 7, pp. 1102–1112, Jul. 2002.
[27] G. P. Agrawal, *Nonlinear Fiber Optics*, 3rd ed. San Diego, CA: Academic Press, 2001.
[28] R. J. Essiambre, M. Mestre, R. Ryf, and A. Gnauck, “Experimental investigation of inter-modal four-wave mixing in multimode fibers,” *IEEE Photon. Technol. Lett.*, vol. 25, no. 6, pp. 539–542, 2013.
[29] 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, 2018
[30] T. Hayashi *et al.*, “Field-deployed multi-core fiber testbed,” in *Proc. 24th Opto Electronics Commun. Conf. Int. Conf. Photon. Switching Comput.*, Jul. 2019, Paper PDP3.
[31] C. W. Gardiner, *Handbook of Stochastic Methods,* 2nd ed. New York: Springer Verlag, 1985.
[32] P. Serena and A. Bononi, “A time-domain extended Gaussian noise model,” *J. Lightw. Technol.*, vol. 33, no. 7, pp. 1459–1472, 2015.
[33] 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, 2011.
[34] C. Antonelli, O. Golani, M. Shtaif, and A. Mecozzi, “Propagation effects in few-mode fibers,” in *Proc. Eur. Conf. Opt. Commun*, Goteborg, Sweden, 2017, Paper W.1.B.1.
[35] R. Ryf and C. Antonelli, “Space-division Multiplexing,” in Springer *Handbook of Optical Networks*, B. Mukherjee, I. Tomkos, M. Tornatore, P. Winzer, and Y. Zhao, Eds. Cham: Springer International Publishing, 2020, pp. 353–394.
[36] C. Antonelli, A. Mecozzi, M. Shtaif , N.- K. Fontaine, H. Chen , and R. Ryf, ”Stokes-Space Analysis of Modal Dispersion of SDM Fibers With Mode-Dependent Loss: Theory and Experiments,” *J. Lightw. Technol.*, vol. 38, no. 7, pp. 1668–1677, 2020.
[37] A. Vannucci and A. Bononi, “Statistical characterization of the Jones matrix of long fibers affected by polarization mode dispersion (PMD),” *J. Lightw. Technol.*, vol. 20, no. 5, pp. 811–821, 2002.
[38] P. K. A. Wai and C. R. Menyuk, “Polarization mode dispersion, decorrelation, and diffusion in optical fibers with randomly varying birefringence,” *J. Lightw. Technol*., vol. 14, no. 2, pp. 148–157, 1996.
[39] T. Hayashi, *Multi-core Fibers for Space Division Multiplexing*, in: Peng GD. (eds) Handbook of Optical Fibers. Springer, Singapore. https://doi.org/10.1007/978-981-10-1477-2_66-3.