Nonlinearity Mitigation in WDM Systems: Models, Strategies, and Achievable Rates

(article starts on next page)

Nonlinearity Mitigation in WDM Systems: Models, Strategies, and Achievable Rates

Marco Secondini *Senior Member, IEEE*, Erik Agrell *Fellow, IEEE*, Enrico Forestieri *Senior Member, IEEE*, Domenico Marsella, Menelaos Ralli Camara

Abstract—After reviewing models and mitigation strategies for interchannel nonlinear interference (NLI), we study its characteristics and coherence properties. Based on this study, we devise an NLI mitigation strategy which exploits the synergic effect of phase and polarization noise compensation (PPN) and subcarrier multiplexing with symbol-rate optimization. This synergy persists even for high-order modulation alphabets and Gaussian symbols. A particle method for the computation of the resulting achievable information rate and spectral efficiency (SE) is presented and employed to lower-bound the channel capacity. The dependence of the SE on the link length, amplifier spacing, and presence or absence of in-line dispersion compensation is studied. Single-polarization and dual-polarization scenarios with either independent or joint processing of the two polarizations are considered. Numerical results show that, in links with ideal distributed amplification, an SE gain of about 1 bit/s/Hz/polarization can be obtained (or, in alternative, the system reach can be doubled at a given SE) with respect to single-carrier systems without PPN mitigation. The gain is lower with lumped amplification, increases with the number of spans, decreases with the span length, and is further reduced by in-line dispersion compensation. For instance, considering a dispersion-unmanaged link with lumped amplification and an amplifier spacing of 60 km, the SE after 80 spans can be be increased from 4.5 to 4.8 bit/s/Hz/polarization, or the reach raised up to 100 spans (+25%) for a fixed SE.

Index Terms—Optical fiber communication, hidden Markov models, nonlinearity mitigation, perturbation methods, achievable information rate.

I. INTRODUCTION

In optical fiber communications, the problems of channel modeling, nonlinearity mitigation, and capacity evaluation are strongly connected and still open. A good channel model is needed to understand nonlinear effects and devise effective nonlinearity mitigation strategies, with the final goal of optimizing the modulation format and the detection strategy. In turn, such an optimization is required to maximize the achievable information rate (AIR) over the channel and compute its capacity.

Ideally, a good channel model, suitable for capacity analysis, should be: i) information-theory friendly [2], including an

M. Secondini, E. Forestieri, and M. Ralli Camara are with the Institute of Communication, Information and Perception Technologies, Scuola Superiore Sant'Anna, Pisa, Italy, and with the National Laboratory of Photonic Networks, CNIT, Pisa, Italy (email: marco.secondini@sssup.it; forestieri@sssup.it; m.rallicamara@sssup.it).

E. Agrell is with the Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, Sweden (email: agrell@chalmers.se).

D. Marsella is with Nokia, Vimercate, Italy (email: domenico.marsella@nokia.com).

appropriate discrete representation of the input and output waveforms and providing an explicit and mathematically tractable expression of the conditional distribution of the output symbols given the input ones; ii) physically accurate for a wide range of (ideally, for *any*) input distributions. Unfortunately, the equations governing the propagation of light in optical fibers—namely, the nonlinear Schrödinger equation (NLSE), the Manakov equation, and their generalizations [3], [4]—meet only the second requirement. Thus, over the years, a number of different approximated models of the optical fiber channel have been developed [5]–[23], often with a conflicting view about the nature of nonlinear effects. As a matter of fact, the available models achieve different trade-offs between the two requirements, being either more accurate but less information-theory friendly, or the other way around [24], so that the problems of channel modeling and, in turn, of capacity evaluation remain open [25], [26].

The simplest channel models are based on the assumption that nonlinear interference (NLI) can be accurately represented by an additive white Gaussian noise (AWGN) term [5], [16], [19], [22], [23]. This assumption greatly simplifies the analysis and provides a good accuracy when applied to conventional systems. The main reason for this accuracy becomes apparent when replacing the ambiguous concept of "conventional" with the more explicit one of "designed for the AWGN channel". However, extending this approach to capacity analysis, which entails considering the best possible combination of modulation, coding, and detection, is not likely to give the desired results. In fact, it only provides a capacity lower bound which is actually achievable by conventional systems, but possibly very loose [27], [28].

No matter how accurate AWGN-like models might seem, if we aim for tighter capacity lower bounds or better modulation and detection strategies, we have to drop the AWGN assumption and look for any non-AWGN channel characteristics that can be exploited to improve system performance.

Deviations of NLI from the AWGN model have been observed and theoretically predicted in some previous works. In particular, the importance of phase noise (PN) and polarization rotation noise induced by NLI in dispersion-unmanaged links have been pointed out in [18], [21], [29], [30] and in [31], [32], respectively, showing that their temporal correlation properties can be exploited for the mitigation of interchannel NLI.

In this work, after reviewing some popular channel models and discussing their suitability for capacity analysis, we focus on a linear time-variant model [18], [20], which describes interchannel NLI in optical fibers as a kind of doubly dispersive
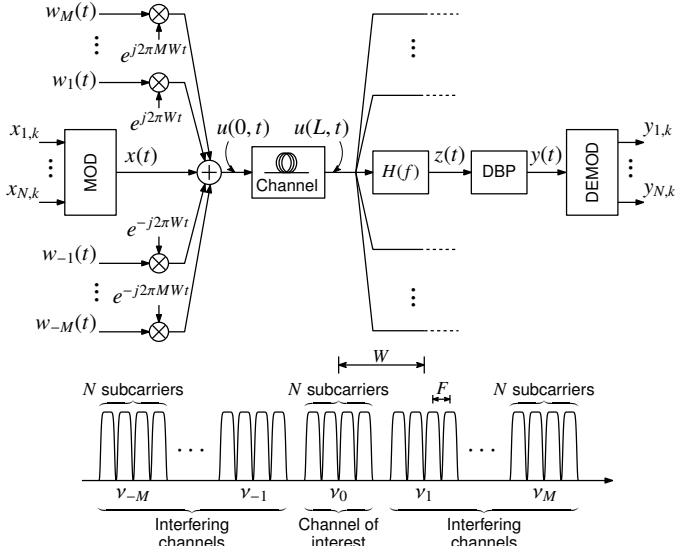
Fig. 1. Equivalent lowpass system model and corresponding WDM spectrum.

fading (see [33], [34], and references therein) which, at a given time and frequency, appears mainly as *phase and polarization noise* (PPN). Hence, we study the time and frequency coherence properties of this PPN—how fast it changes with time and frequency—and, based on it, we devise a simple NLI mitigation strategy that relies on the synergic effect of PPN compensation and subcarrier multiplexing (SCM) with symbol rate optimization. The PPN evolution is modeled by a first-order hidden Markov model (HMM), corresponding to the combination of a Wiener phase noise (PN) and an isotropic random walk over the Poincaré sphere [35]. Eventually, we verify the effectiveness of the proposed strategy in terms of achievable spectral efficiency (SE) and reach, resorting to a particle method for the numerical computation of the AIR [36]. Contrary to what is predicted by AWGN-like models, symbol rate optimization remains effective even for high-order modulation alphabets and Gaussian symbols if combined with PPN compensation [37]–[39], which allows us to derive tighter capacity lower bounds compared to the ones based on the AWGN assumption. The dependence of the achievable SE on the link length, amplifier spacing, and presence or absence of in-line dispersion compensation is also studied.

The paper is organized as follows. The system is described in Section II. Channel models and nonlinearity mitigation strategies are discussed in Section III and IV, respectively. The AIR computation method is detailed in Section V. Numerical results are presented in Section VI. Conclusions are finally drawn in Section VII.

II. SYSTEM DESCRIPTION

We consider a WDM system in which $2M+1$ signals (channels)—with channel spacing W and optical center frequencies $\nu_m = \nu_0 + mW$ for $m = -M, \ldots, M$—are independently generated and detected. Each WDM channel is further decomposed into N subchannels (subcarriers), all electronically generated by the same transmitter and jointly processed by the same receiver. The central WDM channel is arbitrarily selected

to be the channel of interest (COI), whose performance is under investigation, while the others are considered as interfering channels (IC) affecting the performance of the COI. The low-pass equivalent model with respect to the COI is depicted in Fig. 1, where only the block diagram for the COI has been fully represented. Here, the $2M$ complex baseband IC signals are denoted by $w_m(t)$, $m = \pm 1, \ldots, \pm M$, while the COI is simply denoted by $x(t)$. The transmitted signals can be also polarization multiplexed, in which case they should be denoted by bidimensional vectors whose components represent the signals on two orthogonal polarizations. For example, the COI signal should be interpreted as[1] $x(t) = [x_1(t), x_2(t)]^T$, where $x_1(t)$ and $x_2(t)$ are the aforementioned signal components on two orthogonal polarizations. In the following, for simplicity and in order to avoid using too many indices, all signals are to be interpreted as scalar values in the single-polarization (1-pol) case or as bidimensional vectors in the polarization-multiplexed (2-pol) case. When not clear from the context, the dimensionality of the signals will be explicitly stated. Using this convention and assuming a linear modulation format, the signal transmitted on the COI can be written as

$$x(t) = \sum_{n=1}^{N} \sum_{k=1}^{K} x_{n,k} p(t - kT) e^{j2\pi f_n t} \tag{1}$$

where $x_{n,k}$ is the k-th symbol (possibly a bidimensional vector) transmitted in the n-th subcarrier, $p(t)$ is the (scalar) supporting pulse (assumed to be the same for all subcarriers and polarizations of the COI), and $f_n = [n - (N + 1)/2]F$, $F \leq W/N$ being the frequency separation between the subcarriers.

The COI signal (1), which will be translated to an optical frequency ν_0, is then multiplexed with the IC signals generated by the other WDM transmitters to obtain the input waveform $u(0, t)$ (see Fig. 1). During propagation, the WDM signal undergoes attenuation due to fiber loss and amplification due to optical amplifiers. So, the signal power varies along the link according to $P(z) = a(z)P(0)$, where $P(0)$ is the launch power, and $a(z) = P(z)/P(0)$ is the link power profile. For example, for a link with ideal distributed amplification (IDA), $a(z) = 1$; for lumped amplification (LA), N_s spans of the same length L_s, and attenuation exactly compensated after each span, $a(z) = \exp[-\alpha(z \bmod L_s)]$. In the following we will normalize the complex envelope of the propagating signal according to the power profile. Specifically, denoting by $v(z, t)$ the complex envelope, the normalized one is $u(z, t) = v(z, t)/\sqrt{a(z)}$. Unless otherwise stated, we assume that all WDM channels are transmitted with the same power and use the same modulation format. Thus, the corresponding processes are all independent and statistically equivalent to (1), but translated to different central frequencies ν_m.

After propagation through the optical fiber link, the COI is selected by a demultiplexer filter $H(f)$, whose output $z(t)$ is processed by digital backpropagation (DBP) (see Section IV-A

---

[1] In the following, $(\cdot)^T$ denotes transpose, $(\cdot)^H$ transpose conjugate, and $(\cdot)^*$ complex conjugate. Boldface letters are used to denote row vectors collecting a sequence of (scalar or dual-polarization) symbols, with a subscript denoting the number of their elements, according to the notation $\mathbf{x}_k \triangleq (x_1, \ldots, x_k)$. Capital sans serif letters are used to denote matrices.

Parameter	Symbol	Value
Channel spacing	W	50 GHz
Number of subcarriers	N	variable
Symbol time	T	$T = N/W$
Number of WDM channels	$2M + 1$	5
Pulse shape	$p(t)$	$\mathrm{sinc}(t/T)$
Attenuation	α	$0.2\,\mathrm{dB/km}$
Dispersion	β_2	$-21.7\,\mathrm{ps^2/km}$
Nonlinear coefficient	γ	$1.27\,\mathrm{W^{-1} \cdot km^{-1}}$
Spontaneous emission coeff.	η	1 (IDA) or 1.6 (LA)

for details). The backpropagated signal $y(t)$ is then sent to a bank of N matched filters to obtain the output samples

$$y_{n,k} = y(t) \otimes \left[p^*(-t)e^{-j2\pi f_n t} \right]\big|_{t=kT} \qquad (2)$$

where $\otimes$ denotes convolution. The received samples are finally processed to recover the transmitted information according to one of the detection strategies described in Section V-A.

The propagation of the normalized complex envelope $u(z,t)$ of the WDM signal through the fiber link is governed by the NLSE in the 1-pol case [3], or the Manakov equation in the 2-pol case [4]. By using our convention on the signals, both can be written as

$$\frac{\partial u}{\partial z} = j\frac{\beta_2}{2}\frac{\partial^2 u}{\partial t^2} - ja(z)\gamma\|u\|^2 u + n(z,t) \qquad (3)$$

where β_2 is the group-velocity-dispersion parameter, γ is the nonlinear coefficient, $a(z)$ is the power profile along the link due to attenuation or amplification, and $n(z,t)$ is the (normalized) noise injected by optical amplifiers. The nonlinear coefficient is expressed as $\gamma = (8/9)2\pi\nu_0 n_2/(cA_{\mathrm{eff}})$, where n_2 is the Kerr coefficient, c the speed of light, A_{eff} the fiber effective area, and the factor of 8/9 accounts for rapid mixing of the polarization state on the Poincaré sphere [40]. In the 1-pol case, $u(z,t)$ is a scalar signal and $\|u\|^2 = |u|^2$. The system parameters considered in this work are reported in Table I.

III. CHANNEL MODELS

In the following, we discuss some approximated channel models, highlighting their differences, similarities, and suitability for capacity analysis.

A. Split-step Fourier method (SSFM)

Rather than a channel model, the SSFM is an efficient numerical method to solve (3) for any specific realization of the input signal and amplifier noise. It is commonly employed for the numerical simulation of optical fiber systems. On the other hand, its use for information theoretical analyses appears much more challenging, if not for the mere collection of signal statistics. A relevant exception is the derivation of a capacity upper bound for the optical fiber channel in [41]. The SSFM is also used for the implementation of DBP, as discussed in Section IV-A. In this work, the SSFM is used both for the simulation of fiber propagation according to (3) and for the implementation of DBP.

B. Regular perturbation (RP)

Linearization techniques are often employed to tackle the study of nonlinear effects. A widely used technique is the so-called RP method. It consists in expanding the complex envelope $u(z,t)$ in power series in γ, so that $u(z,t) = \sum_{k=0}^{\infty} \gamma^k u_k(z,t)$, where $u_0(z,t)$ is the linear solution. Replacing this expression in the NLSE/Manakov equation (3) and equating the terms with equal powers of γ that appear on both sides, one obtains a system of recursive linear differential equations to compute all u_k's starting from u_0. Often the expansion is truncated to the leading order, giving the approximation $u(z,t) \approx u_0(z,t) + \gamma u_1(z,t)$. The term $\gamma u_1(z,t)$ provides an approximate description of what is commonly referred to as *nonlinear interference (NLI) noise*—though, as we shall see, its interpretation as noise might be, in general, misleading. In general, this NLI can be further decomposed into several different terms, depending on the signal and/or noise components involved in its generation. In this work, we are primarily interested in interchannel NLI, generated by the interaction of the COI and IC signals.

Due to its simplicity, the RP method has been used in many papers to model fiber nonlinearity [8], [12]–[14], [17], [21], [42], [43]. Moreover, it has been shown that the order n RP solution coincides with the order $2n+1$ Volterra series solution [10]. Although taking into account only the leading term limits the applicability to the realm of small nonlinearity, the RP method is very general and, in the case of interchannel NLI, provides a simple deterministic relationship between the first-order term $\gamma u_1(z,t)$ and the IC and COI signals involved in its generation. The statistical channel model that is obtained from this approach depends on the additional approximations and assumptions that are commonly made to simplify the analysis, e.g., by assuming that the perturbation term is Gaussian or not, white or not, correlated with the COI signal or not, and so on. For instance, starting from the RP model, the presence of a PN component was recognized in [17], a significant time correlation of this PN component was highlighted in [21], and a representation of NLI as intersymbol interference (ISI) was given in [44]. As we shall see, all these features are naturally embedded in the frequency resolved logarithmic perturbation (FRLP) model [18], [20].

On the other hand, by ignoring these features and making more simplified assumptions, the RP method leads to the GN and EGN models described in the next subsection.

C. Gaussian noise (GN) and enhanced GN (EGN) models

A quite common trend is that of representing complicated effects, such as fiber nonlinearity, as an additional source of AWGN. This is the case, for instance, of the GN and EGN models [16], [22], [23]. Both models can be derived by using the RP method, with some additional assumptions and approximations. The GN model, in particular, relies on two main assumptions: i) NLI can be modeled as an additive Gaussian noise, nearly white over the COI bandwidth, and independent of the COI signal; ii) the NLI variance can be computed by assuming that, during propagation, dispersion turns all the signals involved in the NLI generation into

Gaussian processes. This approach, besides simplifying the analysis, usually provides a good accuracy when used to predict the performance of conventional systems.

Indeed, the second assumption entails that NLI is independent of the modulation format of the WDM signals, while a certain dependence on it has been observed and theoretically predicted by other models [17], [20], [21]. This is because dispersion ensures that the signal samples become only marginally but not jointly Gaussian [20], [21], [28]. As a result, the GN model turns out to be slightly pessimistic for low-order modulation alphabets and, in particular, for constant-envelop modulations. On the other hand, the EGN model drops this second assumption and relies only on the first one. Thus, it is more complex but also more accurate, correctly predicting the NLI dependence on modulation format [23] and allowing some nonlinear optimization (e.g., of the symbol rate [45]).

Despite their good features, the GN and EGN models have some limitations, so that the research of alternative models for the nonlinear optical fiber channel is still in progress. The key factor behind these limitations (but also behind the simplicity of the models) is the aforementioned AWGN assumption for the NLI. In fact, making this assumption entails that we accept a priori the impossibility of mitigating NLI and improving the performance beyond that of conventional detectors (optimized for the AWGN channel). This implies that the search for improved detectors, performance, and capacity bounds (or of a proof of impossibility thereof, should that be the case) cannot take place within the framework of this hypothesis. And that capacity calculations, based on it, should be interpreted only as capacity lower bounds, achievable by a mismatched decoder optimized for the AWGN channel [28].

Deviations from the AWGN model, while hard to recognize when considering only marginal channel statistics, become more evident when considering joint statistics [28]. The presence of a PN component with a long time correlation, the dependence of NLI on the symbols transmitted on the COI and IC signals, are all examples of such deviations. Finding and modeling those deviations, no matter how small they are, is the key factor to devise NLI mitigation strategies and design improved detectors. A possible approach is discussed in the next subsection.

D. Linear time-variant model

Another approach to the solution of (3) is that of replacing the nonlinear potential $\|u\|^2$ with a linearized approximation [9], [18], [20], [46]. By rewriting the optical propagating field as the sum of a COI and IC component $u(z,t) = s(z,t) + w(z,t)$, with $s(0,t) = x(t)$ and $w(0,t) = \sum_{m=\pm 1,\ldots,\pm M} w_m(t)e^{j2\pi mWt}$, and neglecting four-wave-mixing (FWM) terms, the evolution of the COI term can be approximated as[2]

$$\frac{\partial s}{\partial z} = j\frac{\beta_2}{2}\frac{\partial^2 s}{\partial t^2} - ja(z)\gamma(s^H s\, \mathsf{I} + w^H w\, \mathsf{I} + ww^H)s + n(z,t) \tag{4}$$

[2]The compact form of the nonlinear term is equivalent to the one provided in [15], in which cross-phase modulation (XPM) and cross-polarization modulation terms are explicitly defined.

where I is the 2×2 identity matrix in the 2-pol case, and the scalar unity in the 1-pol case. By further neglecting the intrachannel nonlinearity term $s^H s$ (as it is compensated for by single-channel DBP), and replacing the IC term $w(z,t)$ with its linearly propagated approximation $w_L(z,t)$, (4) reduces to

$$\frac{\partial s}{\partial z} = j\frac{\beta_2}{2}\frac{\partial^2 s}{\partial t^2} - ja(z)\gamma(w_L^H w_L\,\mathsf{I} + w_L w_L^H)s + n(z,t) \tag{5}$$

that is, a linear Schrödinger equation with a space- and time-variant potential. In the 1-pol case, all the quantities in (4) and (5) become scalar, so that $s^H s\,\mathsf{I} = |s|^2$, $w^H w\,\mathsf{I} + ww^H = 2|w|^2$, and $w_L^H w_L\,\mathsf{I} + w_L w_L^H = 2|w_L|^2$. Finally, assuming that the demultiplexer filter $H(f)$ removes the IC component without affecting the COI, the received signal after single-channel DBP can be expressed as

$$y(t) = \int_{-\infty}^{\infty} \mathsf{H}_w(f,t)X(f)e^{j2\pi ft}df + n(t) \tag{6}$$

where $\mathsf{H}_w(f,t)$ is a 2×2 time-variant transfer matrix, defined by a straightforward extension of the theory of linear time-variant systems [33] to the 2-pol case, and $n(t)$ is the AWGN obtained by integrating along z the optical amplifier noise $n(z,t)$. Given the characteristics of (5), H_w must be unitary.

The time-variant transfer matrix $\mathsf{H}_w(f,t)$ depends on the other WDM channel(s) $w(t)$ and accounts for XPM and cross-polarization modulation. The AWGN $n(t)$ accounts for the optical amplifier noise, whose statistics, in the linearized model (5), remain unchanged during propagation. Other non-explicitly modeled propagation effects are assumed to have been either compensated for (e.g., chromatic dispersion and deterministic intrachannel NLI) or included in the AWGN term by increasing its power spectral density (e.g., FWM). From the COI viewpoint, the nonlinear time-invariant optical fiber channel is seen as a linear time-variant one. This seeming paradox is explained by the fact that channel nonlinearity is accounted for by the dependence of $\mathsf{H}_w(f,t)$ on $w(t)$ [20]. However, assuming that $w(t)$ is unknown to both the transmitter and the receiver, such nonlinearity remains hidden, and the effect of $\mathsf{H}_w(f,t)$ is simply perceived as a linear distortion. Moreover, since $w(t)$ depends on time, also the transfer matrix $\mathsf{H}_w(f,t)$ depends on time.[3]

The model in (6) is substantially that of a doubly dispersive fading channel, often used in wireless communications, whose key features are the coherence time and bandwidth over which the channel remains strongly correlated [34]. This analogy may help to better understand the channel characteristics and behavior, as well as to devise improved transmission and detection strategies. For instance, a wide coherence bandwidth (compared to the COI) implies that the frequency dependence of $\mathsf{H}_w(f,t)$ can be neglected, with (6) corresponding to a

[3]If $w(t)$ were known, then the distortion could be seen as nonlinear, time-invariant, and fully deterministic, and could be exactly compensated for, e.g., by multiplying the received signal by the inverse matrix $\mathsf{H}_w^{-1}(f,t)$. This would be practically equivalent to performing an approximated multi-channel DBP. In principle, if $\mathsf{H}_w(f,t)$ were unitary—as suggested by the approximated propagation equation (5)—its compensation would not alter the statistics of the AWGN term and would thus completely remove the impact of nonlinearity. In this situation, signal–noise interaction, not explicitly modeled in (5), would become dominant and limit system performance.
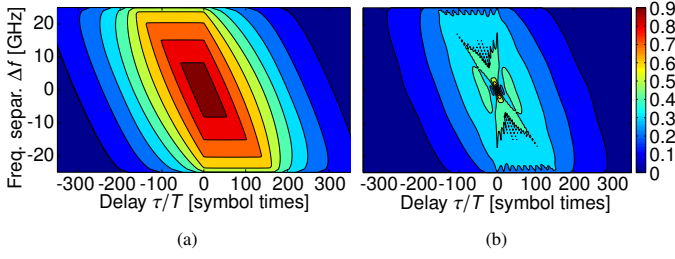
Fig. 2. XPM coherence: (a) IDA link; (b) 10x100km LA-DU link.

random time-variant phase and polarization rotation, here referred to as PPN. On the other hand, a long coherence time (many symbol times) implies that the time dependence of $H_w(f,t)$ can be neglected, with (6) being analogous to a channel affected by polarization mode dispersion.

In order to illustrate the coherence properties of the channel, we consider the 1-pol case, for which the transfer function can be analytically approximated through the FRLP model[4] as [20]

$$H_w(f,t) = e^{-j\theta(f,t)} \tag{7}$$

where the XPM term $\theta(f,t)$ depends on the IC signal $w(t)$ and link characteristics. Some general expressions for $\theta(f,t)$ and its statistics are provided in [18], [20]. The contour plots in Fig. 2 show the correlation between the values $\theta(f,t)$ and $\theta(f + \Delta f, t + \tau)$ of the XPM term as a function of the delay τ and frequency separation Δf.[5] Fig. 2(a) and (b) refer to a 1000 km IDA link and a $10 \times 100\,\mathrm{km}$ LA dispersion unmanaged (DU) link. Only the XPM term generated by the couple of closest ICs (i.e., those located at $\nu_{\pm 1} = \nu_0 \pm 50\,\mathrm{GHz}$) is considered. The coherence is quite substantial in the IDA link, but significantly reduced in the LA link.

IV. NONLINEARITY MITIGATION

In the following, we briefly discuss the nonlinearity mitigation strategies that are relevant for the present investigation, and their relations to the models presented in the previous section.

A. Digital backpropagation

DBP is a channel inversion technique, obtained by a digitally-emulated propagation of the received signal through a link with opposite coefficients and reversed spatial configuration [47], [48]. According to (3), such an inversion technique can perfectly cancel all the nonlinear effects accumulated

---

[4]In the FRLP model [20], the solution $s(z,t)$ of (5) is *resolved* into its *frequency* components. A *perturbation* approach is then separately applied to the *logarithm* of each component to study its propagation. The components are finally recombined by exploiting the linearity of (5). The temporal correlation properties of PPN can be studied also using a regular perturbation (RP) approach [21], [31]. While a comparison between the RP and FRLP approach is outside the scope of this work, the similar conclusions that can be achieved about the relevance and temporal coherence of PPN confirm the importance of the latter in the characterization of the optical fiber channel.

[5]The XPM process $\theta(f,t)$ is stationary in time but not in frequency (its statistics depend on the frequency distance from the ICs that generated it) [20]. Thus, the correlation is independent of t but depends on f, and is analyzed by holding f fixed in the middle of the COI bandwidth (conventionally set to $f = 0$), and letting Δf vary inside it ($-W/2 < \Delta f < W/2$).

during propagation, with the exclusion of signal–noise interaction. Its implementation is usually based on the SSFM for its good characteristics in terms of performance and complexity, discussed in Section III-A.

In principle, DBP may be either single-channel or multi-channel, depending on how many WDM channels are jointly backpropagated. The latter has relevant complexity issues, both in terms of hardware and processing resources, and is implementable only when all the involved channels are routed through the same optical path and jointly processed by the same receiver. Moreover, DBP can be implemented at the transmitter, at the receiver, or both (splitting the compensation between transmitter and receiver). In terms of performance, the split configuration is the most effective to mitigate the impact of signal–noise interaction, though the advantage may be negligible when considering single-channel DBP [49].

In this work, we employ single-channel DBP at the receiver, as shown in Fig. 1, to remove deterministic intrachannel nonlinear effects (the term $s^H s$ in (4)) from the received COI signal $z(t)$. With single-channel we mean that each WDM channel is independently backpropagated (indeed, only the COI is actually processed in our simulations), though each WDM channel can be further decomposed into many subcarriers that are jointly backpropagated. The implementation of DBP is based on the SSFM. Both the step size and sampling time are selected to obtain the best possible performance (i.e., they are reduced until the impact of a further reduction on the performance becomes negligible). Note that though the expansion bandwidth over which $z(t)$ is represented equals the inverse of the sampling time and may exceed the COI bandwidth, the demultiplexer filter $H(f)$ ensures that only the COI is actually backpropagated.

Some current challenges about DBP are the reduction of computational complexity [50], [51] and the inclusion of time-varying polarization effects [52], [53] and amplifier noise [54], [55]. None of those issues will be considered in this work.

B. Interchannel nonlinearity mitigation

The way in which nonlinearity is modeled has a great importance in determining how and if its impact can be mitigated. As mentioned in Section III, the available models often provide a conflicting view on the nature of nonlinear impairments. Nevertheless, when considering a simple receiver designed for an AWGN channel, all those models usually predict a similar performance. It does not matter if nonlinearity generates additive noise, PN, or interference: the impact is eventually the same and the distinction appears merely ontological. This observation plays in favor of the simplest available models, such as the GN and EGN ones. Yet, some models, by providing a different view on the nature of nonlinear impairments, also suggest the possibility of designing a better receiver that can at least partially mitigate them. In this case, the optimal detection strategy and the corresponding performance could actually depend on the adopted model. A trivial but instructive example is that of intrachannel nonlinearities: if we exclude DBP, their presence can be included in a GN-type model and their impact on a conventional receiver correctly evaluated [22]. On the

other hand, by recognizing their nature of nonlinear signal distortions induced by (3), it is possible to compensate them almost exactly through DBP [47], [48]. Further improvements are also achievable by accounting for signal–noise interaction, e.g., through maximum likelihood sequence detection [56] or stochastic DBP [54].

Similar considerations hold also in the case of interchannel nonlinearities, which are of primary interest in this work. A fairly common belief (perhaps at the base of considering some available capacity lower bounds as ultimate capacity limits) is that, without resorting to joint channel processing at the transmitter or receiver, it is impossible to mitigate interchannel NLI, which is perceived as an exogenous noise source, outside the user control. Although it is undoubtedly true that a complete compensation of interchannel NLI by multichannel DBP requires joint processing, there are at least two weak points in this kind of reasoning, which naturally lead to two different classes of nonlinearity mitigation strategies.

The first weak point is that, as indicated by several channel models and discussed in Section III, interchannel NLI is not truly equivalent to AWGN. For instance, its independence of the COI signal is just a useful working assumption made in the GN and EGN model to simplify the analysis. In fact, its generation involves, to a large extent, also the COI signal, on which it depends, causing the emergence of PPN and ISI, as discussed in Section III. This effects can be partly mitigated by some classical algorithms—commonly employed also in the linear regime to counteract the "linear version" of this kind of impairments, e.g., due to laser PN, chromatic dispersion and polarization mode dispersion—which can be considered in all respects as effective nonlinearity mitigation strategies. Examples of this first class of nonlinearity mitigation techniques, usually operating at the receiver end, are reported in [30], [57]–[59]. The main challenge is that channel variations induced by nonlinear impairments are typically much faster than those observed in linear channels, requiring algorithms with a higher adaptation speed.[6]

The second weak point is that, though we have excluded the possibility to deterministically control or cancel such interference by jointly processing the COI and IC signals, it is nevertheless possible to influence the statistics of the generated NLI by employing a proper combination of coding and modulation. Examples of this second class of NLI mitigation techniques, usually operating at the transmitter end, are the design of nearly constant-envelope modulation formats [60]; the ad hoc probabilistic shaping of QAM constellations over one [61]–[63] or several time slots [64]; the use of SCM with low-order modulation alphabets (this approach will be further discussed in the next subsection) [45]; and the use of twin waves [65] or conjugate data repetition [66] to partly cancel NLI. The main challenge with these techniques is that, in most of the aforementioned works, NLI is reduced at the expense of

introducing specific shaping, constraints, or redundancy in the transmitted signals, which often conflict with the optimization required to maximize the SE over an AWGN channel.

As a final remark, we note that the two classes of mitigation techniques have been usually developed and studied separately. Therefore, it is not guaranteed that the gains obtained by their separate use will add up when combined in the same system. For instance, all the shaping techniques aimed at minimizing the amplitude fluctuations of the signal mainly remove the same NLI component that is mitigated by PN compensation, so that little additional gains are obtained by combining the two techniques. On the other hand, it is also possible to use the two approaches synergistically, with an overall gain that is higher than the sum of the gains obtained by their separate use. This possibility will be further discussed in the next subsection.

C. Subcarrier multiplexing with symbol rate optimization

The idea of symbol-rate optimization in SCM, i.e., optimizing the symbol rate of each subcarrier to maximize the performance, has been investigated in several previous works [37], [38], [45], [67], [68]. From a theoretical point of view, and neglecting nonlinear effects, using one or several subcarriers is perfectly equivalent in terms of performance, given that the amplitude of the transfer function of the fiber and of the amplifiers is almost constant over each WDM channel bandwidth. At the same time, dividing each WDM channel into many low-rate subcarriers may have some practical advantages and disadvantages: it entails a natural parallelization of digital signal processing, which runs at a lower clock rate, and provides a finer granularity for mapping tributary channels on transport resources; but it also requires some additional processing effort and reduces the tolerance to laser PN. While these aspects are outside the scope of this work, we are particularly interested in the impact of SCM and symbol rate optimization in the presence of nonlinear effects.

One effect of decreasing the symbol rate of each subcarrier is that of changing the relative impact of XPM and FWM, which respectively decreases and increases, and vice versa. A possible mitigation strategy is, therefore, that of selecting the symbol rate that provides the best trade-off between XPM and FWM, hence minimizing the overall NLI [45]. This effect is maximally relevant for constant-envelope modulations, while it becomes negligible for high-order modulations and vanishes for Gaussian modulation symbols.[7]

Changing the symbol rate has another potential effect that is not related to the amount of generated NLI and does not vanish when considering high-order modulation alphabets or Gaussian symbols. This effect can be explained only by considering the non-AWGN nature of NLI, and can be exploited only by optimizing the receiver accordingly. Indeed, as explained

---

[6]As an example, typical time scales for polarization drifts due to thermal changes and mechanical perturbations in optical fibers range from days to milliseconds, while phase variations induced by laser PN usually take place over microseconds (see [35] and references therein). On the other hand, channel variations induced by interchannel nonlinearity are much faster (nanoseconds), as can be inferred from the channel coherence properties reported in Fig. 2.

[7]For Gaussian modulation symbols, $w(0, t)$ is a Gaussian process, whose statistical properties are fully and univocally determined by its power spectral density. Therefore, assuming that the power spectral density of $w(0, t)$ is independent of the number of subcarriers N (this is the case with ideal rectangular Nyquist pulses), also the statistical properties of the NLI generated by $w(0, t)$ must be independent of N. In fact, the dependence of NLI on symbol rate is correctly predicted by the EGN model, but completely absent in the GN model [45].

in Section III-D, the effect of NLI can be modeled as a doubly dispersive fading, characterized by the time-varying transfer matrix $H_w(f,t)$ in (6). The impact of $H_w(f,t)$ on the transmitted signal and its possible mitigation (e.g., by the compensation and equalization strategies discussed in the previous subsection) depend on the coherence properties of the channel—i.e., on the speed with which $H_w(f,t)$ changes with time and frequency—compared to the duration and bandwidth of the supporting pulse $p(t)$ in (1). Such duration and bandwidth are related in a reciprocal way and scale with the symbol rate of each subcarrier.

For instance, when a high number of low-rate subcarriers is considered, $p(t)$ has a narrow bandwidth and a long duration. In this case, $H_w(f,t)$ is practically constant over each subcarrier bandwidth, such that it can be considered as a simple phase and polarization rotation. At the same time, it changes rapidly over time, being hard to track and causing spectral broadening of the pulses and interchannel interference (ICI) between subcarriers. On the other hand, when a low number of high-rate subcarriers is considered, $p(t)$ has a large bandwidth and a short duration. In this case, $H_w(f,t)$ remains nearly constant over many symbol times, such that its slow variations cause no spectral broadening and can be more easily tracked. At the same time, it becomes highly selective in frequency, causing a long and harmful ISI. Under such circumstances, a possible NLI mitigation strategy is the optimization of the symbol rate to find the best trade-off between time and frequency coherence (hence, between ISI and ICI), making more effective the compensation of $H_w(f,t)$ by a detector with a limited complexity—in other words, minimizing the mismatch between the channel and the detector.

In this work, in an attempt to maximize the AIR and find tighter capacity lower bounds, we will only consider Gaussian modulation symbols and, therefore, we will exploit only the latter strategy and not the XPM–FWM trade-off effect.

V. ACHIEVABLE INFORMATION RATE AND SPECTRAL EFFICIENCY

The capacity of realistic fiber channels, notably those represented by the NLSE or Manakov equation (3), is not known, neither analytically nor numerically, due to the unavailability of an exact expression for the conditional distribution of the output samples given the input symbols. Therefore, optical information theory has been pursued along two main tracks: To characterize the capacity of simplified channel models and to establish bounds on the capacity of more realistic channel models.

One of the most important simplified channel models is when the nonlinear term in (3) is neglected. Then the channel is fully linear and the exact capacity is given by Shannon's famous formula [69]. The linear capacity is often used as a benchmark for more sophisticated capacity results, and it is accurate at low launch powers. Another simplified model is obtained by neglecting the dispersion in (3). This leads to a memoryless channel model, which has been extensively studied under an IDA assumption. Although the exact capacity is not known, several upper and lower bounds have been derived [70]–[72].

With more realistic channel models, the most common approach to capacity, and the one that will be investigated in this paper, is considering a fixed input distribution and a mismatched detector—a detector that is optimized for an approximated channel, known as the *auxiliary channel*, rather than for the true channel. Then the AIR (also known as auxiliary-channel lower bound) [73] in bits per channel use can be computed as[8]

$$I_q(X;Y) \triangleq \lim_{K \to \infty} \frac{1}{K} E\left\{ \log_2 \frac{q(\mathbf{y}_K|\mathbf{x}_K)}{q(\mathbf{y}_K)} \right\} \qquad (8)$$

Here, X and Y represent the discrete-time stochastic processes of which $\mathbf{x}_K$ and $\mathbf{y}_K$, respectively, are length-K realizations. For every $K = 1, 2, \ldots$, $p(\mathbf{x}_K)$ is the selected input distribution; $p(\mathbf{y}_K|\mathbf{x}_K)$ and $q(\mathbf{y}_K|\mathbf{x}_K)$ are the conditional distributions of the true and auxiliary channel, respectively; and $q(\mathbf{y}_K) = \int q(\mathbf{y}_K|\mathbf{x}_K)p(\mathbf{x}_K)d\mathbf{x}_K$ is the output distribution obtained by connecting the input process to the auxiliary channel. The expectation $E\{\cdot\}$ is taken with respect to the actual (unknown) joint distribution $p(\mathbf{x}_K, \mathbf{y}_K) = p(\mathbf{y}_K|\mathbf{x}_K)p(\mathbf{x}_K)$ of the input and output processes of the true channel.

The AIR (8) has some important properties, which hold for any true and auxiliary channel: i) It is a lower bound to the information rate $I(X;Y)$ and, therefore, to channel capacity; ii) its maximization over the auxiliary channel distribution $q(\mathbf{y}_K|\mathbf{x}_K)$ leads to the actual information rate; iii) its further maximization over the input distribution $p(\mathbf{x}_K)$ leads to channel capacity; iv) it is achievable over the true channel by the maximum a posteriori probability detector designed for the selected auxiliary channel; and v) it can be evaluated through numerical simulations, without an explicit knowledge of the statistics of the true channel, provided that $q(\mathbf{y}_K|\mathbf{x}_K)$ can be computed.

The most common auxiliary channel is the AWGN channel, in a fiber-optic context often motivated by the GN or EGN models. Lower bounds using this approach were presented in, e.g., [20], [74]–[76]. Even some of the earliest evaluations of channel capacity [5], [46], [77] were later identified as belonging to this class of lower bounds [2], [28]. Other lower bounds were obtained by numerically estimating the information rate using uniform binary input [78] or ring constellations [79].

In contrast to the abundance of lower bounds, no upper bounds on the capacity of (3) were known until 2015, when a closed-form upper bound on the per-sample capacity of the sampled NLSE was presented [41]. There is however still a sizeable gap between this upper bound and known lower bounds. This is because the upper bound, if interpreted as bit rate or SE, is proportional to the simulation sampling rate, which should be large for an accurate representation of the NLSE (several samples per symbol) [24].

The definition of AIR in (8) is based on the rather abstract concept of *channel use*. In practical cases, the channel is used many times by many users (employing different time slots and frequency bands), so that a more interesting quantity is

---

[8]In the following, we consider only the case in which all the subcarriers are independently modulated and (after DBP) detected. Thus, we focus on a generic subcarrier and, for simplicity of notation, drop the subcarrier index.

the SE, measured in bit per second per hertz and defined as the ratio between the information rate and the *time* and *bandwidth* required for each channel use. While required time and bandwidth are fuzzy notions when referred to a single pulse (channel use) [80], in the WDM scenario of Fig. 1 they are clearly identified by the symbol time T and channel spacing W (accommodating N subcarriers with spacing $F \leq W/N$). In this case, possible time and spectral broadening of the transmitted pulses outside these values are directly accounted for by a corresponding reduction of the information rate caused by the resulting ICI and ISI. Thus, given the AIRs $I_q^{(n)}$, for $n = 1, \ldots, N$, over the N different subcarriers, the achievable SE over the COI is

$$\text{SE}_q = \frac{\sum_{n=1}^{N} I_q^{(n)}}{WT} \tag{9}$$

For orthogonal signaling, the minimum time-frequency separation product between subcarriers is $FT = 1$ (obtained, for instance, with rectangular or sinc pulses). In this case, assuming no additional guard bands between WDM channels, $WT = N$ and the achievable SE in (9) (in bit per second per hertz) equals the average AIR (in bit per channel use) over the subcarriers. This is the configuration considered in the numerical results of Section VI.

When computing (8), both the input distribution and the auxiliary channel model can be arbitrarily selected. For what concerns the latter, it should be selected to provide a good approximation of the true channel, while still guaranteeing the computability of the conditional distribution $q(\mathbf{y}_K|\mathbf{x}_K)$. Three different discrete-time auxiliary channels, based on the approximated models described in Section III, are considered in the next subsection. On the other hand, the optimization of the input distribution is not specifically addressed in this work, and the practical choice of considering i.i.d. circularly-symmetric complex Gaussian (CSCG) samples on each subcarrier and polarization is made. This choice, though suboptimum at high power, is capacity-achieving at low power (in the linear regime). Moreover, it offers some advantages for the computation of the output distribution $q(\mathbf{y}_K)$ (as shown in the next subsection) and represents a limit that is practically achievable by dense quadrature amplitude modulation (QAM) constellations with probabilistic shaping.

A. Auxiliary discrete-time channel models

In the following, we consider three different discrete-time models for the COI of Fig. 1. The first one is a simple AWGN model and is inspired by the GN or EGN models. The others include some phase and/or polarization noise and are inspired by the FRLP model. Following [73], we refer to these models as *auxiliary channel* models to mean that they are used to compute the argument of the log function in (8) (and to design the detector), but not compute the expectation (that is, to simulate channel propagation), which is always done by using an accurate and complete numerical description of the *true channel* based on the SSFM. This ensures the validity and achievability of the bound (8), regardless of the assumptions

and approximations—often motivated by heuristic considerations or simply aimed at simplifying the model—on which the auxiliary channels are based.

The auxiliary models include some constant parameters whose values can, in principle, be derived from the equations describing the EGN and FRLP models. Since such a derivation is not essential to our aims, and would not guarantee the most accurate approximation of the channel, we prefer to resort to the numerical estimation procedure described in Section V-B. Such a procedure aims at minimizing the mismatch between the true and auxiliary channel, offering the advantage that it does not require any knowledge of the physical channel, making it suitable also for an experimental environment, possibly accounting for other effects which are neglected in the equations.

All the models refer to a generic subcarrier, such that the corresponding parameters depend on the subcarrier index. However, as explained in footnote 8, the subcarrier index is always omitted. The first two models are 1-pol models, but can also be independently applied to each polarization of a 2-pol system—in fact, neglecting any polarization effect during propagation. The third model, on the other hand, considers the two polarization components jointly, and includes some polarization effects.

1) AWGN model: When neglecting nonlinear effects, or when representing them through the GN or EGN model, the channel of Fig. 1 reduces to an AWGN channel. The input–output relationship for a 1-pol system and a generic subcarrier can thus be modeled as

$$y_k = cx_k + n_k \tag{10}$$

where $c = ae^{j\theta}$ is a complex constant representing a generic attenuation/amplification a and a phase rotation θ, and n_k are i.i.d. CSCG variables with variance σ_n^2. This is a memoryless channel, whose joint conditional distribution factorizes into the product of the marginal distributions as

$$q(\mathbf{y}_K|\mathbf{x}_K) = \prod_{k=1}^{K} \frac{1}{\pi\sigma_n^2} \exp\left(-\frac{|y_k - cx_k|^2}{\sigma_n^2}\right) \tag{11}$$

For i.i.d. CSCG input variables with variance σ_x^2, also the output variables are i.i.d. CSCG, with variance $\sigma_y^2 = a^2\sigma_x^2 + \sigma_n^2$, and their joint distribution factorizes into the product of the marginal distributions as

$$q(\mathbf{y}_K) = \prod_{k=1}^{K} \frac{1}{\pi\sigma_y^2} \exp\left(-\frac{|y_k|^2}{\sigma_y^2}\right) \tag{12}$$

2) PN model: A more accurate representation of XPM can be obtained by using the FRLP model (7). The frequency dependence of the XPM term $\theta(f, t)$ within each subcarrier bandwidth and its time dependence within a pulse duration cause ISI and ICI, respectively. By neglecting such ISI and ICI—whose minimization is the purpose of optimizing the number of subcarriers—the input–output relationship for a 1-pol system and a generic subcarrier can be modeled as

$$y_k = ae^{j\theta_k}x_k + n_k \tag{13}$$

where a and n_k are as in the AWGN channel (10), and θ_k is the k-th sample of a discrete-time process that accounts for XPM. In principle, the statistics of the XPM process could be derived from the FRLP equations. In particular, an important property of the process is its long temporal coherence which, as shown in Fig. 2, might span many symbol times. Moreover, as explained in [20], the process can be approximately assumed to be real-valued and have Gaussian samples. For the sake of simplicity, we model it as a random walk (the discrete-time analog of a Wiener process) on a circle

$$\theta_k = \theta_{k-1} + \Delta\theta_k \quad \mod 2\pi \tag{14}$$

with i.i.d. real zero-mean Gaussian increments $\Delta\theta_k$ with variance σ_θ^2.

The resulting channel model (13) is a first-order HMM, with θ_k playing the role of the (hidden) state variable and taking continuous values in the range $[0, 2\pi)$. In this case, the conditional distribution does not factorize into the marginal distributions, but can be expressed as

$$q(\mathbf{y}_K|\mathbf{x}_K) = \int q(\mathbf{y}_K, \theta_K|\mathbf{x}_K)d\theta_K \tag{15}$$

where the term in the integral is recursively computed as

$$q(\mathbf{y}_k, \theta_k|\mathbf{x}_k) =$$
$$\int q(\mathbf{y}_{k-1}, \theta_{k-1}|\mathbf{x}_{k-1})q(y_k|x_k, \theta_k)q(\theta_k|\theta_{k-1})d\theta_{k-1} \tag{16}$$

for $k = 1, \ldots, K$, with

$$q(y_k|x_k, \theta_k) = \frac{1}{\pi\sigma_n^2}\exp\left(-\frac{|y_k - ae^{j\theta_k}x_k|^2}{\sigma_n^2}\right) \tag{17}$$

$$q(\theta_k|\theta_{k-1}) = \frac{1}{\sigma_\theta\sqrt{2\pi}}\exp\left(-\frac{(\theta_k - \theta_{k-1})^2}{2\sigma_\theta^2}\right) \tag{18}$$

No prior knowledge about θ_0 is assumed, so the recursion is initiated by setting $q(\mathbf{y}_0, \theta_0|x_0) \triangleq q(\theta_0)$ uniform in $[0, 2\pi)$. On the other hand, considering i.i.d. CSCG input variables, the output distribution is not affected by the presence of PN and can be expressed as in (12).

The random walk in (14) is the same model commonly adopted for laser PN. In fact, (14) could even represent the combination of XPM and laser PN, with σ_θ^2 being the sum of the variances related to the two processes. This is, however, a clear oversimplification of the actual evolution of the XPM term. In a random walk, the process value at a given time is equally determined by all the previous steps, so that its standard deviation keeps increasing with time. When the random walk is wrapped to a finite interval—the $[0, 2\pi)$ interval in our case—this yields a uniform distribution of the process value after a few steps. While this is a reasonable assumption for laser PN, it is not so for XPM. In fact, the XPM term at a given time is determined only by the values of the IC signals over a finite window around that time, so that its standard deviation remains constant over time—in practice, the phase fluctuates in a small interval around its mean value. Such a behavior could be modelled by adding a mean-reverting term (proportional to the difference between the mean and current values of the phase) to (14), obtaining a first-order

autoregressive (AR) process. Eventually, a higher-order AR process could be employed to better account for the actual XPM statistical properties and for the possible combination of XPM and laser PN. In this work, the choice of the simple (but clearly mismatched) model in (14) is justified a posteriori by the performance gains shown in Section VI. Possible improvements achievable by using more refined models will be investigated in a future work.

3) PPN model: The model (13) can be generalized to account for polarization effects induced by the Manakov equation (3). In this case, considering two-dimensional input and output vectors x_k and y_k, and assuming that the overall effect of the channel is still that of inducing a unitary transformation and adding AWGN as in (6), the input–output relationship for a generic subcarrier can be modeled as

$$y_k = ae^{j\theta_k}\mathsf{J}_k x_k + n_k \tag{19}$$

where a and θ_k are as in the PN model (13), n_k is a two-dimensional noise vector, whose elements are independent and distributed as in the AWGN and PN model, and J_k is a 2×2 complex unitary Jones matrix that accounts for a generic rotation of the state of polarization. In analogy to the random-walk model adopted in (14), the evolution of J_k is modeled as an isotropic random walk over the Poincaré sphere [35]

$$\mathsf{J}_k = \exp(j\boldsymbol{\alpha}_k \cdot \vec{\boldsymbol{\sigma}})\mathsf{J}_{k-1}$$
$$= \exp\begin{pmatrix} j\alpha_{1,k} & \alpha_{3,k} + j\alpha_{2,k} \\ -\alpha_{3,k} + j\alpha_{2,k} & -j\alpha_{1,k} \end{pmatrix}\mathsf{J}_{k-1} \tag{20}$$

where $\vec{\boldsymbol{\sigma}}$ is the Pauli spin vector in Stokes space [81], and the real innovation parameters $\boldsymbol{\alpha}_k = (\alpha_{1,k}, \alpha_{2,k}, \alpha_{3,k})$ are drawn independently from a zero-mean Gaussian distribution with variance σ_p^2 at each time instance k.

The obtained model is still a first-order HMM, whose hidden state is represented by the pair (θ_k, J_k). The joint conditional distribution $q(\mathbf{y}|\mathbf{x})$ can thus be expressed as in (15)–(16), but replacing θ_k with (θ_k, J_k) and extending integrations to the corresponding state space. The conditional distribution (17) must be replaced by

$$q(y_k|x_k, \theta_k, \mathsf{J}_k) = \frac{1}{\pi^2\sigma_n^4}\exp\left(-\frac{\|y_k - ae^{j\theta_k}\mathsf{J}_k x_k\|^2}{\sigma_n^2}\right) \tag{21}$$

and the transition probability $q(\mathsf{J}_k|\mathsf{J}_{k-1})$ could be derived from (20).[9] As in the 1-pol case, for i.i.d. CSCG input symbols, the output distribution is not affected by the unitary transformation in (20) and can be expressed as the product of two distributions as in (12), one per each polarization component.

We remark that the same considerations expressed at the end of Section V-A2—the mismatch between the random-walk model (14) and the actual PN evolution, and the better modelling options offered by AR processes—hold also for (20) and the actual PPN evolution.

---

[9]For the particle method adopted in the next section to compute (15)–(16), an explicit expression of the transition probability is not required, as the simple algorithmic representation based on (20) is sufficient.

B. Numerical computations

The problem of computing the AIR (8) and the corresponding achievable SE (9) cannot be solved analytically, as the joint input–output distribution $p(\mathbf{x}_K, \mathbf{y}_K)$ of the true channel—with respect to which the expectation in (8) must be computed—is generally unknown. However, an accurate numerical estimate can be obtained by relying on the asymptotic equipartition property [82, Ch. 3] and following the procedure described in [73]. One additional difficulty is that the HMMs in (13) and (19) are characterized by a continuous state space, such that the recursive computation of (16) would require the computation of a long series of nested integrals. Among several possible approaches to address this issue[10], we resort to the particle method proposed in [36], because it can be easily extended to general HMMs, its accuracy can be arbitrarily increased by increasing the number of particles, and its complexity scales fairly well with the dimensionality of the state space. The overall procedure for the computation of (8) (for a 1-pol system, a single subcarrier, and the PN model) is briefly summarized in the following.

1) Draw two long input sequences $\mathbf{x}_K = (x_1, \ldots, x_K)$ and $\mathbf{x}'_{K'} = (x'_1, \ldots, x'_{K'})$ of i.i.d. CSCG samples.
2) Compute the corresponding output sequences $\mathbf{y}_K = (y_1, \ldots, y_K)$ and $\mathbf{y}'_{K'} = (y'_1, \ldots, y'_{K'})$ by using the SSFM to simulate the true channel.
3) Estimate the parameters a, σ_n^2, σ_θ^2 of the auxiliary channel (13) by using the sequences $\mathbf{x}'_{K'}$ and $\mathbf{y}'_{K'}$.
4) Compute the conditional distribution $q(\mathbf{y}_K|\mathbf{x}_K)$ in (15) by using the particle method [36]. In particular, for $k = 1, \ldots, K$:
   a) Generate the new particle set for the representation of θ_k by drawing samples from (18).
   b) Update the state metric for each particle by using (17).
   c) Use normalization and resampling [36] to avoid numerical issues.
5) Compute the corresponding output distribution $q(\mathbf{y}_K)$ by using (12).
6) Estimate the AIR as

$$\hat{I}_q(X; Y) = \frac{1}{K} \log_2 \frac{q(\mathbf{y}_K|\mathbf{x}_K)}{q(\mathbf{y}_K)} \quad (22)$$

and the corresponding achievable SE in (9).

Some remarks follow.

- For the AWGN model, the particle approach is not required, as the conditional and output distributions can be directly computed from (11) and (12), respectively.
- For 2-pol systems, the AWGN or PN models can be independently applied to each polarization component of each subcarrier. Moreover, the PPN model can be jointly applied to vector samples, with a straightforward extension of the described numerical procedure, i.e., including a particle representation of the matrix J_k, replacing (18) with (20) and (17) with (21), and estimating the additional parameter σ_p^2 of the PPN model.
- In Step 3, channel parameters are estimated by using a two-step procedure. Given each input variable x'_k, the output squared modulus $\|y'_k\|^2$ has a noncentral chi-squared distribution that does not depend on the presence of PPN. Thus, in the first step, a and σ_n^2 are estimated by numerically maximizing the corresponding likelihood

$$\hat{\sigma_n^2} = \underset{\sigma_n^2}{\operatorname{argmax}} \sum_{k=1}^{K'} \log \left[\frac{1}{\sigma_n^2} \exp \left(-\frac{\|y'_k\|^2 + a^2\|x'_k\|^2}{\sigma_n^2} \right) \right.$$
$$\left. \left(\frac{\|y'_k\|^2}{a^2\|x'_k\|^2} \right)^p I_{2p} \left(\frac{2a\|x'_k\|\|y'_k\|}{\sigma_n^2} \right) \right] \quad (23)$$

where $a = \sqrt{(\sigma_y^2 - \sigma_n^2)/\sigma_x^2}$; σ_x^2 and σ_y^2 are the variances of the input and output samples, respectively; $p = 0$ or 0.5 for the 1-pol and 2-pol case, respectively; and $I_m(\cdot)$ is the modified Bessel function of the first kind of order m. In the second step, the remaining parameter σ_θ^2 (and, in the 2-pol case, σ_p^2) is estimated by numerically maximizing the AIR, which corresponds to minimizing the mismatch between the true and auxiliary channels.

- For systems with N subcarriers, N independent input sequences are drawn to generate the input signal (1). The corresponding output signal is computed by using the SSFM, and the N output sequences are then extracted by the bank of matched filters according to (2). The AIR is finally computed independently for each subcarrier and averaged according to (9).

VI. NUMERICAL RESULTS

In this section, by following the procedure detailed in Section V-B, we compute the achievable SE for the COI of the system described in Section II under constrained modulation (CSCG input symbols) and mismatched decoding, considering the system parameters provided in Table I, 1-pol and 2-pol transmissions, and different link configurations. We compare the SE that can be achieved by using three different detectors, matched to the AWGN, PN, and PPN auxiliary channel models provided in Section V-A, and by optimizing the number of subcarriers (or equivalently, the symbol rate per subcarrier). For the considered pulse shape, according to (9), the achievable SE over the COI equals the average AIR over the subcarriers. All the plots report the SE per polarization. DBP is implemented as explained in Section IV-A. SE values and channel parameters are estimated by averaging over $K = 100000$ and $K' = 10000$ symbols per subcarrier, respectively. The transmitted waveform is sampled at $8N$ samples per subcarrier symbol, corresponding to a fixed sampling rate of $8W$, sufficiently wider than the WDM signal bandwidth of $5W$ to account for spectral broadening during propagation. No relevant performance changes were observed by considering larger sampling rates. The procedure adopted for the numerical simulation entails a periodization

---

[10]A typical approach is the quantization of the state space which, however, becomes infeasible in high-dimensional spaces—in fact, it provides good results for the monodimensional state space of (13) but is already too complex for the four-dimensional state space of (19). An alternative low-complexity approach is the use of properly selected parametrized canonical distributions (e.g., Gaussian or Tikhonov) to represent (16), but it usually entails some approximations and must be specifically developed and tested for each model. For a general discussion, see [83] and references therein.

of the transmitted signal, with more than $K + K' = 110000$ symbols per subcarrier in each period. The resulting period is always much longer than the overall channel memory, so that numerical artifacts due to this periodization are avoided.

First, we consider a 1000 km IDA link with a 2-pol transmission (the 1-pol case was shown in [1]). Fig. 3(a) shows the SE as a function of the launch power (per each polarization of each WDM channel) for the PPN detector and different numbers of subcarriers N. As a comparison, we also report the SE obtained with the AWGN detector and the SE lower bound computed as in [29] and based on an independent processing of the two polarizations. Note that, for CSCG input symbols, changing N does not alter the input distribution. As a consequence, also the amount of generated NLI, the AIR with the AWGN detector, and the (unknown) information rate remain unchanged. However, when considering the PPN detector, the mismatch between channel and detector depends on N. In fact, ISI and ICI, unaccounted for by the model, respectively decreases and increases with increasing N. An optimum value of N exists, which, by minimizing the overall interference, minimizes the mismatch and maximizes the AIR and, hence, the SE. Thanks to the large coherence time and bandwidth of IDA links (see Fig. 2), a significant SE gain (the difference between the SE values obtained by the PPN and AWGN detectors at their respective optimal launch power) of about 0.8 bit/s/Hz/pol is achieved for $N = 4$. This is about twice the SE gain of 0.4 bit/s/Hz/pol predicted by the lower bound of [29].

The gain obtained with the PPN detector is due to its capability to accurately track the time evolution of the state of the auxiliary channel model (19), represented by the phase θ_k and the rotation matrix J_k. An example of such evolution, extracted as a byproduct of the AIR computation algorithm,[11] is represented in Fig. 3(b) and (c) for one of the central subcarriers of a 4-subcarrier system and an optimum launch power of $-7\,\mathrm{dBm}$. In particular, Fig. 3(b) reports the evolution of the phase θ_k over 20000 symbol times (together with a zoomed version and its estimated autocorrelation function), while the red points on the Poincaré sphere in Fig. 3(c) represent the rotations of the Cartesian basis vectors S_1, S_2, and S_3 in the Stokes space induced by the estimated J_k over the same time interval. As anticipated in Sections V-A2 and V-A3, both θ_k and J_k do not evolve as true random walks, as postulated in (18) and (20), but rather show limited fluctuations around their mean value, being therefore more similar to AR processes. Despite such an apparent mismatch with respect to the true channel, the adopted PPN model provides a relevant gain in terms of achievable SE, thanks to the significant temporal correlation of the phase values (and of the polarization rotations, whose correlation is not shown in the figure). The possibility to get higher gains by adopting a more refined channel model—which describes more accurately the PPN evolution observed in Fig. 3(b)–(c) and the coherence properties shown in Fig. 2—is left as a future work.

For the same configuration as in Fig. 3(a)–(c) and a fixed

launch power of $-7\,\mathrm{dBm}$, Fig. 3(d) shows how the characteristics of the channel seen by one of the central subcarriers—the estimated parameters σ_n^2 (reduced by the contribution σ_{ASE}^2 of the linearly accumulated amplifier noise and normalized to the signal power $|a|^2\sigma_x^2$), σ_θ^2, and σ_p^2 (tripled to account for the approximately triple impact that the three polarization rotation terms $\alpha_{1,k}$, $\alpha_{2,k}$, and $\alpha_{3,k}$ have on the signal compared to the PN term θ_k) of the PPN model (19)—depend on N. By increasing N, the frequency coherence of PPN over each subcarrier increases (as the bandwidth becomes narrower), so that a larger portion of NLI can be actually represented as pure (frequency-independent) PPN. This explains the decrease of σ_n^2 and, in part, the increase of σ_θ^2 and σ_p^2. The latter is also due to the decrease of the symbol rate of each subcarrier, which reduces the temporal coherence of PPN over a symbol time. The best tradeoff is obtained for $N = 4$ subcarriers, corresponding to the highest SE in Fig. 3(a).

The maximum achievable SE—i.e., the SE obtained by optimizing the launch power and the number of subcarriers—depends on the link and system configuration and on the detection strategy. For instance, Fig. 4 shows the maximum SE achievable by different detectors in DU links as a function of (a) the link length (for an IDA link) and (b) the amplifier spacing (for a 1000 km LA link). For the sake of comparison, both 1-pol and 2-pol transmissions are considered. For both cases, the SE lower bound of [29] is also reported. For IDA links, relevant gains are obtained both in the 1-pol and 2-pol scenarios by the PN and PPN detectors, respectively. At 4000 km, the SE gain is about 1.1 bit/s/Hz in the 1-pol case, and 1 bit/s/Hz/pol in the 2-pol case with PPN detection. In the latter case, the gain reduces to about 0.6 bit/s/Hz/pol if the two polarizations are independently processed by PN detection, neglecting polarization effects. By comparison, the gains predicted by the lower bound of [29] are 0.6 and 0.4 bit/s/Hz/pol in the 1-pol and 2-pol cases, respectively. The gains slightly increase for longer distances, and appear even more relevant if measured in terms of reach: for a target SE of 7 bit/s/Hz/pol, the reach is doubled in the 2-pol case (with PPN detection), and more than doubled in the 1-pol case. On the other hand, lower gains are obtained in LA links, due to their lower time and frequency coherence (see, for instance, Fig. 2). For the 1000 km link considered in Fig. 4(b), the SE gain in the 2-pol case gradually decreases from 0.8 to 0.2 bit/s/Hz/pol when increasing the amplifier spacing from zero (actually corresponding to an IDA configuration) to 100 km. For such a long spacing, similar gains are predicted also by the lower bound of [29]. A similar behavior is observed in the 1-pol case.

Fig. 5 provides the same comparison as in Fig. 4(a) but for an $N_s \times 60\,\mathrm{km}$ (a) LA-DU and (b) LA dispersion-managed (DM) link. In the DM link, dispersion is fully compensated after each span by $10.2\,\mathrm{km}$ of dispersion-compensating fiber (DCF) with parameters $\alpha = 0.57\,\mathrm{dB/km}$, $\beta_2 = 127.5\,\mathrm{ps}^2/\mathrm{km}$, and $\gamma = 6.5\,\mathrm{W}^{-1} \cdot \mathrm{km}^{-1}$. An additional optical amplifier with $\eta = 1.6$ is inserted at the input of each DCF to set the launch power in the DCF always at 4 dB below that in the transmission fiber. For the DU link, at 4800 km, an SE gain of about 0.3 bit/s/Hz/pol is obtained in both the

---

[11]At each sampling time, the algorithm generates a particle representation of the state distribution. The actual estimate is then computed as the expectation over the particles.
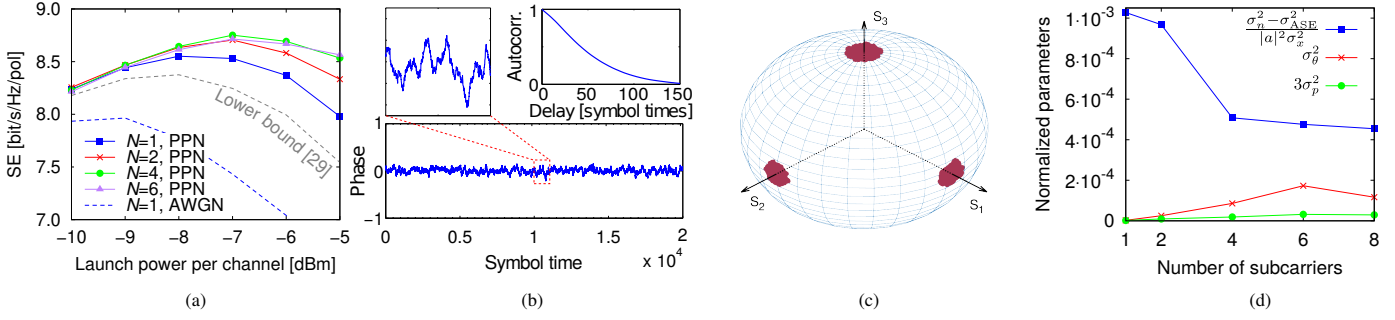
Fig. 3. 1000 km IDA link with 2-pol transmission and PPN detection: (a) SE vs. launch power for different number of subcarriers; (b) phase (and its autocorrelation in the inset) and (c) polarization rotations induced by fiber nonlinearity at $-7\,\mathrm{dBm}$ and $N=4$; (d) estimated parameters of the PPN auxiliary channel vs. N at $-7\,\mathrm{dBm}$.
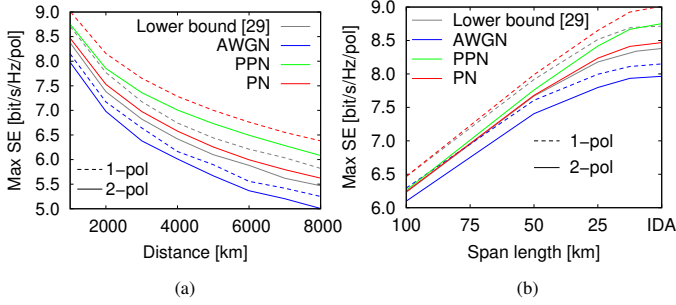

Fig. 4. Dependence of the achievable SE on: (a) distance, for an IDA link; (b) amplifier spacing for a 1000 km LA link.
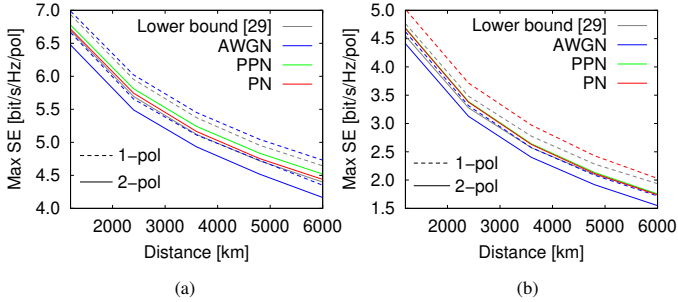

Fig. 5. Dependence of the achievable SE on distance for an $N_s \times 60\,\mathrm{km}$ (a) LA-DU and (b) LA-DM link.

1-pol and 2-pol cases, which can be alternatively exploited to increase the reach by 25% (up to 6000 km) keeping the SE fixed. The gain reduces to 0.2 bit/s/Hz/pol (+20% reach) in the 2-pol case if independent PN detection on each polarization is employed. By comparison, the gains predicted by the lower bounds in [29] are about 0.1 bit/s/Hz/pol lower. On the other hand, in the DM link, the gain provided by the PPN strategy is reduced (about 0.2 bit/s/Hz/pol at 4800 km) due to the lower temporal coherence. This is in contrast with the predictions made in [38] in terms of potential peak SNR gain for single-span links, which should behave similarly to DM links. As mentioned in [38], those "potential" SNR gains are based on ideal PPN removal and do not account for the temporal correlation of PPN, which is quite short in DM links. As a result, they cannot be fully translated into *achievable* SE gains.

The optimum number of subcarriers considered in the

previous cases depends on the link configuration: it increases with distance from 4 to 8 in Fig. 4(a) and from 2 to 6 in Fig. 5(a), while it is $N=1$ at any distance in Fig. 5(b).

VII. DISCUSSION AND CONCLUSIONS

We have discussed the importance of channel models for nonlinearity mitigation in WDM fiber-optic systems, and the need to drop the AWGN assumption to devise improved detection strategies. A step toward this direction is done by resorting to a linear time-variant model, which suggests considering interchannel NLI as linear doubly dispersive (in time and in frequency) fading. This analogy helps to understand the characteristics of the channel in terms of its coherence time and bandwidth, and to devise improved transmission schemes borrowed from wireless communications. On this basis, we have derived some simple discrete-time auxiliary channel models and employed SCM, optimizing the number of subcarriers, to maximize the AIR and found some improved SE lower bounds. For IDA links, thanks to the high time–frequency coherence of interchannel NLI, the detection strategies based on the considered channel models provide about 1 bit/s/Hz/pol SE gain at 4000 km, and double the reach for a fixed SE of 7 bit/s/Hz/pol compared to a detection based on the AWGN model. The gain increases with distance. In LA links, the coherence time and bandwidth of interchannel NLI are reduced, so that lower gains are obtained. The gain increases with the number of spans, decreases with the span length, and is further reduced by in-line dispersion compensation. Similar behaviors are observed in 1-pol and 2-pol systems, provided that, in the latter, the two polarizations are jointly detected to account for polarization effects. All the gains were measured for a 5-channel WDM system to keep the computational complexity low, but are expected to increase when more channels are present [30].

Another interesting result of this study is the observed synergy of SCM and PPN compensation. When detection is optimized for the AWGN channel, the benefits provided by SCM in terms of nonlinearity mitigation are reduced for high-order QAM modulations, and vanish as the modulation alphabet approaches a Gaussian distribution [45]. On the other hand, when employing PN or PPN detection, SCM allows to obtain the best trade-off between coherence time and bandwidth by optimizing the number of subcarriers, hence fully revealing

the NLI mitigation capabilities of these detection strategies and maximizing the achievable SE and reach.

This study also suggests that further optimizations and gains are possible, which are left as a future work. Potential improvements include, for instance, accounting for phase and polarization correlation among subcarriers or considering higher-order AR processes to model the evolution of the channel state. This can be done either analytically, based on the FRLP model, or from a statistical analysis of the HMM parameters based on simulated/experimental data. Finally, unlike conventional fading, the channel response derived from the FRLP model does not depend on environmental conditions, out of the user control (e.g., multipath and shadowing effects), but rather on the actual signals transmitted by the other users (channels, according to WDM terminology). Even excluding unpractical joint modulation or detection, this peculiarity gives the possibility of controlling the channel response by adopting a proper combination of coding and modulation on each channel. In this context, an interesting but still unexplored possibility is that of using constellation shaping to increase channel coherence, rather than to decrease NLI. This approach might have a better synergy with the ones discussed in this work and provide additional SE gains.

REFERENCES

[1] M. Secondini, E. Agrell, E. Forestieri, and D. Marsella, "Fiber nonlinearity mitigation in WDM systems: Strategies and achievable rates," in *Proc. Eur. Conf. Opt. Commun. (ECOC)*, Sept. 2017.

[2] E. Agrell, G. Durisi, and P. Johannisson, "Information-theory-friendly models for fiber-optic channels: A primer," in *IEEE Information Theory Workshop (ITW)*, Apr. 2015.

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

[4] E. Wang and C. R. Menyuk, "Polarization evolution due to the Kerr nonlinearity and chromatic dispersion," *J. Lightwave Technol.*, vol. 17, pp. 2520–2529, Dec. 1999.

[5] A. Splett, C. Kurtzke, and K. Petermann, "Ultimate transmission capacity of amplified optical fiber communication systems taking into account fiber nonlinearities," in *Proc. Eur. Conf. Opt. Commun. (ECOC)*, vol. 2, pp. 41–44, 1993.

[6] K. V. Peddanarappagari and M. Brandt-Pearce, "Volterra series transfer function of single-mode fibers," *J. Lightwave Technol.*, vol. 15, pp. 2232–2241, Dec. 1997.

[7] A. Cartaxo, "Cross-phase modulation in intensity modulation-direct detection WDM systems with multiple optical amplifiers and dispersion compensators," *J. Lightwave Technol.*, vol. 17, pp. 178–190, Feb. 1999.

[8] R. Holzlöhner, V. S. Grigoryan, C. R. Menyuk, and W. L. Kath, "Accurate calculation of eye diagrams and bit error rates in optical transmission systems using linearization," *J. Lightwave Technol.*, vol. 20, pp. 389–400, Mar. 2002.

[9] A. G. Green, P. B. Littlewood, P. P. Mitra, and L. G. L. Wegener, "Schrödinger equation with a spatially and temporally random potential: Effects of cross-phase modulation in optical communication," *Phys. Rev. E*, vol. 66, Oct. 2002.

[10] A. Vannucci, P. Serena, and A. Bononi, "The RP method: a new tool for the iterative solution of the nonlinear Schrödinger equation," *J. Lightwave Technol.*, vol. 20, pp. 1102–1112, July 2002.

[11] K.-P. Ho, "Error probability of DPSK signals with cross-phase modulation induced nonlinear phase noise," *J. Sel. Topics Quantum Electron.*, vol. 10, pp. 421–427, Mar.-Apr. 2004.

[12] S. Kumar and D. Yang, "Second-order theory for self-phase modulation and cross-phase modulation in optical fibers," *J. Lightwave Technol.*, vol. 23, pp. 2073–2080, June 2005.

[13] P. Serena, A. Bononi, J.-C. Antona, and S. Bigo, "Parametric gain in the strongly nonlinear regime and its impact on 10-Gb/s NRZ systems with forward-error correction," *J. Lightwave Technol.*, vol. 23, pp. 2352–2363, Aug. 2005.

[14] M. Secondini, E. Forestieri, and C. R. Menyuk, "A combined regular-logarithmic perturbation method for signal-noise interaction in amplified optical systems," *J. Lightwave Technol.*, vol. 27, pp. 3358–3369, Aug. 2009.

[15] M. Winter, C. A. Bunge, D. Setti, and K. Petermann, "A statistical treatment of cross-polarization modulation in DWDM systems," *J. Lightwave Technol.*, vol. 27, pp. 3739–3751, Sept. 2009.

[16] P. Poggiolini, "The GN model of non-linear propagation in uncompensated coherent optical systems," *J. Lightwave Technol.*, vol. 30, pp. 3857–3879, Dec. 2012.

[17] A. Mecozzi and R.-J. Essiambre, "Nonlinear Shannon limit in pseudolinear coherent systems," *J. Lightw. Technol.*, vol. 30, pp. 2011–2024, June 2012.

[18] M. Secondini and E. Forestieri, "Analytical fiber-optic channel model in the presence of cross-phase modulation," *IEEE Photon. Technol. Lett.*, vol. 24, pp. 2016–2019, Nov. 2012.

[19] L. Beygi, E. Agrell, P. Johannisson, M. Karlsson, and H. Wymeersch, "A discrete-time model for uncompensated single-channel fiber-optical links," *IEEE Trans. Commun.*, vol. 60, pp. 3440–3450, Nov. 2012.

[20] M. Secondini, E. Forestieri, and G. Prati, "Achievable information rate in nonlinear WDM fiber-optic systems with arbitrary modulation formats and dispersion maps," *J. Lightw. Technol.*, vol. 31, pp. 3839–3852, Dec. 2013.

[21] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "Properties of nonlinear noise in long, dispersion-uncompensated fiber links," *Opt. Exp.*, vol. 21, pp. 25685–25699, Nov. 2013.

[22] 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. Lightwave Technol.*, vol. 32, pp. 694–721, Feb. 2014.

[23] A. Carena, G. Bosco, V. Curri, Y. Jiang, P. Poggiolini, and F. Forghieri, "EGN model of non-linear fiber propagation," *Opt. Exp.*, vol. 22, pp. 16335–16362, June 2014.

[24] E. Agrell, "Capacity bounds in optical communications," in *Proc. Eur. Conf. Opt. Commun. (ECOC)*, Sept. 2017.

[25] E. Agrell, M. Karlsson, A. R. Chraplyvy, D. J. Richardson, P. M. Krummrich, P. Winzer, K. Roberts, J. K. Fischer, S. J. Savory, B. J. Eggleton, M. Secondini, F. R. Kschischang, A. Lord, J. Prat, I. Tomkos, J. E. Bowers, S. Srinivasan, M. Brandt-Pearce, and N. Gisin, "Roadmap of optical communications," *J. Optics*, vol. 18, no. 6, p. 063002, 2016.

[26] E. Agrell, A. Alvarado, and F. R. Kschischang, "Implications of information theory in optical fibre communications," *Phil. Trans. Roy. Soc. A*, vol. 374, no. 2062, 2016. Art. no. 20140438.

[27] E. Agrell and M. Karlsson, "Influence of behavioral models on multiuser channel capacity," *J. Lightwave Technol.*, vol. 33, pp. 3507–3515, Sept. 2015.

[28] M. Secondini and E. Forestieri, "Scope and limitations of the nonlinear Shannon limit," *J. Lightw. Technol.*, vol. 35, pp. 893–902, Feb. 2017.

[29] R. Dar, M. Shtaif, and M. Feder, "New bounds on the capacity of the nonlinear fiber-optic channel," *Opt. Lett.*, vol. 39, pp. 398–401, Jan 2014.

[30] M. Secondini and E. Forestieri, "On XPM mitigation in WDM fiber-optic systems," *IEEE Photon. Technol. Lett.*, vol. 26, no. 22, pp. 2252–2255, 2014.

[31] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "Pulse collision picture of inter-channel nonlinear interference in fiber-optic communications," *J. Lightw. Technol.*, vol. 34, pp. 593–607, Jan. 2016.

[32] R. Dar, M. Feder, M. Shtaif, P. Winzer, *et al.*, "Inter-channel nonlinear interference noise in fully loaded WDM systems," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, pp. W3I–1, Mar. 2016.

[33] P. Bello, "Characterization of randomly time-variant linear channels," *IEEE Transactions on Communications Systems*, vol. 11, pp. 360–393, Dec. 1963.

[34] K. Liu, T. Kadous, and A. M. Sayeed, "Orthogonal time-frequency signaling over doubly dispersive channels," *IEEE Trans. Inform. Theory*, vol. 50, no. 11, pp. 2583–2603, 2004.

[35] C. B. Czegledi, M. Karlsson, E. Agrell, and P. Johannisson, "Polarization drift channel model for coherent fibre-optic systems," *Scientific Reports*, vol. 6, 2016. Art. no. 21217.

[36] J. Dauwels and H.-A. Loeliger, "Computation of information rates by particle methods," *IEEE Trans. Inf. Theory*, vol. 54, pp. 406–409, Jan. 2008.

[37] D. Marsella, M. Secondini, E. Agrell, and E. Forestieri, "A simple strategy for mitigating XPM in nonlinear WDM optical systems," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, 2015.

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

[39] F. Guiomar, A. Carena, G. Bosco, A. Nespola, L. Bertignono, and P. Poggiolini, "Effectiveness of symbol-rate optimization with PM-16QAM subcarriers in WDM transmission," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, p. W3J.3, 2017.

[40] P. K. A. Wai, C. R. Menyuk, and H. H. Chen, "Stability of solitons in randomly varying birefringent fibers," *Opt. Lett.*, vol. 16, pp. 1231–1233, Aug. 1991.

[41] G. Kramer, M. I. Yousefi, and F. R. Kschischang, "Upper bound on the capacity of a cascade of nonlinear and noisy channels," in *IEEE Information Theory Workshop (ITW)*, Apr. 2015.

[42] M. Karlsson, "Modulational instability in lossy optical fibers," *J. Opt. Soc. Am. B*, vol. 12, pp. 2071–2077, Nov. 1995.

[43] A. Carena, V. Curri, R. Gaudino, P. Poggiolini, and S. Benedetto, "New analytical results on fiber parametric gain and its effects on ASE noise," *IEEE Photon. Technol. Lett.*, vol. 9, pp. 535–537, Apr. 1997.

[44] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "Time varying ISI model for nonlinear interference noise," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, Mar. 2014.

[45] P. Poggiolini, A. Nespola, Y. Jiang, G. Bosco, A. Carena, L. Bertignono, S. M. Bilal, S. Abrate, and F. Forghieri, "Analytical and experimental results on system maximum reach increase through symbol rate optimization," *J. Lightw. Technol.*, vol. 34, pp. 1872–1885, Apr. 2016.

[46] P. P. Mitra and J. B. Stark, "Nonlinear limits to the information capacity of optical fiber communications," *Nature*, vol. 411, pp. 1027–1030, June 2001.

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

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

[49] D. Lavery, R. Maher, G. Liga, D. Semrau, L. Galdino, and P. Bayvel, "On the bandwidth dependent performance of split transmitter-receiver optical fiber nonlinearity compensation," *Opt. Exp.*, vol. 25, pp. 4554–4563, Feb. 2017.

[50] L. B. Du and A. J. Lowery, "Improved single channel backpropagation for intra-channel fiber nonlinearity compensation in long-haul optical communication systems," *Opt. Exp.*, vol. 18, pp. 17075–17088, June 2010.

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

[52] C. B. Czegledi, G. Liga, D. Lavery, M. Karlsson, E. Agrell, S. J. Savory, and P. Bayvel, "Digital backpropagation accounting for polarization-mode dispersion," *Opt. Exp.*, vol. 25, pp. 1903–1915, Feb. 2017.

[53] G. Liga, C. B. Czegledi, and P. Bayvel, "A PMD-adaptive DBP receiver based on SNR optimization," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, 2018.

[54] N. V. Irukulapati, H. Wymeersch, P. Johannisson, and E. Agrell, "Stochastic digital backpropagation," *IEEE Trans. Commun.*, vol. 62, pp. 3956–3968, Nov. 2014.

[55] N. V. Irukulapati, M. Secondini, E. Agrell, P. Johannisson, and H. Wymeersch, "Improved lower bounds on mutual information accounting for nonlinear signal–noise interaction," *J. Lightw. Technol.*, to appear, 2018.

[56] D. Marsella, M. Secondini, and E. Forestieri, "Maximum likelihood sequence detection for mitigating nonlinear effects," *J. Lightwave Technol.*, vol. 32, pp. 908–916, Mar. 2014.

[57] P. Serena, A. Ghazisaeidi, and A. Bononi, "A new fast and blind cross-polarization modulation digital compensator," in *Proc. Eur. Conf. Opt. Commun. (ECOC)*, Sept. 2012.

[58] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "Inter-channel nonlinear interference noise in WDM systems: modeling and mitigation," *J. Lightw. Technol.*, vol. 33, pp. 1044–1053, Mar. 2015.

[59] M. P. Yankov, T. Fehenberger, L. Barletta, and N. Hanik, "Low-complexity tracking of laser and nonlinear phase noise in WDM optical fiber systems," *J. Lightw. Technol.*, vol. 33, pp. 4975–4984, Dec. 2015.

[60] K. Kojima, D. S. Millar, T. Koike-Akino, and K. Parsons, "Constant modulus 4D optimized constellation alternative for DP-8QAM," in *Proc. Eur. Conf. Opt. Commun. (ECOC)*, Sept. 2014.

[61] L. Beygi, E. Agrell, J. M. Kahn, and M. Karlsson, "Coded modulation for fiber-optic networks: Toward better tradeoff between signal processing complexity and optical transparent reach," *IEEE Signal Process. Mag.*, vol. 31, pp. 93–103, Mar. 2014.

[62] M. P. Yankov, D. Zibar, K. J. Larsen, L. P. Christensen, and S. Forchhammer, "Constellation shaping for fiber-optic channels with QAM and high spectral efficiency," *IEEE Photon. Technol. Lett.*, vol. 26, pp. 2407–2410, Dec. 2014.

[63] T. Fehenberger, A. Alvarado, G. Böcherer, and N. Hanik, "On probabilistic shaping of quadrature amplitude modulation for the nonlinear fiber channel," *J. Lightw. Technol.*, vol. 34, pp. 5063–5073, Nov. 2016.

[64] O. Geller, R. Dar, M. Feder, and M. Shtaif, "A shaping algorithm for mitigating inter-channel nonlinear phase-noise in nonlinear fiber systems," *J. Lightw. Technol.*, vol. 34, pp. 3884–3889, Aug. 2016.

[65] X. Liu, A. Chraplyvy, P. Winzer, R. Tkach, and S. Chandrasekhar, "Phase-conjugated twin waves for communication beyond the Kerr nonlinearity limit," *Nature Photonics*, vol. 7, pp. 560–568, July 2013.

[66] H. Eliasson, P. Johannisson, M. Karlsson, and P. A. Andrekson, "Mitigation of nonlinearities using conjugate data repetition," *Opt. Exp.*, vol. 23, pp. 2392–2402, Feb. 2015.

[67] C. Behrens, R. I. Killey, S. J. Savory, M. Chen, and P. Bayvel, "Nonlinear transmission performance of higher-order modulation formats," *IEEE Photon. Technol. Lett.*, vol. 23, pp. 377–379, Mar. 2011.

[68] L. B. Du and A. J. Lowery, "Optimizing the subcarrier granularity of coherent optical communications systems," *Opt. Exp.*, vol. 19, pp. 8079–8084, Apr. 2011.

[69] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423/623–656, July/Oct. 1948.

[70] K. S. Turitsyn, S. A. Derevyanko, I. V. Yurkevich, and S. K. Turitsyn, "Information capacity of optical fiber channels with zero average dispersion," *Phys. Rev. Lett.*, vol. 91, Nov. 2003.

[71] M. I. Yousefi and F. R. Kschischang, "On the per-sample capacity of nondispersive optical fibers," *IEEE Trans. Inf. Theory*, vol. 57, pp. 7522–7541, Nov. 2011.

[72] K. Keykhosravi, G. Durisi, and E. Agrell, "A tighter upper bound on the capacity of the nondispersive optical fiber channel," in *Proc. Eur. Conf. Opt. Commun. (ECOC)*, Sept. 2017.

[73] D. M. Arnold, H.-A. Loeliger, P. O. Vontobel, A. Kavčić, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Trans. Inform. Theory*, vol. 52, pp. 3498–3508, Aug. 2006.

[74] T. Fehenberger, A. Alvarado, P. Bayvel, and N. Hanik, "On achievable rates for long-haul fiber-optic communications," *Opt. Exp.*, vol. 23, pp. 9183–9191, Apr. 2015.

[75] T. A. Eriksson, T. Fehenberger, P. A. Andrekson, M. Karlsson, N. Hanik, and E. Agrell, "Impact of 4D channel distribution on the achievable rates in coherent optical communication experiments," *J. Lightw. Technol.*, vol. 34, pp. 2256–2266, May 2016.

[76] G. Liga, A. Alvarado, E. Agrell, and P. Bayvel, "Information rates of next-generation long-haul optical fiber systems using coded modulation," *J. Lightw. Technol.*, vol. 35, pp. 113–123, Jan. 2017.

[77] L. G. L. Wegener, M. L. Povinelli, A. G. Green, P. P. Mitra, J. B. Stark, and P. B. Littlewood, "The effect of propagation nonlinearities on the information capacity of WDM optical fiber systems: Cross-phase modulation and four-wave mixing," *Physica D: Nonlinear Phenomena*, vol. 189, pp. 81–99, Feb. 2004.

[78] I. Djordjevic, B. Vasic, M. Ivkovic, and I. Gabitov, "Achievable information rates for high-speed long-haul optical transmission," *J. Lightwave Technol.*, vol. 23, pp. 3755–3763, Nov. 2005.

[79] R.-J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity limits of optical fiber networks," *J. Lightwave Technol.*, vol. 28, pp. 662–701, Feb. 2010.

[80] D. Slepian, "On bandwidth," *Proc. IEEE*, vol. 64, pp. 292–300, 1976.

[81] J. Gordon and H. Kogelnik, "PMD fundamentals: Polarization mode dispersion in optical fibers," *Proceedings of the National Academy of Sciences*, vol. 97, no. 9, pp. 4541–4550, 2000.

[82] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ: Wiley, 2nd ed., 2006.

[83] G. Colavolpe, A. Barbieri, and G. Caire, "Algorithms for iterative decoding in the presence of strong phase noise," *IEEE J. Sel. Areas Commun.*, vol. 23, pp. 1748–1757, Sept. 2005.