Fiber Nonlinearity and Optical System Performance

# Fiber Nonlinearity and Optical System Performance

Alberto Bononi, Ronen Dar, Marco Secondini, Paolo Serena and Pierluigi Poggiolini

***Abstract*—This chapter aims at providing a comprehensive picture of the impact of fiber non-linear effects on modern coherent WDM systems performance. First, the main non-linearity models currently available are introduced and discussed in depth. Then, various specific aspects are addressed, such as the interplay of PMD/PDL and non-linearity, or the dependence of non-linear effects on modulation format. The important topic of non-linear effects mitigation is then dealt with. Finally, system performance metrics and capacity are discussed extensively, as to they are fundamentally influenced and limited by fiber non-linearity.**

***Index Terms*—coherent systems, uncompensated transmission, fiber non-linearity, perturbation models, GN-model, EGN-model, time-domain model, spatially-resolved model, LP-model, non-linearity mitigation, performance metrics, optical system capacity, non-linear Shannon limit, LOGON strategy, achievable information rate.**

OPTICAL propagation within an optical fiber involves complex linear and nonlinear interactions among traveling optical signals which manifest themselves as intersymbol as well as inter-channel interference among the digital data. To shed light on such effects, a large amount of analytical, numerical and experimental investigations have been carried out. Scientific modeling is an essential tool for any investigation as a way to make it easier to understand and reproduce experimentally observed results. Models help system designers in predicting the behavior of an optical network, and show their greatest effectiveness in their ability to isolate subsets of test conditions not easily obtainable otherwise.

The main aim of this chapter is to review the most important models for performance prediction in optical fibers, to point out their strength and weaknesses and provide both theoretical background and practical indications on their use.

## I. Modeling the optical fiber

Any model captures part of reality at a given abstraction level, hence each model can always be generalized into a more complete model able to explain the previous one, at the cost of an increased complexity. Usually in science the most complete models are local models, i.e., general descriptions of what happens at each space coordinate. Such models are usually written in terms of differential equations. In many cases such equations cannot be solved exactly, nevertheless they contain all the information about the evolution of the signal of interest. In optical fiber propagation the reference model is the nonlinear Schrödinger equation (NLSE), which is a partial differential equation that describes the local behavior of a propagating electromagnetic wave [1]. In optical communications, an electromagnetic wave is typically written as $A(z,t)\cdot\exp(j(\omega t-\beta z))$. Such notation describes a wave of amplitude $|A|$, phase $\arg[A]$, propagating at frequency $f=\omega/(2\pi)$ at phase velocity $\omega/\beta$ in the $z$ direction. $A(z,t)$ is assumed to be 'slowly-varying', in both time $t$ and distance $z$, with respect to the faster-oscillating exponential factor. Henceforth, for compactness, we will omit to explicitly indicate the $t$ and $z$ dependence of $A$.

In its simplest form the NLSE can be written as [1]:

$$\frac{\partial A}{\partial z}=-\frac{\alpha_{\rm p}}{2}A+j\frac{\beta_2}{2}\frac{\partial^2 A}{\partial t^2}-j\gamma|A|^2A\ . \tag{1}$$

The first term on the right side in (1) causes the wave to experience a *power loss* at a rate set by $\alpha_{\rm p}$. The small value of $\alpha_{\rm P}$ in silica compared to copper cable has been one of the greatest achievements of optical communications, celebrated by the Nobel prize to the pioneering studies made by sir Charles Kao.

The second term in (1) expresses dispersion in the time domain $t$. To understand the implications of dispersion we should first generalize the basic idea of plane wave. The reader familiar with the Fourier transform should not be surprised that the true electromagnetic field propagating within an optical fiber can be described by a superposition of waves. However, such waves travel at different speeds with a group velocity dispersion (GVD) proportional to $\beta_2$. This is the implication of the second term on the right side of (1), which is usually best understood in the frequency domain as a term in quadrature with the signal and proportional to the square of the frequency. As a result, GVD alone on a global scale manifests as a pure phase shift in frequency domain. GVD makes the different waves making up the signal reach the receiver at different times, so that their sum is different than what it was at the transmitter. The result is signal distortion, generally referred to as inter-symbol interference at the receiver. This problem was eventually solved by means of optical or electronic dispersion compensation.

The third term in (1) is a nonlinear term, i.e., a contribution for which the superposition principle does not hold anymore. It is called the nonlinear Kerr effect [1], [2]. As a consequence, because of nonlinearity the traveling waves interact. We observe that nonlinearity is weighed by a factor $|A|^2$, hence proportional to the *power* of the propagating signal. We thus expect the nonlinear effect to be relevant only at high power. This observation is reminiscent of the idea of Taylor series in mathematics, where any well-behaved function is linear in a small neighborhood of a given value. In this case, we could say that the fiber is linear close to the zero value of the amplitude $A$. It is worth noting, though, that from a naive similarity with Taylor series, one would expect a nonlinear quadratic term, whereas the nonlinear Kerr term in (1) is *cubic* in $A$. The reason of the cubic behavior is strictly related to the odd symmetry of the silica molecule [1].

The nonlinear Kerr term in (1) is a memoryless effect since it impacts the propagating signal at the present-time

only. This is clearly unrealistic, since any physical effect has some memory. However, this assumption is justified by the extremely small time-lag over which nonlinearity occurs, on the order of femtoseconds.

There is a very interesting duality between nonlinear Kerr effect in time domain and dispersion in frequency domain. For instance, GVD alone does not make different frequencies interact, as for any linear time invariant effect, while it creates inter-symbol interference in the time domain; the opposite occurs for nonlinearity: different times do not interact, while different frequencies do. Moreover, while GVD acts in quadrature with the signal in the frequency domain, the same occurs for the nonlinear Kerr effect, but in the time domain. Hence, for similar arguments, both effects manifest as phase rotations: GVD in the frequency domain, the nonlinear Kerr effect in time domain. The latter is generally called self phase modulation (SPM).

Modeling nonlinearity has been one of the greatest challenges of the last decades because of the complexity of the problem. Thanks to the similarity between optical waves propagation in optical fibers and water waves propagation in the sea, it is helpful to visualize nonlinearity by the waves interaction in a heavy sea. The chaotic behavior of such interaction is self-explaining of the difficulty of undoing it at the receiver as well of the difficulty of a full statistical description.

Linear and nonlinear effects are experienced both by the information signal but also by the unavoidably propagating noise, generally introduced by optical amplifiers in the form of amplified spontaneous emission (ASE).

To understand the implications of ASE let us ignore nonlinear effects for a moment. At the receiver we know how to compensate for dispersion, thus we are left with the problem of noise cumulated along the link. The solution appears simple: increase power to make noise be negligible with respect to the information signal. Figure 1 sketches this idea by showing the typical behavior of the receiver signal to noise ratio (SNR) versus signal power, at the output of a fiber transmission link. The SNR is a typical quality-of-transmission parameter, as we will discuss more in detail later in this chapter. In the absence of nonlinear effects, the SNR would grow indefinitely for increasing power, see the dashed line “linear asymptote” in Fig. 1. However, this ceases to hold when nonlinearity enters the game. The typical behavior of SNR versus power is depicted in Fig 1 by a solid line. From the figure we observe that an ‘optimal power’ exists, corresponding to the maximum achievable SNR. Beyond the optimal power the SNR drops, with a slope which is actually steeper than the slope of increase at small power [43], [119].

We thus recognize two asymptotic behaviors at small/high powers, which in agreement with the previous discussion define what we call the ASE-limited region and the nonlinearity-limited region, respectively.

The optimal power of Fig. 1 refers to a link of a given length. We may ask what happens by increasing the link length. In this scenario both the linear and the nonlinear asymptotes run lower, because of more accumulated ASE and nonlinearity along transmission. If we fix a reference SNR,

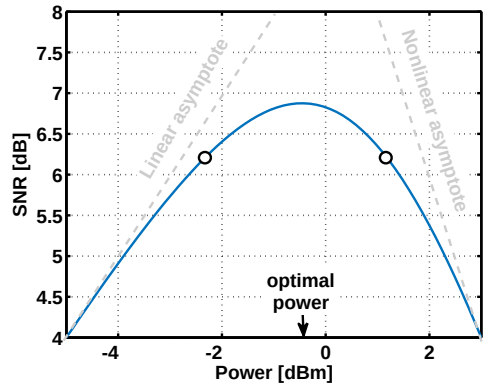


Fig. 1. Example of behavior SNR vs. Power after a fiber-optic link.

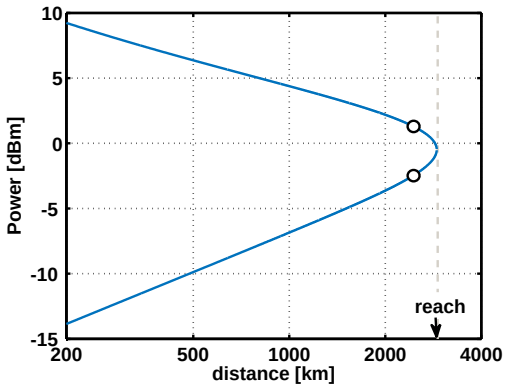


Fig. 2. Example of contour levels at fixed SNR. The lower part is set by accumulation of ASE; the higher part is set by accumulation of nonlinearity. The reach is the maximum distance for the given SNR.

which may be the threshold of the forward error correcting code under use, we can identify two power values, one in the ASE dominated regime and one in the nonlinearity dominated regime. Such values are sketched in Fig. 1 at a fixed distance, and in Fig. 2 at all distances up to the maximum distance, or *maximum reach* (MR), at which they merge into a single power value.

In this chapter we discuss some models that are able to describe with good approximation such important features as the optimal power and the reach. Of course their knowledge is of great importance in the study, design and even real-time management of optical networks, a topic that will be covered in Section X-C.

The NLSE can be further generalized. So far we have described the electromagnetic wave in terms of its amplitude, frequency and propagation constant. However, if this wave collides with an electron along propagation, such an electron starts oscillating around a preferred direction. The information about this direction has not been accounted for, so far. We thus need to accept that our signal wave is described by a *vector*. When such a direction matters for investigating propagation, we say that the optical fiber shows *birefringence*. Such vectors can be fully described by two components. The scalar impairments such as loss and dispersion become matrices in the presence of birefringence.

The implications of the vector nature of of the waves will be described more in detail in Section IX-C. In essence, the nonlinear term of the NLSE becomes more complex [2], [15], [16]. However, it is possible to simplify it with reasonable approximations, based on the following argument.

The two components of the wave, from now on referred to as *polarizations*, experience different propagation constants. Such a difference is the main manifestation of fiber birefringence. Because of it, the two polarizations periodically phase-shift and return in phase after the so called *beat length* [93]. For typical telecom fibers, the beat length is on the order of meters, a length scale usually much smaller than the length scale over which nonlinearity induces significant changes in the wave. We are thus justified to average out such fast phase variations.

The result yields an isotropic nonlinear effect, that turns the scalar nonlinear term $\gamma |A|^2 A$ into $\gamma \frac{8}{9} ||\vec{A}||^2 \vec{A}$, where $\vec{A} = [A_x,\ A_y]$ is the vector containing the amplitudes of the two polarizations in the fiber, while $||\vec{A}||^2$ is the overall power of the wave. The factor $8/9$ is the net result of the averaging. The corresponding equation is referred to as the *Manakov equation* [19], [99].

The Manakov equation is currently the most widely used signal-propagation equation, the starting point for the derivation of most analytical models quantifying the system impact of nonlinearity. This chapter will mostly focus on such analytical models. However, another approach is also possible towards system study and design.

It is based on the brute-force numerical integration of the signal propagation equation, be it the NLSE, the Manakov equation or more complex ones. While analytical modeling has recently gained widespread adoption, numerical integration is and will likely be a mainstay of system studies. One of the main reasons is that, as mentioned, all analytical models are based on assumptions and approximations, and must be validated. System simulation, by means of direct numerical integration of the propagation equations and Monte-Carlo system performance estimation, provides the indispensable tool for model validation. In addition, non-conventional system set-ups may fall outside the scope of models and necessarily need numerical integration. Because of the key role of system simulations, the next section is devoted to this topic.

## II. Fiber propagation: numerical methods

The NLSE (1) introduced in Section I has no general closed-form solution. While some approximated analytical solutions (discussed in Section III-A) are available and widely used for modeling purposes, the accurate simulation of optical fiber systems largely relies on numerical methods.

Over the years, several different numerical methods have been developed to solve the NLSE and its generalizations, based on different approaches for the discretization of the time and space variables, such as classical explicit and implicit finite-difference methods and pseudospectral methods (see, for instance, [176], [1] and references therein). Among them, the split-step Fourier method (SSFM) [174], [175] has largely prevailed thanks to its simplicity, flexibility (it can be easily extended to include other propagation effects and systems

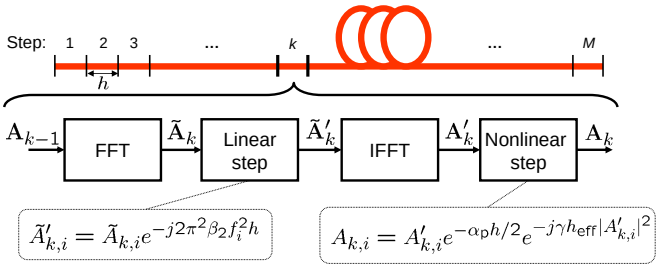


Fig. 3. Basic implementation of the SSFM algorithm.

elements), numerical stability, and computational efficiency [176], becoming the reference method in the optical fiber community [1].

The key idea behind the SSFM is that of dividing the propagation along a link of length $L$ into $M$ steps of length $h = L/M$, over which the effects of the linear and nonlinear parts of the equation are split (hence, the attribute *split step*). Equation (1) is rewritten in the simple general form

$$\frac{\partial A}{\partial z} = (\mathcal{L} + \mathcal{N})A \tag{2}$$

where the operators $\mathcal{L} = j(\beta_2/2)(\partial^2/\partial t^2)$ and $\mathcal{N} = -\alpha_p/2 - j\gamma |A|^2$ represent the linear and nonlinear part of the equation, respectively [180]. (Though fiber attenuation is a linear effect and should naturally be included in the linear operator $\mathcal{L}$, moving it into the nonlinear operator $\mathcal{N}$ makes the resulting SSFM more accurate [180].)

The propagation over a step of length $h$ can be formally approximated as

$$A(z+h,t) \simeq \exp\left[\int_z^{z+h} \mathcal{N}(\zeta) d\zeta\right] \exp(h\mathcal{L}) A(z,t) \tag{3}$$

The approximation (3) is particularly convenient, as the separate effects of the linear and nonlinear operators can be easily expressed in closed form, the former in the frequency domain (hence the attribute *Fourier*), and the latter in the time domain.

The algorithm operates on a discrete-time representation of the propagating waveform $A(z,t)$. Assuming a sampling time $T_s$ and a total duration $T = NT_s$, the signal at distance $z = kh$ (i.e., after after $k$ propagation steps) is represented by the vector $\mathbf{A}_k = (A_{k,1}, \ldots, A_{k,N})$, with $A_{k,i} = A(kh, iT_s)$. Transformations from time to frequency domain and back are based on the direct and inverse discrete Fourier transforms, which are efficiently computed by the fast Fourier transform (FFT) algorithm. This entails assuming periodic boundary conditions both in time and frequency, with periods $T$ and $B = 1/T_s$, respectively. Therefore, when selecting $T$ and $B$, special care must be taken to ensure that they are sufficiently wider than the actual continuous-time signal duration and (two-sided) bandwidth, respectively, to avoid time and frequency aliasing [179]. In the basic implementation of the algorithm, the input vector $\mathbf{A}_0$ is propagated through the $M$ fiber sections according to the scheme in Fig. 3. In particular, the propagation through the $k$-th section of the fiber consists of the following four operations:

1) Computation of the FFT of $\mathbf{A}_{k-1}$ to obtain the frequency-domain vector $\tilde{\mathbf{A}}_k = (\tilde{A}_{k,1}, \ldots, \tilde{A}_{k,N})$;
2) linear step to compute the vector $\tilde{\mathbf{A}}'_k = (\tilde{A}'_{k,1}, \ldots, \tilde{A}'_{k,N})$ according to

$$\tilde{A}'_{k,i} = \tilde{A}_{k,i} e^{-j2\pi^2 \beta_2 f_i^2 h}, \quad i = 1, \ldots, N \tag{4}$$

where $f_i$ is the frequency of the $i$-th component;
3) computation of the inverse FFT (IFFT) of $\tilde{\mathbf{A}}'_k$ to obtain the time-domain vector $\mathbf{A}'_k = (A'_{k,1}, \ldots, A'_{k,N})$;
4) nonlinear step to compute the vector $\mathbf{A}_k = (A_{k,1}, \ldots, A_{k,N})$ according to

$$A_{k,i} = A'_{k,i} e^{-\alpha_\mathrm{p} h/2} e^{-j\gamma h_\mathrm{eff} |A'_{k,i}|^2}, \quad i = 1, \ldots, N \tag{5}$$

where $h_\mathrm{eff} = (1 - e^{-\alpha_\mathrm{p} h})/\alpha_\mathrm{p}$ is the effective length of the step, which accounts for the dependence of the nonlinear operator on the propagation variable due to the attenuation.

The algorithm complexity is $O(N \log N)$ for the FFTs, and $O(N)$ for (4) and (5). The actual number of operations depends on the specific implementation of the algorithm and will be shortly discussed in Section XI-A. The algorithm can be extended to the Manakov equation for dual-polarization signals by replacing signal samples with two-element vector samples (each element representing a polarization component). All the described operations should then be applied to both polarizations, interpreting the squared modulus $|A'_{k,i}|^2$ in (5) as the squared norm $\|A'_{k,i}\|^2$ of the vector sample.

Since the linear and nonlinear operators in (2) do not commute, (3) is only an approximation. However, an argument based on the Baker–Campbell–Hausdorff formula ensures that the local error induced by (3) is of the order of $h^2$, such that the global error accumulated at the output of the link is of the order of $h$, i.e., it decreases linearly with the number of steps $M$. A more accurate approximation can be obtained by using (3) for half the step, and applying the operators in reverse order for the other half, such that error terms of the order of $h^2$ cancel out. This results in a symmetric version of the algorithm that is based on the approximation

$$A(z+h) \simeq \exp(\frac{h}{2}\mathcal{L}) \exp\left[\int_z^{z+h} \mathcal{N}(\zeta) d\zeta\right] \exp(\frac{h}{2}\mathcal{L}) A(z) \tag{6}$$

and has local and global errors of the order of $h^3$ and $h^2$, respectively. The adjacent linear steps of two consecutive sections can be merged into a single linear step of length $h$. Therefore, the symmetric algorithm is almost equal to the basic one depicted in Fig. 3—but for halving the length of the first linear step and for adding one more linear step of length $h/2$ after the last section—and achieves a higher accuracy with a negligible increase of complexity.

The accuracy and complexity of the algorithm can be traded off by selecting the step size $h$, which can also be adaptively changed along the link to increase the accuracy for a fixed complexity. Some possible criteria for the selection of the step size, as well as a strategy for its adaptation based on the local error, are discussed and compared in [177]. A simple strategy for the adaptation of the step size based on the global error is presented in [178].

When processing a very long sequence of samples, the overlap-and-save technique is typically employed to minimize the complexity of the algorithm and keep its latency within reasonable limits [179]. The input sequence of samples is thus divided into several overlapping blocks of length $N$, which are separately processed according to the algorithm in Fig. 3. The output sequence is eventually reconstructed by concatenating the output blocks and discarding the overlapping samples.

The number of overlapping samples should be at least equal to the overall memory $N_m$ of the fiber-optic channel, which depends on the maximum dispersion experienced during propagation. In a dispersion-uncompensated link of total length $L$, assuming a constant dispersion parameter $\beta_2$, the channel memory can be approximated as $N_m \simeq 2\pi |\beta_2| B^2 L$ [78].

## III. Fiber propagation: analytical perturbation modeling

### A. Perturbation methods

The absence of closed-form solutions of the NLSE, except for peculiar cases, stimulated the search for approximated analytical solutions. Among the plethora of methods that mathematics can provide, perturbation theory [115], [116] has been the most widely used in approximating the NLSE.

The basic idea of perturbation theory is to identify a term in a differential equation that is expected to be a small disturbance to the exact solution of the equation, i.e., it is a perturbation. Most of the times the small perturbative term is weighed by a parameter, which represents a key variable for perturbation theory. If such a parameter is not clearly visible, one can always multiply the perturbative term by a coefficient $\epsilon$, apply the perturbation theory, and finally set $\epsilon = 1$.

Once the small perturbation and its related parameter $\epsilon$, which is based on physical intuition, are identified, one can use math to properly approximate the solution for small $\epsilon$. The idea is to write the unknown exact solution of the differential equation as a Taylor series in $\epsilon$. In the NLSE case, we identify the nonlinear Kerr effect as the small perturbation, and $\gamma$ as the corresponding small parameter [23], [117], [24]. We thus search the solution of the NLSE as a Taylor series with respect to $\gamma$:

$$A(z,t) \simeq A_0(z,t) + A_1(z,t)\gamma + \ldots$$

We do not know $A(z,t)$, but thanks to perturbation theory we can evaluate $A_n(z,t)$ for all integers $n \geq 1$. They can be evaluated recursively by solving the following differential equations [115]:

$$\frac{\partial A_n}{\partial z} = \mathcal{L} A_n + \mathcal{N}(A_{n-1}) \tag{7}$$

By a closer look we note that such differential equations are linear, with a forcing term $\mathcal{N}(A_{n-1})$. Each of them can be thus solved exactly once we know the solution at the previous order.

Although we can truncate the Taylor series up to any order, all $A_n$ with $n \geq 2$ are extremely complex and, with very few exceptions, the NLSE has typically been approximated

to first order in $\gamma$. Incidentally, the NLSE is a well behaved differential equation and setting $\gamma = 0$ does not cause divergence problems, so that we say that we have a regular perturbation (RP) solution [115]. We are then mainly interested in RP1, i.e., in RP up to $\gamma^1$.

The first order perturbative term in frequency domain takes the following general expression [116], [23], [22], [34]:

$$\tilde{A}_1(z,f) = \iint_{-\infty}^{\infty} \eta(f_1,f_2,f) \times \tilde{A}(0,f+f_1)\tilde{A}(0,f+f_2)\tilde{A}^*(0,f+f_1+f_2)\mathrm{d}f_1\mathrm{d}f_2 \tag{8}$$

where $\eta(\omega_1,\omega_2,\omega)$ is the fiber kernel and contains all the main properties of the link. From a numerical point of view the RP solution (8) does not yield significant speed up with respect to the numerical solution provided by SSFM. However, its main interest is in providing an end-to-end analytical solution that can be statistically manipulated, as we will see in the following sections. Alternatively, the RP1 solution can be expressed as a propagating perturbative term along distance, whose implications will be discussed in Section VI.

For the NLSE, the RP has been showed to coincide with the Volterra series expansion [22]. Nevertheless, the physical intuition behind RP is of great help in suggesting ways to improve the solution. First, it has been shown that even in the presence of dispersion part of the perturbation is in quadrature with the unperturbed signal. Therefore, besides observing amplitude distortion at the receiver, we do have phase distortion that we generally categorize as SPM and XPM. Even if SPM and XPM are small phase distortions, they may induce a significant low-frequency contribution that is usually removed by the carrier phase estimator (CPE) in coherent detection, or by the photodiode in direct detection. The basic RP solution (8) is 'unaware' of such observations, and thus its basic expression (8) often fails to be satisfactorily accurate. Special tricks to remove such problems are thus necessary to improve the range of validity of the RP1 approximation.

A first improvement is to search the perturbation in the reference system of the *average nonlinear phase rotation*, yielding the enhanced RP (eRP) method [22]. For first order eRP the approximated solution of the NLSE is:

$$A(z,t) \simeq (A_0 + \gamma A_1)\, e^{-j\Phi_{\mathrm{NL}}}$$

where $\Phi_{\mathrm{NL}}$ is the average nonlinear phase cumulated along the link. For instance, for a single span of effective length $L_{\mathrm{eff}}$ and nonlinear length $L_{\mathrm{NL}}$ it is $\Phi_{\mathrm{NL}} = \gamma L_{\mathrm{NL}}/L_{\mathrm{eff}}$. This approach corresponds to perturbing a modified NLSE with average power $P$ removed in the nonlinear term, i.e., with $\mathcal{N}(A) = -j\gamma(|A|^2 - P)A$ [91]. In Sections IV,V,VI we will calculate and discuss the variance of the perturbation in the eRP reference system.

Another improvement to RP is the logarithmic perturbation (LP), where the perturbation is searched in a logarithmic domain [38]. At first order for $A_0 \neq 0$ we have:

$$A(z,t) \simeq A_0 e^{\gamma \frac{A_1}{A_0}} \ .$$

LP coincides with the exact solution of NLSE in the limit of zero dispersive effect, where the nonlinear Kerr effect is a pure phase rotation. With GVD the exponent in LP is a complex number, hence it does not manifest itself just as a phase rotation. This class of models will be discussed in Sect. VII.

## IV. From FWM to the GN and EGN models

In this section, two of the most well-known and widely used nonlinearity models, the GN and EGN-model, will be introduced by deriving them from the basic Kerr-related nonlinear fiber effect called Four-Wave Mixing (FWM). Historically, this is one of the ways in which these models have been derived, and perhaps it is one of the most intuitive. Interestingly, this FWM-based derivation can be shown to be equivalent to the RP1 method described in the previous section.

In the end, although some of the details will have to be skipped, the reader should gain a rather detailed appreciation of the derivation procedure, of the inner structure of these models and a general understanding of how fiber nonlinear effects can be described in terms of the interaction of many discrete signal spectral components with one-another.

The GN-model was actually derived several times, in slightly different versions, over the years. The earliest instance dates back to 1993 [30], where a FWM-based derivation was used. Though limited to single-polarization, ideal-lossless fiber and a rectangular overall WDM spectrum, the equations in [30] essentially coincide with those of the incoherent GN-model, or iGN-model (see Sect. IV-G).

Later, in 2003, it was shown in [35] that results similar to [30] could also be derived using a different perturbation approach, earlier proposed in [25]. Similar equations were also independently derived in [34] using truncated Volterra-series in frequency domain, as introduced in [23].

Subsequently (2008-2010), a derivation approach based again on FWM was used in [40], [42] to derive a GN-model version limited to optical OFDM. A similar approach was independently exploited in 2010-2011 to address not just OFDM, but all WDM systems [44], [48], [49]. The derivation yielded the frequency-continuous, dual-polarization integral GN-model formula which is currently considered the reference formula for the model (called GNRF, or GN-model reference formula). The name 'GN-model' was first introduced in [48].

Detailed re-derivations of the GN-model were published in [52], [58] based on different methods. Both independently confirmed the GN-model equations and provided substantial extensions and insightful generalizations. Various follow-up papers have been published since, providing further extensions as well as closed-form approximate solutions of the integral GN-model formula, for instance [50], [53], [56], [60], [62].

More recently, based on a better understanding of the features and limitations of the GN-model, an improved model called the *enhanced* GN-model (or EGN-model) has been proposed. For references and a detailed discussion of the EGN-model, see Sect. IV-H.

### A. Kerr nonlinearity and spectral-line signals

We consider a signal $E(f)$ which is made up of $Q$ spectral 'lines', ideally Dirac's deltas, written as:

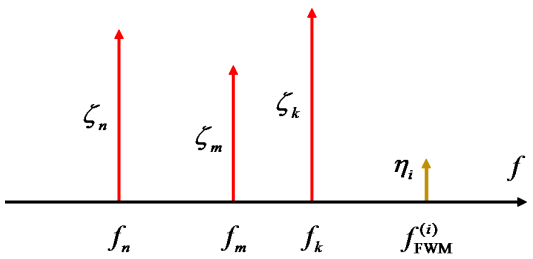


Fig. 4. Generation of a FWM spectral line at frequency $f_{\mathrm{FWM}}^{(i)} = (f_m - f_n + f_k)$ due to Kerr nonlinearity acting upon the signal spectral lines at frequencies $f_m$, $f_n$ and $f_k$, called 'pumps'.

$$E(f) = \sum_{q=1}^{Q} \zeta_q \delta\left(f - f_q\right) \tag{9}$$

where the $\zeta_q$'s are the complex amplitudes of each one of the lines. On such spectrally delta-like signal, fiber Kerr nonlinearity (see Sect. I) acts mainly through the FWM effect.

The literature on FWM is vast. A comprehensive treatment can be found in [1]. In the next few paragraphs, we are going to recall only what is strictly needed for our purposes.

When $E(f)$ is propagated along a stretch of optical fiber, then *each* set of *three* out of the $Q$ signal lines in Eq. (9), say those located at frequencies $f_k$, $f_m$ and $f_n$, generates a FWM contribution in the form of a *fourth* spectral line at frequency:

$$f_{\mathrm{FWM}}^{(i)} = f_k + f_m - f_n \tag{10}$$

as shown in Fig. 4. Note that $k, m, n$ run each independently over $\{1 \ldots Q\}$ and therefore there are $Q^3$ possible triples $(k, m, n)$. We use the index $i$ to identify each one of these triples, through the notation $(k, m, n)_i$, where $i$ runs over $1 \ldots Q^3$. A given triple $(k, m, n)_i$ may actually generate a FWM line at the same frequency as the FWM line generated by another triple $(k', m', n')_{i'}$. That is, it can be that: $f_{\mathrm{FWM}}^{(i)} = f_{\mathrm{FWM}}^{(i')}$. The reason is that it is possible that $(f_k + f_m - f_n) = (f_{k'} + f_{m'} - f_{n'})$, even though $(k, m, n)_i \neq (k', m', n')_{i'}$.

We now write the FWM generation formulas for a single stretch of fiber. A FWM spectral line generated by the triple $(k, m, n)_i$ at a specific frequency $f_{\mathrm{FWM}}^{(i)}$ can be written as:

$$E_i\left(z, f\right) = \eta_i\left(z\right) \delta\left(f - f_{\mathrm{FWM}}^{(i)}\right) \tag{11}$$

Its complex amplitude $\eta_i$ grows along the fiber, depending on several parameters, including the power of the signal spectral lines, their relative spectral position, fiber dispersion, loss and the strength of the Kerr nonlinearity of that fiber.

Significantly, such growth law takes on a relatively simple form, provided that the so-called 'undepleted pump assumption' (UPA) can be made. Such terminology stems from the fact that the signal lines in Eq. (9) are called 'pumps' according to jargon inherited from Physics. The UPA consists of this: due to energy conservation, the FWM-generated spectral lines draw their growing power from the 'pumps', whose power is therefore 'depleted'. However, if the overall power of the generated FWM lines is small vs. the pumps' power, then the pumps can be assumed unperturbed, or *undepleted*. It can be shown that the UPA is equivalent to what is called a 'first-order regular perturbation' (RP1) approach to finding the solution of the NLSE (see Sect. III-A).

We assume that the FWM amplitude value at the start of the fiber, $z = 0$, be equal zero, so that $\eta_i\left(0\right) = 0$. Then, using the single-polarization NLSE, under the UPA, the FWM line amplitude $\eta_i$ can be shown to be:

$$\eta_i\left(z\right) = -j\zeta_m\zeta_n^*\zeta_k \cdot e^{\int_0^z \kappa\left(f_{\mathrm{FWM}}^{(i)}, z'\right) dz'} \cdot \int_0^z \gamma\left(z'\right) e^{\int_0^{z'} \left[-\kappa\left(f_{\mathrm{FWM}}^{(i)}, z''\right) + \kappa\left(f_m, z''\right) + \kappa^*\left(f_n, z''\right) + \kappa\left(f_k, z''\right)\right] dz''} dz' \tag{12}$$

where:

$$\kappa\left(f, z\right) = -j\beta\left(f, z\right) - \alpha\left(f, z\right) + g\left(f, z\right) \tag{13}$$

Note that this equation can handle all-order dispersion through $\beta(f, z)$, as well as frequency-dependent distributed-gain or loss along the fiber through $g(f, z)$ and $\alpha(f, z)$, respectively, $g$ and $\alpha$ being both positive numbers by definition.

The overall FWM created by the signal of Eq. (9) at the location $z$ along the fiber is actually given by the superposition of all the individually generated FWM contributions, that is:

$$E_{\mathrm{FWM}}\left(z, f\right) = \sum_i \eta_i\left(z\right) \delta\left(f - f_{\mathrm{FWM}}^{(i)}\right) \tag{14}$$

Focusing onto the quite typical case of a *homogenous* stretch of fiber, whereby $\gamma$ is a constant vs. the spatial coordinate $z$, this allows one to re-write Eq. (12) in the compact form:

$$\eta_i\left(z\right) = -j\gamma\zeta_k\zeta_m\zeta_n^* \, \mathrm{H}\left(f_{\mathrm{FWM}}^{(i)}, z\right) L_{\mathrm{eff}}\left(z, \bar{f}_i\right) \mu\left(z, \bar{f}_i\right) \tag{15}$$

where $\bar{f}_i = (f_k, f_m, f_n)$ is a shorthand for the $i$-th triple of frequencies from Eq. (4) generating a FWM line at $f_{\mathrm{FWM}}^{(i)}$ as per Eq. (10), and:

$$\mathrm{H}\left(f, z\right) = e^{\int_0^z \kappa\left(f, z'\right) dz'} \tag{16}$$

is the *linear fiber transfer function* at a generic frequency $f$, from the fiber start till the length $z$. In addition:

$$L_{\mathrm{eff}}\left(z, \bar{f}_i\right) = \int_0^z e^{\int_0^{z'} \mathrm{Re}\left\{-\kappa\left(f_{\mathrm{FWM}}^{(i)}, z''\right) + \kappa\left(f_m, z''\right) + \kappa^*\left(f_n, z''\right) + \kappa\left(f_k, z''\right)\right\} dz''} dz' \tag{17}$$

$$\mu\left(z, \bar{f}_i\right) = L_{\mathrm{eff}}^{-1}\left(z, \bar{f}_i\right) \cdot \int_0^z e^{\int_0^{z'} \left[-\kappa\left(f_{\mathrm{FWM}}^{(i)}, z''\right) + \kappa\left(f_m, z''\right) + \kappa^*\left(f_n, z''\right) + \kappa\left(f_k, z''\right)\right] dz''} dz' \tag{18}$$

where $\mathrm{Re}\{\cdot\}$ means 'real part'.

Eq. (15) shows that the amplitude of the $i$-th generated FWM line, $\eta_i$, is proportional to the factors $L_{\mathrm{eff}}$ and $\mu$. The former is called the 'effective length' of the fiber. It is a *real and positive* number, with units of length. It can be thought of as the length of the fiber as 'felt' by FWM generation. In a perfectly lossless fiber it coincides with the actual physical

length of the fiber, that is: $L_{\text{eff}} = z$. In a lossy fiber, such 'FWM length' is shorter than $z$. Note that when *both* loss and distributed gain are present, the result of Eq. (17) can be either shorter or longer than $z$, depending on their balance.

The quantity $\mu$ is the so-called 'FWM efficiency' and it accounts for the critically important effect of dispersion on FWM generation. If there is no dispersion, i.e., if $\beta(f)$ is a constant vs. $f$, then it can be immediately seen from Eq. (18) that $\mu = 1$. If *any* amount of dispersion is instead present, then $|\mu| < 1$. This shows that dispersion is a mitigator of FWM and, in general, of nonlinearity, an important concept that we will return to, later in the chapter.

Eqs. (15)-(18) constitute an analytical solution to the NLSE in the FWM framework, potentially providing a powerful tool for the derivation of nonlinearity models. In practice, for this to be the case, two conditions must be met: (i) that the propagation problem is within the validity range of RP1, and (ii) that the signal of interest, which is sent into the fiber, can be described as a set of spectral lines like Eq. (9), so that the FWM formalism can be applied.

As for (i), this condition is typically well verified in modern fiber telecom systems. We will later actually transition from RP1 to eRP1 (see Sect. III-A for definition of eRP), which further extends the range of validity of the solution. Regarding (ii), the question is whether actual data signals just can be put in the form Eq. (9). In the following, we briefly address this latter aspect.

Actual physical transmitted signals (henceforth 'data signals'), which we will denote by $E_{\text{S}}(t)$, have these features: they are finite-energy, limited-peak-power signals of finite bandwidth $[f_l, f_h]$ and finite duration $[t_l, t_h]$. This is of course true also for WDM signals consisting of multiple channels. In the WDM case, $[f_l, f_h]$ would represent the overall WDM comb optical bandwidth. Signals with these features can always be Fourier-transformed. Formally: $E_{\text{S}}(f) = \text{F}\{E_{\text{S}}(t)\}$, where $\text{F}\{\cdot\}$ is the Fourier transform. In general, though, $E_{\text{S}}(f)$ *would not be* in the form of Eq. (9): it could contain some spectral lines but it would typically consist mainly of a continuous spectrum. However, given *any* signal with the above features as $E_{\text{S}}(t)$, then the *periodic signal* built by simply repeating it at regular intervals $T_0$:

$$E_{\text{S}}^{\text{per}}(t) = \sum_{q=-\infty}^{\infty} E_{\text{S}}(t - qT_0) \tag{19}$$

acquires a *pure line spectrum* whose expression is:

$$E_{\text{S}}^{\text{per}}(f) = \sum_{q=q_l}^{q_h} \zeta_q \, \delta(f - qf_0) \tag{20}$$

where $f_0 = 1/T_0$ and $\zeta_q = f_0 E_{\text{S}}(qf_0)$. The summation index $q$ would ideally run over $[-\infty, \infty]$ but $E_{\text{S}}(f)$ is band-limited by assumption to $[f_l, f_h]$. This curtails $q$ within $q_l = \min\{q : qf_0 > f_l\}$ and $q_h = \max\{q : qf_0 < f_h\}$. In conclusion, the signal of Eq. (20) is of the form Eq. (9) and therefore the FWM formalism can be applied to it.

Concerns may arise whether the periodicization performed in Eq. (19) may lead to undesired consequences or loss of generality. There are subtleties to be heeded, such as signal continuity issues at the edges of each period, but the answer is that no loss of generality is incurred.

Note that Eq. (20) is actually more constrained than Eq. (9), because the spectral lines in Eq. (20) are not arbitrarily located in frequency, but rather lie on a *grid of fixed pitch* $f_0$. Thanks to this, without any loss of generality, we can from now on identify the frequencies $f_k, f_m, f_n$ in Eq. (10) as $kf_0, mf_0, nf_0$, respectively. When this writing is inserted into Eq. (10), it can be seen that all generated FWM contributions $f_{\text{FWM}}^{(i)}$ forcedly lie *on the same frequency grid* $qf_0$ where the periodicized signal spectral lines lie too, with $q$:

$$q = (k + m - n) \tag{21}$$

For notational compactness, we then define $\{i\}_q$ as the set of all the indices $i$ that identify triples $(k, m, n)_i$ that produce a FWM line *at the same frequency* $qf_0$, according to Eq. (10). In formulas:

$$\{i\}_q = \left\{ i \, : \, f_{\text{FWM}}^{(i)} = qf_0 \right\} \tag{22}$$

This could be rephrased by saying that the index $q$ identifies the class of equivalence $\{i\}_q$ of all possible triples $(k, m, n)_i$ that produce a FWM spectral line at the *same frequency* $qf_0$. Assuming to be looking at a single span of fiber of length $L_s$, then the total signal, including both data signal and FWM, at the end of the span is given by:

$$E_{\text{tot}}(f, L_s) = E_{\text{S}}(f, L_s) + E_{\text{FWM}}(f, L_s) \tag{23}$$

where:

$$E_{\text{S}}(f, L_s) = \sum_{q=q_l}^{q_h} \zeta_q(L_s) \, \delta(f - qf_0) \tag{24}$$

$$E_{\text{FWM}}(f, L_s) = \sum_{q} \sum_{\{i\}_q} \eta_i(L_s) \, \delta(f - qf_0) \tag{25}$$

The factors $\zeta_q(L_s)$ are the data signal spectral line amplitudes from Eq. (20), linearly propagated to the end of the span according to:

$$\zeta_q(L_s) = \zeta_q \cdot \text{H}(f_q, L_s) \tag{26}$$

and H is the fiber linear transfer function defined in Eq. (16).

### *B. Multispan Links*

Extending the previous equations to *multiple spans* requires that the UPA be augmented by two related assumptions.

One is that the total FWM at the end of a multi-span link can be expressed as the sum of the FWM generated in each generic $n$-th span. The other is that the signal that generates FWM in the $n$-th span consists of just the data signal spectral lines injected at the input of the overall link, then linearly propagated to the input of the $n$-th span.

These assumptions neglect the fact that, at the input of each successive span, besides the data signal, there is also the FWM produced in the previous spans, as well as ASE noise. This approximation is viable provided that, at any point along the link, the data signal power can be assumed to be much larger than the accumulated FWM and ASE power. In practice, this is typically the case because otherwise, if the data signal power was comparable to either the ASE

or FWM power, the data signal would be too degraded for successful reception to take place. This argument fails only for systems that can operate at quite low optical signal-to-noise ratio (OSNR) values. Indicatively, no significant effect of co-propagating ASE or FWM are observed down to about 10 dB OSNR [28], [145]. A clear impact can be observed at about 5 dB OSNR where, indicatively, about 10% maximum reach decrease is found between considering co-propagating ASE and FWM vs. not considering them. In this section we neglect this aspect, which is discussed in more detail in Sect. IX-A.

Analytically, Eq. (15) can then be modified as follows, to provide $\eta_i$ at the end of the link (i.e., at $L_{\rm end}$), which we call $\eta_i^{\rm e}$:

$$\begin{aligned}\eta_i^{\rm e} = -j\zeta_k\zeta_m\zeta_n^* \sum_{n_s=1}^{N_s} \mathbf{H}\left(\bar{f}_i;1,n_s-1\right)\\ \gamma\left(n_s\right) L_{\rm eff}\left(\bar{f}_i;n_s\right)\mu\left(\bar{f}_i;n_s\right)\mathrm{H}\left(f_{\rm FWM}^{(i)};n_s,N_s\right)\end{aligned} \tag{27}$$

where the fiber transfer function $\mathrm{H}\left(f_{\rm FWM}^{(i)};n_s,N_s\right)$ linearly propagates the $i$-th FWM line from where it is generated, i.e., the $n_s$-th span, to the last span $N_s$. H can be found by cascading its single-span version Eq. (16) along the link. The factor $\mathbf{H}\left(\bar{f}_i;1,n_s-1\right)$ is a shorthand for the product of the three transfer functions that propagate linearly the three pumps from the first span to the end of the $(n_s-1)$-th span, that is:

$$\begin{aligned}&\mathbf{H}\left(\bar{f}_i;1,n_s-1\right)=\\ &\mathrm{H}\left(f_k;1,n_s-1\right)\mathrm{H}\left(f_m;1,n_s-1\right)\mathrm{H}^*\left(f_n;1,n_s-1\right)\end{aligned} \tag{28}$$

The notation $L_{\rm eff}(\bar{f}_i;n_s)$ indicates the effective length calculated for the $n_s$-th span alone, relative to the $i$-th triple $(k,m,n)_i$. It is found by using Eq. (17) in the local span length coordinate which runs from 0 to the span length $L_s^{(n_s)}$, and setting $z=L_s^{(n_s)}$. Similarly for $\mu(\bar{f}_i;n_s)$ with Eq. (18). Finally, the summation in Eq. (27) sums the overall FWM produced in each span, then linearly propagated to the end of the link.

We could then use Eq. (22)–Eq. (25) to express the overall data and FWM signals at the end of the multi-span link. We will do it in Sect. IV-D after dealing with dual-polarization propagation.

### C. Dual Polarization

Actual fibers carry two polarizations. Non-linear propagation in dual-polarization (DP) obeys a set of two stochastically-coupled nonlinear differential equations, called the dual-polarization NLSE (DP-NLSE) [1]. The stochastic nature of the coupling is due to the random birefringence of the fiber. However, as pointed out in Sect. I, in typical telecom fibers, it is possible to average the DP-NLSE over the spatial evolution of random birefringence, obtaining the so-called Manakov Equation (ME) [18], [19], [20]. The ME provides a very effective means of modeling DP nonlinear propagation and is very widely used in analytical modeling and simulations.

To address DP-FWM, we first redefine the 'pump' amplitudes $\zeta_q$ in Eq. (9) as DP quantities:

$$\zeta_q = \zeta_{\hat{x},q}\,\hat{x} + \zeta_{\hat{y},q}\,\hat{y} \tag{29}$$

where $\zeta_{\hat{x},q},\zeta_{\hat{y},q}$ are the $\hat{x}$ and $\hat{y}$ pump components, respectively. Then, according to the ME, quite remarkably all of the previously shown formulas remain *unchanged*, with the exception of the substitution of the scalar quantity $\zeta_k\zeta_m\zeta_n^*$, in Eq. (27), with the Jones vector (in the $\hat{x}$ and $\hat{y}$ basis) appearing in square brackets below:

$$\begin{aligned}\eta_i^{\rm e} = -j\tfrac{8}{9}\left[\begin{array}{c}\left(\zeta_k^{\hat{x}}\zeta_m^{\hat{x}}\zeta_n^{\hat{x}\,*}+\zeta_k^{\hat{x}}\zeta_m^{\hat{y}}\zeta_n^{\hat{y}\,*}\right)\\ \left(\zeta_k^{\hat{y}}\zeta_m^{\hat{y}}\zeta_n^{\hat{y}\,*}+\zeta_k^{\hat{y}}\zeta_m^{\hat{x}}\zeta_n^{\hat{x}\,*}\right)\end{array}\right]\sum_{n_s=1}^{N_s} L_{\rm eff}\left(\bar{f}_i;n_s\right)\\ \gamma\left(n_s\right)\mu\left(\bar{f}_i;n_s\right)\mathbf{H}\left(\bar{f}_i;1,n_s-1\right)\mathrm{H}\left(f_{\rm FWM}^{(i)};n_s,N_s\right)\end{aligned} \tag{30}$$

Note, in particular, that $L_{\rm eff}$, $\mu$, $\mathbf{H}$ and H are *scalar quantities* and are unaffected by DP. Of course, as a result of the above substitution, the FWM amplitudes $\eta_i^{\rm e}$ acquire a DP nature too, that is $\eta_i^{\rm e}=\eta_{\hat{x},i}^{\rm e}\,\hat{x}+\eta_{\hat{y},i}^{\rm e}\,\hat{y}$.

### D. The statistical modeling approach

The shown analytical FWM formalism lends itself to the derivation of *statistical models*, which are potentially more powerful design/analysis tools than brute-force numerical NLSE or ME integration and Monte-Carlo performance estimation.

We focus on the *end of the link*, identified as $L_{\rm end}$. In agreement with system-related literature, we call the Kerr-induced nonlinear disturbance at this location 'Non-Linear Interference', or NLI, replacing the 'FWM' acronym, that is:

$$E_{\rm NLI}(f) = E_{\rm FWM}\left(f,L_{\rm end}\right) \tag{31}$$

The formula for $E_{\rm NLI}$ at $L_{\rm end}$ is then, similar to Eq. (25):

$$E_{\rm NLI}(f) = \sum_q \sum_{\{i\}_q} \eta_i^{\rm e}\ \delta\left(f-qf_0\right) \tag{32}$$

In essence, the *statistical approach* (SA) consists of viewing the $\eta_i^{\rm e}$ as random variables (RV's), so that the overall NLI disturbance field Eq. (32) becomes a random process (RPR), whose statistical features can be studied to assess its impact on data signal detection. Note that the $\eta_i^{\rm e}$ are *not* RV's if transmission of a specific fully-defined data signal takes place. They become RV's if transmission of random data signals is assumed. Then, by statistical manipulation, one can extract the average features of NLI over *all possible* data signals, which is the great advantage of the SA over any Monte-Carlo strategy.

In this section, we focus on the GN/EGN class of models. These models aim at finding the *power spectral density* (PSD) of $E_{\rm NLI}(f)$, which we will call $G_{\rm NLI}(f)$ (typically in W/Hz). The reason for the focus on this quantity is that the impact of NLI noise on system performance, typically expressed in terms of a degraded estimated OSNR, can be found based on $G_{\rm NLI}(f)$. Other indicators can then be estimated based on such OSNR, such as BER, MI, GMI, to different accuracy.

We first define for convenience a new RV which is the sum of all the FWM amplitudes *contributing to the same NLI line at frequency* $(q\,f_0)$:

$$\nu_q^{\rm e} = \sum_{\{i\}_q} \eta_i^{\rm e} \tag{33}$$

Given Eqs. (32) and (33), $G_{\rm NLI}(f)$ is then directly found as:

$$G_{\rm NLI}(f) = \sum_{q} \mathrm{E}\left\{ |\, \nu_q^{\rm e}\, |^2 \right\} \delta\left(f - q\, f_0\right) \tag{34}$$

To carry out the calculation of $\mathrm{E}\left\{ |\, \nu_q^{\rm e}\, |^2 \right\}$, with reference to Eq. (30) and (33), we define the *data signal factor*:

$$\mathrm{S}_i = \left[ \begin{array}{c} \left( \zeta_k^{\hat{x}} \zeta_m^{\hat{x}} \zeta_n^{\hat{x}\,*} + \zeta_k^{\hat{x}} \zeta_m^{\hat{y}} \zeta_n^{\hat{y}\,*} \right) \\ \left( \zeta_k^{\hat{y}} \zeta_m^{\hat{y}} \zeta_n^{\hat{y}\,*} + \zeta_k^{\hat{y}} \zeta_m^{\hat{x}} \zeta_n^{\hat{x}\,*} \right) \end{array} \right] \tag{35}$$

and the *link factor:*

$$\begin{aligned} \mathrm{LK}_i = -j\tfrac{8}{9} \sum_{n_s=1}^{N_s} \gamma\left(n_s\right) L_{\rm eff}\left(\bar{f}_i; n_s\right) \mu\left(\bar{f}_i; n_s\right) \\ \mathbf{H}\left(\bar{f}_i; 1, n_s - 1\right) \mathrm{H}\left(f_{\rm FWM}^{(i)}; n_s, N_s\right) \end{aligned} \tag{36}$$

This way, we can compactly write: $\nu_q^{\rm e} = \sum_{\{i\}_q} \mathrm{S}_i\, \mathrm{LK}_i$. We then have:

$$\begin{aligned} \mathrm{E}\left\{ \left| \nu_q^{\rm e} \right|^2 \right\} &= \mathrm{E}\left\{ \left( \sum_{\{i\}_q} \mathrm{S}_i\, \mathrm{LK}_i \right)^{\mathrm{T}*} \sum_{\{i'\}_q} \mathrm{S}_{i'} \mathrm{LK}_{i'} \right\} \\ &= \sum_{\{i\}_q} \sum_{\{i'\}_q} \mathrm{E}\left\{ \mathrm{S}_i^{\mathrm{T}*}\, \mathrm{S}_{i'} \right\} \mathrm{LK}_i^*\, \mathrm{LK}_{i'} \end{aligned} \tag{37}$$

where the superscript 'T' means 'transpose'. Note that the link factor, apart from normalizations, is the same quantity appearing as $\eta$ in Eq. (8), which provides a clue that the FWM formalism with the UPA is equivalent to the RP1 approach (see Sect. III-A).

Eq. (37) is at the core of both the GN and EGN models. It shows that the SA reduces to the evaluation of the expectation $\mathrm{E}\left\{ \mathrm{S}_i \mathrm{S}_{i'}^* \right\}$, since $\mathrm{LK}_i \mathrm{LK}_{i'}^*$ is completely deterministic and it is even available in closed-form in various relevant cases. Note also from Eq. (35) that $\mathrm{S}_i$ *depends only on the data signal at the input of the fiber*, and not on the link features, which are instead confined within $\mathrm{LK}_i$.

Despite its apparent compactness, Eq. (37) does not typically generate simple results. The reason is that the RV's $\zeta_q$'s, which are the amplitudes of the overall WDM data signal spectral lines in Eq. (20), interact nonlinearly in Eq. (37) giving rise to complex sixth-order moments, which generate a large number of terms. Only under drastic simplifying assumptions or special circumstances Eq. (37) and, as a result, the NLI spectrum Eq. (34), take on simple forms. This is the case of the GN-model, which we introduce next.

### *E. The GN-model*

The complexity hidden within Eq. (37) drastically simplifies provided that certain approximations are made. Such approximations give rise to the so-called GN-model. They are as follows.

It is assumed that the $\zeta_q$'s in Eq. (9) are *statistically independent (SI) zero-mean complex circular jointly-Gaussian RVs*. As for their variances $\sigma_{\zeta_q}^2$, they directly relate to the average PSD of the periodicized data signal of Eq. (20), which is given by:

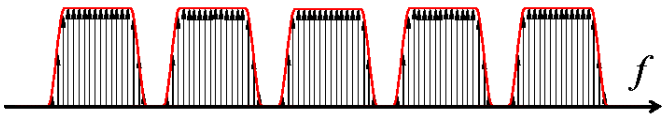


Fig. 5. Solid red line: an example of the PSD $G_{\rm S}(f)$ of a 5-channel aperiodic WDM transmission signal. Black arrows: Dirac's deltas making up the PSD $G_{\rm S}^{\rm per}(f)$ of the periodicized transmission signal Eq. (38), according to Eq. (39).

$$G_{\rm S}^{\rm per}(f) = \sum_{q=q_l}^{q_h} \sigma_{\zeta_q}^2\, \delta\left(f - q\, f_0\right) \tag{38}$$

Then, a logical choice is to assume that the $\sigma_{\zeta_q}^2$ are such that $G_{\rm S}^{\rm per}(f)$ tends to follow the average PSD of the *aperiodic* data signal, $G_{\rm S}(f)$. This is obtained by imposing:

$$\sigma_{\zeta_q}^2 = f_0 \cdot G_{\rm S}\left(q\, f_0\right) \tag{39}$$

where the multiplying constant $f_0$ adjusts dimensions and also ensures signal power consistency. This way, $G_{\rm S}^{\rm per}(f)$ and $G_{\rm S}(f)$ relate as in the pictorial example of Fig. 5.

A key point is whether the main assumptions made so far are reasonable and accurate. As already stated, the signal periodicization of Eq. (19) is of no concern. In addition, in Sect. IV-E1 we will actually show that it can be removed. The PSD matching of Eq. (39), depicted in Fig. 5, is non-problematic either. The zero-mean complex circular Gaussian distribution assumption for each one of the $\zeta_q$'s, *individually*, is actually quite accurate, too. What, in general, is *not* accurate, is the assumption that the $\zeta_q$'s are SI. In fact, assuming conventional memoryless modulation, with independent symbols, and standard PM-QAM constellations, then the $\zeta_q$'s *are not independent* within each single WDM channel. Also, their dependence is not straightforward since, even though they are individually Gaussian, they are not jointly-Gaussian.

The GN-model *neglects* the statistical dependence of the $\zeta_q$'s within each WDM channel and this approximation can be expected to generate some error, which we will discuss later on. On the other hand, it allows for extremely simple handling of the average $\mathrm{E}\left\{ \mathrm{S}_i \mathrm{S}_{i'}^* \right\}$. The very compact result, which can be obtained through Isserlis's theorem, is:

$$\begin{aligned} \mathrm{E}\left\{ \mathrm{S}_i^{\mathrm{T}*}\, \mathrm{S}_{i'} \right\} &= \tfrac{3}{4} \sigma_{\zeta_k}^2 \sigma_{\zeta_m}^2 \sigma_{\zeta_n}^2 \delta_{ii'} \\ &= \tfrac{3}{4} f_0^3 G_{\rm S}\left(k f_0\right) G_{\rm S}\left(m f_0\right) G_{\rm S}\left(n f_0\right) \delta_{ii'} \end{aligned} \tag{40}$$

It should be mentioned that the above formula is valid when the three indices $k$,$m$,$n$ of the $i$-th triple are *all different*. Using it when this is not the case introduces an approximation, which is however of vanishing impact for decreasing $f_0$. We consequently assume that a small-enough $f_0$ is chosen. Interestingly, it can be shown that introducing such approximation and assuming a small-enough $f_0$ also makes the result tend to the more accurate eRP perturbation approach than the RP which would be found otherwise (see Sect. III-A).

Once Eq. (40) is plugged into Eq. (37), the Kronecker's delta $\delta_{ii'}$ reduces the double summation to a single one. As a result, the power carried by the NLI spectral line at frequency $(qf_0)$ is given by:

$$\mathrm{E}\left\{\left|\nu_q^{\mathrm{e}}\right|^2\right\} = \frac{16}{27} f_0^3 \sum_{\{i\}_q} G_{\mathrm{S}}\left(kf_0\right) G_{\mathrm{S}}\left(mf_0\right) G_{\mathrm{S}}\left(nf_0\right) \left|\mathrm{LK}_i\right|^2 \tag{41}$$

We must now make explicit the implicit indexing system used in Eq. (41), defined in Eq. (22), and write equivalently but *explicitly*:

$$\begin{aligned}\mathrm{E}\left\{\left|\nu_q^{\mathrm{e}}\right|^2\right\} = \tfrac{16}{27} f_0^3 \sum_{k=q_l}^{q_h} \sum_{m=q_l}^{q_h} \left|\mathrm{LK}_{km(k+m-q)}\right|^2 \\ G_{\mathrm{S}}\left(kf_0\right) G_{\mathrm{S}}\left(mf_0\right) G_{\mathrm{S}}\left(\left[k+m-q\right] f_0\right)\end{aligned} \tag{42}$$

where the relation $n = (k + m - q)$, derived from Eq. (21), was used. The link factor $\mathrm{LK}_{kmn}$ is now written as:

$$\begin{aligned}\mathrm{LK}_{kmn} = -j \sum_{n_s=1}^{N_s} \gamma\left(n_s\right) L_{\mathrm{eff}}\left(kf_0, mf_0, nf_0; n_s\right) \\ \mu\left(kf_0, mf_0, nf_0; n_s\right) \mathrm{H}\left(qf_0; n_s, N_s\right) \\ \mathbf{H}\left(kf_0, mf_0, nf_0; 1, n_s - 1\right)\end{aligned} \tag{43}$$

The above Eqs. (42) and (43), together with Eq. (34), provide *a general frequency-discrete GN-model equation set*. These formulas could be directly used to calculate the PSD of NLI by executing the indicated summations numerically. However, a *frequency-continuous* set of formulas can be derived from them.

*1) Continuous frequency:* It is possible to re-cast the GN-model in integral form by going from discrete to continuous-frequency. This is useful also because in certain cases it allows analytical closed-form solutions to be obtained.

The procedure consists of lengthening the periodicity $T_0$ of the data signal in Eq. (20). In the limit of $T_0 \to \infty$, that is of the data signal effectively becoming aperiodic and infinitely extended in time, then also $f_0 \to 0$ and through rather straightforward mathematical manipulation, the *frequency continuous GN-model* is found. The PSD of NLI is given by:

$$\begin{aligned}G_{\mathrm{NLI}}\left(f\right) = \tfrac{16}{27} \int_{f_l}^{f_h} \int_{f_l}^{f_h} G_{\mathrm{S}}\left(f_1\right) G_{\mathrm{S}}\left(f_2\right) G_{\mathrm{S}}\left(f_1 + f_2 - f\right) \\ \left|\mathrm{LK}\left(f_1, f_2, f_1 + f_2 - f\right)\right|^2 df_1 df_2\end{aligned} \tag{44}$$

$$\begin{aligned}\mathrm{LK}\left(f_1, f_2, f_3\right) = -j \sum_{n_s=1}^{N_s} \gamma\left(n_s\right) L_{\mathrm{eff}}\left(\left(f_1, f_2, f_3\right); n_s\right) \\ \mu\left(\left(f_1, f_2, f_3\right); n_s\right) \mathrm{H}\left(f_1 + f_2 - f_3; n_s, N_s\right) \\ \mathbf{H}\left(\left(f_1, f_2, f_3\right); 1, n_s - 1\right)\end{aligned} \tag{45}$$

For the readers' convenience, below we replace the shorthands appearing in Eq. (45) with their definitions, and include certain elements, such as lumped gain/loss and possible dispersion-compensating units (DCUs), that have not been explicitly introduced so far. The result is:

$$\begin{aligned}&\mathrm{LK}\left(f_1, f_2, f_3\right) = -j \sum_{n_s=1}^{N_s} \gamma_{n_s} \\ &\int_0^{L_s^{(n_s)}} e^{\int_0^{z'} \left[\kappa_{n_s}\left(f_1, z''\right) + \kappa_{n_s}^*\left(f_3, z''\right) + \kappa_{n_s}\left(f_2, z''\right) - \kappa_{n_s}\left(f_1+f_2-f_3, z''\right)\right] dz''} dz' \\ &\prod_{p=n_s}^{N_s} \Gamma_p^{1/2}\left(f_1 + f_2 - f_3\right) e^{\int_0^{L_s^{(p)}} \kappa_p(f_1+f_2-f_3, z) dz} e^{-j\beta_{\mathrm{DCU}}^{(p)}(f_1+f_2-f_3)} \\ &\prod_{p=1}^{n_s-1} \left[\Gamma_p\left(f_1\right) \Gamma_p\left(f_2\right) \Gamma_p\left(f_3\right)\right]^{\frac{1}{2}} e^{\int_0^{L_s^{(p)}} \left[\kappa_p(f_1, z) + \kappa_p(f_2, z) + \kappa_p^*(f_3, z)\right] dz} \\ &e^{-j\left[\beta_{\mathrm{DCU}}^{(p)}(f_1) + \beta_{\mathrm{DCU}}^{(p)}(f_2) - \beta_{\mathrm{DCU}}^{(p)}(f_3)\right]}\end{aligned} \tag{46}$$

where $\Gamma_p$ is lumped power gain/loss and $\beta_{\mathrm{DCU}}^{(p)}(f)$ is the effect of a DCU element, both located at the end of the $p$-th span. The quantity $\kappa_p$ is the complex propagation constant, defined in Eq. (13), for the $p$-th span. Eq. (46) is quite general as it can handle both distributed and lumped loss/gain, also frequency dependent, as well as the possible presence of DCUs. Note that if $\Gamma_p$ had a phase-shift associated with it, that is its field transfer function was $\Gamma_p^{1/2} e^{j\Phi_\Gamma^{(p)}(f)}$ such phase-shift could be introduced in the formula by formally replacing the DCU phase-shift $\beta_{\mathrm{DCU}}^{(p)}(f)$ with: $\left[\beta_{\mathrm{DCU}}^{(p)}(f) - \Phi_\Gamma^{(p)}(f)\right]$.

### *F. Noteworthy Link Factors*

In this section we focus on closed-forms or simplified forms for the LK factor, which can be found for various cases of interest. LK is a function of the frequencies of the three 'pumps' generating a FWM contribution. This dependence can be cast in the discrete framework of Eqs. (42)-(43) through the indices $k, m, n$ or in the continuous framework of Eqs. (44)-(45) using frequencies $f_1, f_2, f_3$. We use the latter, but it is possible to convert from one to the other through the direct substitutions: $f_1 = kf_0$, $f_2 = mf_0$, $f_3 = nf_0$.

*1) The transparent and uniform link with lumped amplification:* A *transparent span* is such that its optical gain equals its loss exactly, from the input of that span to the input of the next span. A *transparent link* is one where every span is transparent. A link is *uniform if all spans are identical*.

Under the assumptions of a transparent and uniform link, Eq. (45) takes on the closed-form:

$$\begin{aligned}\mathrm{LK}\left(f_1, f_2, f_3\right) = -jN_s\gamma L_{\mathrm{eff}}\left(f_1, f_2, f_3\right) \mu\left(f_1, f_2, f_3\right) \\ e^{-jN_sL_s\beta(f_1+f_2-f_3)} e^{j\frac{\beta_{123}}{2}(N_s-1)L_s} \mathrm{D}\left(L_s\beta_{123}/2, N_s\right)\end{aligned} \tag{47}$$

where the dependence of $L_{\mathrm{eff}}$ and $\mu$ on $n_s$ was removed since they are identical for all spans, and we use the shorthand:

$$\beta_{123} = \beta\left(f_1 + f_2 - f_3\right) - \beta\left(f_1\right) - \beta\left(f_2\right) + \beta\left(f_3\right) \tag{48}$$

$\mathrm{D}(x, N) = \frac{\sin(Nx/2)}{N\sin(x/2)}$ is the periodic-sinc Dirichlet function of order $N$ [258].

If we further assume that the fiber loss coefficient $\alpha$ is frequency-independent and that lumped (non-distributed)

frequency-independent amplification is used, with the amplifier placed at the end of each span, then $L_{\text{eff}}$ and $\mu$ simplify to:

$$L_{\text{eff}} = \left(1 - e^{-2\alpha L_s}\right)/2\alpha \tag{49}$$

$$\mu\left(f_1, f_2, f_3\right) = \frac{1}{L_{\text{eff}}} \cdot \frac{1 - e^{-2\alpha L_s} e^{j\beta_{123} L_s}}{2\alpha - j\beta_{123}} \tag{50}$$

Note that quite often $\beta$ is accounted for through the so-called $\beta_2$ and $\beta_3$ coefficients, as follows:

$$\beta(f) = 2\pi^2 \beta_2 f^2 + \frac{4}{3}\pi^3 \beta_3 f^3 \tag{51}$$

If so, Eq. (48) is re-written as:

$$\beta_{123} = 4\pi^2 \left(f_1 - f_3\right)\left(f_2 - f_3\right)\left[\beta_2 + \pi\beta_3\left(f_1 + f_2\right)\right] \tag{52}$$

*2) The uniform lossless link:* Lossless links are often used in research papers as a limiting case of ideal distributed amplification. A uniform lossless link with no DCUs can actually be viewed as a single span, whose length is that of the whole link. Using Eqs. (17)-(18), with $\alpha = 0$, $g = 0$ and integration length $z = L_{\text{tot}}$, where $L_{\text{tot}}$ is the whole link length, then we have that $L_{\text{eff}} = L_{\text{tot}}$ and:

$$\mu\left(f_1, f_2, f_3\right) = e^{j\beta_{123} L_{\text{tot}}/2} \operatorname{Sinc}\left(\beta_{123} L_{\text{tot}}/2\right) \tag{53}$$

where $\operatorname{Sinc}(x) = \sin(x)/x$. Using these results, the link factor becomes:

$$\text{LK}\left(f_1, f_2, f_3\right) = -j\gamma L_{\text{tot}} \mu\left(f_1, f_2, f_3\right) e^{-j\beta(f_1+f_2-f_3)L_{\text{tot}}} \tag{54}$$

If $\beta$ is expressed as Eq. (51), then Eq. (52) allows to modify Eqs. (53)-(54) accordingly.

Note that there are substantial differences between the link factor of a uniform and transparent lossy multispan link Eq. (47) and a lossless one Eq. (54). Indeed, nonlinearity presents strongly different features in the two set-ups, so a lossless link should never be used as an approximation of a lossy one in practical systems studies.

*3) The uniform transparent link with undepleted backward-pumped Raman amplification:* Raman amplification has become a key staple of modern coherent terrestrial links, where it can provide equivalent noise-figure *gains* on the order of 4-6 dB. Eq. (45) is capable of handling any distributed gain profile, and therefore Raman amplification too, but it would typically require numerical integration of Eqs. (17) and (18) and in general the evaluation of LK would be quite involved.

A closed-form expression for LK is available provided that the following simplifying assumptions and approximations are made: the link is uniform and transparent; gain is due to a single counter-propagating Raman pump; the Raman pump is *undepleted*; Raman gain and fiber loss are frequency-independent. Note that the Raman gain need not balance exactly fiber loss: we assume however transparency so either lumped loss or lumped gain are assumed to be present at the end of each span to bring the total span gain/loss to balance, from which $\mid \text{H}\left(f; n_s, N_s\right) \mid = 1$, for any $n_s$.

The distributed (field) gain due to a single undepleted backward-propagating Raman pump at a location $z$ within a span whose local spatial coordinate is $z \in [0, L_s]$, is:

$$g\left(z\right) = C_{\text{r}} P_{\text{r}_0} e^{2\alpha_{\text{r}} z}/2 \tag{55}$$

where $C_{\text{r}}$ is the Raman gain coefficient 1/(W km), $\alpha_{\text{r}}$ (1/km) is the (field) attenuation of the fiber at the Raman pump frequency and $P_{\text{r}_0}$ is the power of the pump at $z = 0$, i.e., at the start of the span. Note that, since the pump is injected at the end of the span and propagates backwards, its power-profile actually grows exponentially in the forward direction, as shown by Eq. (55). The calculations are fully reported in [50] and the relevant results are:

$$\begin{aligned} L_{\text{eff}} = {} & \frac{e^{-\frac{C_{\text{r}} P_{\text{r}_0}}{2\alpha_{\text{r}}}}}{2\alpha_{\text{r}}} \cdot \left(-\frac{C_{\text{r}} P_{\text{r}_0}}{2\alpha_{\text{r}}}\right)^{\frac{\alpha}{\alpha_{\text{r}}}} \\ & \left[\Gamma\left(-\frac{\alpha}{\alpha_{\text{r}}}, -\frac{C_{\text{r}} P_{\text{r}_0}}{2\alpha_{\text{r}}}\right) - \Gamma\left(-\frac{\alpha}{\alpha_{\text{p}}}, -\frac{C_{\text{r}} P_{\text{r}}}{2\alpha_{\text{r}}}\right)\right] \end{aligned} \tag{56}$$

$$\begin{aligned} \mu\left(f_1, f_2, f_3\right) = {} & \frac{e^{-\frac{C_{\text{r}} P_{\text{r}_0}}{2\alpha_{\text{r}}}}}{2\alpha_{\text{r}} L_{\text{eff}}} \left(-\frac{C_{\text{r}} P_{\text{r}_0}}{2\alpha_{\text{r}}}\right)^{\frac{2\alpha - j\beta_{123}}{2\alpha_{\text{r}}}} \\ & \left[\Gamma\left(-\frac{2\alpha - j\beta_{123}}{2\alpha_{\text{r}}}, -\frac{C_{\text{r}} P_{\text{r}_0}}{2\alpha_{\text{r}}}\right) - \Gamma\left(-\frac{2\alpha - j\beta_{123}}{2\alpha_{\text{r}}}, -\frac{C_{\text{r}} P_{\text{r}}}{2\alpha_{\text{r}}}\right)\right] \end{aligned} \tag{57}$$

where $P_{\text{r}} = P_{\text{r}_0} e^{2\alpha_{\text{r}} L_s}$ is the power of the pump as it is injected at the end of the span, and:

$$\Gamma\left(x, y\right) = \int_y^\infty w^{x-1} e^{-w} dw$$

is the *upper incomplete* Gamma function. Here too, if $\beta$ is expressed as Eq. (51), then Eq. (52) allows to modify Eq. (57) accordingly. The overall expression of LK is still Eq. (47).

### G. The incoherent GN-model

Among the assumptions leading to the GN-model, there is the premise that the NLI, generated by FWM in each span, adds up at the *field* level at the end of the link, that is:

$$E_{\text{NLI}}^{\text{end}}\left(f\right) = \sum_{n_s=1}^{N_s} E_{\text{NLI}}^{(n_s)}\left(f\right) \tag{58}$$

Eqs. (44)-(45) were derived accordingly. A possible alternative assumption is that NLI adds up in *power* at the end of the link, that is:

$$G_{\text{NLI}}^{\text{end}}\left(f\right) = \sum_{n_s=1}^{N_s} G_{\text{NLI}}^{(n_s)}\left(f\right) \tag{59}$$

This assumption, is called the *incoherent accumulation approximation*, or IAA. The GN-model version that uses it is the *incoherent GN-model*, or iGN-model. The iGN-model equations are Eq. (44) and the following (replacing Eq. (45))

$$\begin{aligned} & \left|\text{LK}\left(f_1, f_2, f_3\right)\right|^2 = \\ & \sum_{n_s=1}^{N_s} \left|\gamma\left(n_s\right) L_{\text{eff}}\left(\left(f_1, f_2, f_3\right); n_s\right) \mu\left(\left(f_1, f_2, f_3\right); n_s\right)\right|^2 \\ & \left|\text{H}\left(f_1 + f_2 - f_3; n_s, N_s\right) \mathbf{H}\left(\left(f_1, f_2, f_3\right); 1, n_s - 1\right)\right|^2 \end{aligned} \tag{60}$$

The IAA is a rather drastic approximation. On the other hand, Eq. (59) brings about substantial simplification which is greatly advantageous in the practical use of the GN-model. There are three main aspects in which it helps.

First, if the GN-model is evaluated by numerical integration of Eqs. (44)-(45), the $|\text{LK}|^2$ factor typically turns out to be very rapidly oscillating vs. the integration variables, requiring

a fine integration grid or sophisticated integration strategies [50]. This can be appreciated by looking for instance at a closed-form LK, such as for the transparent and uniform link, where the Dirichlet function in Eq. (47) has a rapidly oscillatory behavior. Such behavior is due to the phase interference of the NLI contributions from different spans, occurring at the end of the link (see Fig. 21 in [50]). If the IAA is made, then this interference does not occur and, in the example in question, the Dirichlet function factor disappears altogether.

Secondly, the simplified analytical form of the LK factor makes it easier to find fully closed-form solutions for the overall NLI PSD $G_{\mathrm{NLI}}(f)$, see Sect. IV-J.

Thirdly, Eq. (59) makes it easy to assess NLI within complex reconfigurable networks, because the computation of NLI within each span, as well as the optimization of certain aspects such as launch power, become dependent only on span-local features. NLI is then easily propagated from that span on, by accounting only for loss and gain along the path. This was recognized early on and has developed, among others, into the LOGO network optimization and control strategy [88], [89], [90], which is currently in use in physical-layer-aware optical networking products of some vendors.

Of course the above advantages would be moot if the IAA induced too large errors. However, despite the drastic nature of the IAA, the results of the iGN vs. the GN-model are typically quite close. In fact, in many cases the iGN-model appears more accurate than the GN-model, due to the fortuitous circumstance that the error due to the Gaussian data signal assumption made by the GN-model partially cancels out the error induced by the IAA. This aspect was studied in detail in [62] (Sects. III-D and V-D) and in [260]. It is dealt with in the next section (Sect. IV-G1) too.

*1) GN-model accuracy and validity range:* For theoretical reasons whose discussion can be found in the rich literature on the topic, such as [54], [55], [57], [61], [91], [92], the GN-model approximations tend to work well and produce accurate results when:

(1) the symbol rate is large
(2) fiber dispersion is large
(3) loss per span is not too small
(4) amplification is prevalently lumped
(5) the number of traversed spans is sufficiently large.

In agreement with the above, the GN-model has shown excellent predictive capability in ultra-long-haul (ULH) submarine transmission over PSCF at large symbol rate. Various level of inaccuracy have instead been found for instance over very low dispersion fibers, at low symbol rates, with short reach (1-3 spans). Note, though, that the model is rather robust in the sense that, typically, several of the above conditions must be simultaneously or severely violated in order to see large errors in NLI prediction by the model. Also, the GN-model error is *always* towards underestimating system performance, i.e., it is always conservative. Such underestimation error is typically in the range -5% to -20% of maximum reach.

Several dedicated GN-model test and validation studies have been carried out over the years, both simulative and experimental. Examples of the experimental ones are [69], [71] [72], [73], [74], [75], [76], [90], where the GN-model consistently proved quite accurate. Two of these experiments were specifically designed to test the model over different fiber types [69], [74]. In particular, the latter addressed seven quite different fiber types. A massive ULH experiment [71] showed a very good accuracy of GN-model predictions for a 106-channel system at 30 GBaud channel rate, with PM-16QAM transmission, over 10290 km, with and without ideal DBP.

Regarding simulative tests, a great many papers have been published. A recent broad study [260] addressed a very wide range of scenarios, including five QAM formats (from PM-QPSK to PM-64QAM) at 32 GBaud, three fibers (NZ-DSF, SMF, PSCF), three channel spacings (33.6, 27.5 and 50 GHz) and two span lengths (60 and 100 km). Mixed Raman/EDFA amplification was also addressed. The range of maximum reach (MR) probed spanned from 200 to 10,000 km. Criteria (1)-(5) above were confirmed and a gradual error increase was found when such criteria were departed from. Along similar lines, two newer studies [261], [262], have explored the realm of high symbol rates (from 64 to 512 GBaud), showing quite compellingly that, as symbol rates go up, the GN-model becomes more accurate. We will come back to this aspect after introducing the EGN model, in the next section.

Here, in Fig. 6a, we provide an original set of results where we focused on 15-channel systems operating at the next-generation industry-standard symbol-rate of 64 GBaud. The test was run on either single-mode-fiber (SMF) or a challenging non-zero dispersion-shifted-fiber (NZDSF) with high nonlinearity coefficient and very low dispersion. Six QAM formats were tested, from PM-8QAM to PM-256QAM, with three different spacings. The detailed system data are reported in the figure caption. MR at optimum launch power is shown, found based on reaching a target minimum generalized mutual information (GMI) equal to 87% of each format entropy, corresponding to about 15% FEC coding overhead. For instance, the entropy of PM-32QAM is 10 bit/symbol and the target GMI was set to $0.87 \cdot 10 = 8.7$ bit/symbol. The abscissa is the net system spectral efficiency in bit/(s Hz), calculated as shown in the caption.

Fig. 6a confirms that the GN-model always underestimates MR, in the SMF case by 7% for PM-8QAM, up to to 17% for PM-256QAM. This larger error is due to the very short reach of PM-256QAM (1 span), severely violating condition (5). Errors are somewhat larger in the case of the NZDSF (which violates condition (2)), again especially at short reach (violating (2) and (5) together). The incoherent GN model is better overall, for the reasons explained at the bottom of Sect. IV-G.

In summary, the GN-model and its incoherent version provide a low-complexity, conservative tool for the estimation of system performance across a very wide range of system scenarios. As we will later show, it also lends itself to the derivation of fully-closed form approximate solutions. These aspects justify its success and widespread use. Later, we will also show that the possible adoption of Gaussian-shaped constellation for optical transmission may actually provide a renewed boost to the use of the GN-model.

### H. The EGN-model

As mentioned above, the GN-model loses accuracy when criteria (1)-(5) are violated. Also, because of the Gaussian data signal assumption, it does not account for certain non-linearity features, such as the dependence of NLI generation on modulation format. These problems were recognized early on [54], [55], [57], and in [57] a procedure was proposed by which the GN-model could be modified to avoid them. In [63] the so-called *EGN-model* was fully derived following up on the procedure introduced in [57]. Extended calculation details are reported in [64].

Both the GN and the EGN model descend from Eq. (37). They differ as to the *result of the expectation* $\mathrm{E}\left\{\mathrm{S}_i\mathrm{S}_{i'}^*\right\}$, where $\mathrm{S}_i$ is given by Eq. (35). This is because the RV's $\zeta_n$, which make up $\mathrm{S}_i$, are assumed SI for the GN-model, whereas in the EGN-model their statistical dependence is taken into account. As a consequence, the sixth-order moments of the $\zeta_n$'s, which appear within $\mathrm{E}\left\{\mathrm{S}_i\mathrm{S}_{i'}^*\right\}$, yield a much more complex result than Eq. (40). When such result is put into Eq. (37), and the transition from discrete frequency to continuous frequency is made (see Sect. IV-E1), not just one, as in the GN-model case, but a total of *twelve* integral terms are generated, each with different integrand functions and different integration domains.

Interestingly, one of these integral terms *corresponds to the GN-model itself*, so that it is possible to write the EGN-model as:

$$G_{\mathrm{NLI}}^{\mathrm{EGN}}(f) = G_{\mathrm{NLI}}^{\mathrm{GN}}(f) - G_{\mathrm{NLI}}^{\mathrm{corr}}(f) \tag{61}$$

where $G_{\mathrm{NLI}}^{\mathrm{GN}}(f)$ is the GN-model term and $G_{\mathrm{NLI}}^{\mathrm{corr}}(f)$ collects the other eleven integral terms. It can be thought of as a 'correction' which originates from the statistical dependence of the $\zeta_n$'s, not taken into account in $G_{\mathrm{NLI}}^{\mathrm{GN}}(f)$. Incidentally, such statistical dependence changes depending on the modulation format, so while $G_{\mathrm{NLI}}^{\mathrm{GN}}(f)$ is format-independent, $G_{\mathrm{NLI}}^{\mathrm{corr}}(f)$ is. Note also that $G_{\mathrm{NLI}}^{\mathrm{corr}}(f)$ is intentionally presented with a minus sign in Eq. (61), to stress the fact that the EGN correction typically *decreases* NLI. In fact, it *always* decreases NLI if PM-QAM signals are assumed and this explains the observed feature of $G_{\mathrm{NLI}}^{\mathrm{GN}}(f)$ always overestimating NLI.

A complete version of the quite complex EGN-model formulas is reported in [63], [259]. We will not reproduce them here for lack of space. We will focus instead on an approximate *reduced* EGN-model in Sect. IV-H2.

*1) EGN model accuracy:* The EGN-model accuracy has been extensively investigated simulatively (see for instance [260], [261], [262]). The results have consistently shown that the accuracy of the EGN-model is excellent, even when one or more of the conditions (1)-(5), which are critical for the GN-model (see top of Sect. IV-G1), are not met. As we did for the GN-model, we show in Fig. 6a a collection of tests on QAM systems operating at 64 GBaud, over SMF and NZDSF, with 100 km spans (see Sect. IV-G1 and the caption of Fig. 6a for setup details). The very remarkable feature of the plot is the flawless coincidence of the EGN model MR prediction with the simulation results, across all system configurations, MR values, channel spacings and fibers.

Note that the star markers in Fig. 6a are simulations that use a receiver that is optimum in AWGN and *makes no attempt* at mitigating the component of NLI which can be classified as *nonlinear phase-noise* (NLPN). Such NLI component turns out to be correlated over time, to an extent that roughly grows as the accumulated dispersion along the link and as the symbol rate squared. If enough time-correlation is present, NLPN can be removed by means of relatively simple carrier-phase estimation (CPE) algorithms and a somewhat better performance than predicted by the EGN model can be achieved. This aspect is discussed extensively in [260], [261], [262] and is dealt with in Sect. V-2 of this chapter.

*2) The reduced EGN model:* The EGN model is very accurate, but quite complex. On the other hand, in virtually all cases of practical interest, very good accuracy is still achieved by using the so-called EGN-SCI-X1 [259] approximation, which we call here *reduced EGN-model*. It consists of neglecting those EGN-model terms whose contribution is typically the smallest. The remaining terms are still substantially more complex, as a whole, than the GN-model, but about only 1/3 as complex as the complete EGN-model. With reference to Fig. 6a, the curves of the reduced EGN-model would be completely superimposed to those of the EGN-model (less than 1% MR difference), in all cases presented there.

To deal with the EGN-model, we have to introduce a different notation from that used for the GN-model in Eq. (44). The reason is that the term $G_{\mathrm{NLI}}^{\mathrm{corr}}(f)$, cannot be expressed in terms of the WDM signal PSD. Rather, the Fourier transforms of the individual WDM channel *pulses* are called into play. Please refer to Appendix B for a detailed list of symbol definitions.

The overall WDM data signal is written in time-domain as:

$$\begin{aligned} s_{\mathrm{WDM}}(t) = & \sum_{n_{\mathrm{ch}}=1}^{N_{\mathrm{ch}}} \sum_{r} \left(a_{x,n_{\mathrm{ch}}}^{r}\hat{x} + a_{y,n_{\mathrm{ch}}}^{r}\hat{y}\right) \\ & s_{n_{\mathrm{ch}}}\left(t - rT_{n_{\mathrm{ch}}}\right) e^{j2\pi f_{n_{\mathrm{ch}}}t} \end{aligned} \tag{62}$$

The $a_{x,n_{\mathrm{ch}}}^{r}$ and $a_{y,n_{\mathrm{ch}}}^{r}$ are the symbols sent on the $\hat{x}$ and $\hat{y}$ polarizations in the $r$ signaling time-slot, on the $n_{\mathrm{ch}}$ the WDM channel. We then assume that all quantities are normalized so that the power carried by each WDM channel is given by:

$$P_{n_{\mathrm{ch}}} = \mathrm{E}\left\{\left|a_{x,n_{\mathrm{ch}}}^{r}\right|^2 + \left|a_{y,n_{\mathrm{ch}}}^{r}\right|^2\right\} \tag{63}$$

We also define the following quantities related to the fourth and sixth moments of the channel symbols:

$$\Phi = 2 - \frac{\mathrm{E}\left\{|a|^4\right\}}{\mathrm{E}^2\left\{|a|^2\right\}} \quad , \quad \Psi = -\frac{\mathrm{E}\left\{|a|^6\right\}}{\mathrm{E}^3\left\{|a|^2\right\}} + 9\frac{\mathrm{E}\left\{|a|^4\right\}}{\mathrm{E}^2\left\{|a|^2\right\}} - 12 \tag{64}$$

where $a$ is any of the $a_{x,n_{\mathrm{ch}}}^{r}$ or of the $a_{y,n_{\mathrm{ch}}}^{r}$, which are assumed to be all identically distributed. The exact values of $\Phi$ and $\Psi$ for the most commonly used constellations are shown in Table I. They vary substantially among low-cardinality constellations, whereas they change little among high-cardinality ones. We also report the limit values for a QAM constellation made up of infinitely many signal points, uniformly distributed within a square region whose center is the origin (the PM-$\infty$-QAM entry in Table I). Note that the values for PM-64QAM are already very close to such limit.

| format | $\Phi$ | $\Psi$ |
|---|---|---|
| PM-BPSK | 1 | -4 |
| PM-QPSK | 1 | -4 |
| PM-8QAM | 2/3 | -2 |
| PM-16QAM | 17/25 | -52/25 |
| PM-QPSK | 69/100 | -211/100 |
| PM-64QAM | 13/21 | -5548/3087 |
| PM-128 | 1105/1681 | -135044/68921 |
| PM-256 | 257/425 | -12532/7225 |
| PM-$\infty$-QAM | 3/5 | -12/7 |
| PM-Gaussian | 0 | 0 |

TABLE I
EXACT VALUES OF THE $\Phi$ AND $\Psi$ PARAMETERS.

To compute the GN-model contribution $G_{\mathrm{NLI}}^{\mathrm{GN}}(f)$ according to the signal notation introduced above, Eq. (44) can still be used, with the substitution:

$$G_{\mathrm{S}}(f) = \sum_{n_{\mathrm{ch}}=1}^{N_{\mathrm{ch}}} P_{n_{\mathrm{ch}}} R_{n_{\mathrm{ch}}} |\tilde{s}_{n_{\mathrm{ch}}}(f)|^2 \tag{65}$$

where $\tilde{s}_{n_{\mathrm{ch}}}(f) = \mathrm{F}\left\{ s_{n_{\mathrm{ch}}}(t)\, e^{j2\pi f_{n_{\mathrm{ch}}} t} \right\}$. Regarding the correction term $G_{\mathrm{NLI}}^{\mathrm{corr}}(f)$, its overall expression, for the *reduced* EGN model, is:

$$\begin{aligned} G_{\mathrm{NLI}}^{\mathrm{corr}}(f) = {} & P_{m_{\mathrm{c}}}^3 \left[ \Phi_{m_{\mathrm{c}}}\, \rho_{\mathrm{SCI}}^{m_{\mathrm{c}}}(f) + \Psi_{m_{\mathrm{c}}}\, \tau_{\mathrm{SCI}}^{m_{\mathrm{c}}}(f) \right] \\ & + P_{m_{\mathrm{c}}} \sum_{\substack{n_{\mathrm{ch}}=1 \\ n_{\mathrm{ch}} \neq m_{\mathrm{c}}}}^{N_{\mathrm{ch}}} P_{n_{\mathrm{ch}}}^2\, \Phi_{n_{\mathrm{ch}}} \rho_{\mathrm{X1}}^{n_{\mathrm{ch}}}(f) \end{aligned} \tag{66}$$

where we have assumed that the channel under test (CUT) is the $m_{\mathrm{c}}$-th channel, *not necessarily* the center channel in the WDM comb.

The terms in Eq. (66) bearing the subscript 'SCI' (Self-Channel Interference) are EGN correction terms to the NLI produced by the CUT onto itself. Their expression is:

$$\begin{aligned} & \rho_{\mathrm{SCI}}^{m_{\mathrm{c}}}(f) = \tfrac{80}{81} R_m^2 \int_{f_{m_{\mathrm{c}}}-B_{m_{\mathrm{c}}}/2}^{f_{m_{\mathrm{c}}}+B_{m_{\mathrm{c}}}/2} df_1 \int_{f_{m_{\mathrm{c}}}-B_{m_{\mathrm{c}}}/2}^{f_{m_{\mathrm{c}}}+B_{m_{\mathrm{c}}}/2} df_2 \int_{f_{m_{\mathrm{c}}}-B_{m_{\mathrm{c}}}/2}^{f_{m_{\mathrm{c}}}+B_{m_{\mathrm{c}}}/2} df_2^{'} \cdot \\ & |\tilde{s}_{m_{\mathrm{c}}}(f_1)|^2 \tilde{s}_{m_{\mathrm{c}}}(f_2) \tilde{s}_{m_{\mathrm{c}}}^*(f_2^{'}) \tilde{s}_{m_{\mathrm{c}}}^*(f_1+f_2-f) \tilde{s}_{m_{\mathrm{c}}}(f_1+f_2^{'}-f) \cdot \\ & \mathrm{LK}(f_1, f_2, f_1+f_2-f)\, \mathrm{LK}^*\left(f_1, f_2^{'}, f_1+f_2^{'}-f\right) \\ & + \tfrac{16}{81} R_{m_{\mathrm{c}}}^2 \int_{f_{m_{\mathrm{c}}}-B_{m_{\mathrm{c}}}/2}^{f_{m_{\mathrm{c}}}+B_{m_{\mathrm{c}}}/2} df_1 \int_{f_{m_{\mathrm{c}}}-B_{m_{\mathrm{c}}}/2}^{f_{m_{\mathrm{c}}}+B_{m_{\mathrm{c}}}/2} df_2 \int_{f_{m_{\mathrm{c}}}-B_{m_{\mathrm{c}}}/2}^{f_{m_{\mathrm{c}}}+B_{m_{\mathrm{c}}}/2} df_2^{'} \cdot \\ & |\tilde{s}_{m_{\mathrm{c}}}(f_1+f_2-f)|^2 \tilde{s}_{m_{\mathrm{c}}}(f_1) \tilde{s}_{m_{\mathrm{c}}}(f_2) \tilde{s}_{m_{\mathrm{c}}}^*(f_1+f_2-f_2^{'}) \tilde{s}_{m_{\mathrm{c}}}^*(f_2^{'}) \cdot \\ & \mathrm{LK}(f_1, f_2, f_1+f_2-f)\, \mathrm{LK}^*\left(f_1+f_2-f_2^{'}, f_2^{'}, f_1+f_2-f\right) \end{aligned} \tag{67}$$

$$\begin{aligned} & \tau_{\mathrm{SCI}}^{m_{\mathrm{c}}}(f) = \tfrac{16}{81} R_{m_{\mathrm{c}}} \\ & \int_{f_{m_{\mathrm{c}}}-B_{m_{\mathrm{c}}}/2}^{f_{m_{\mathrm{c}}}+B_{m_{\mathrm{c}}}/2} df_1 \int_{f_{m_{\mathrm{c}}}-B_{m_{\mathrm{c}}}/2}^{f_{m_{\mathrm{c}}}+B_{m_{\mathrm{c}}}/2} df_2 \int_{f_{m_{\mathrm{c}}}-B_{m_{\mathrm{c}}}/2}^{f_{m_{\mathrm{c}}}+B_{m_{\mathrm{c}}}/2} df_1^{'} \int_{f_{m_{\mathrm{c}}}-B_{m_{\mathrm{c}}}/2}^{f_{m_{\mathrm{c}}}+B_{m_{\mathrm{c}}}/2} df_2^{'} \cdot \\ & \tilde{s}_{m_{\mathrm{c}}}(f_1) \tilde{s}_{m_{\mathrm{c}}}(f_2) \tilde{s}_{m_{\mathrm{c}}}^*(f_1+f_2-f) \tilde{s}_{m_{\mathrm{c}}}^*(f_1^{'}) \tilde{s}_{m_{\mathrm{c}}}^*(f_2^{'}) \tilde{s}_{m_{\mathrm{c}}}(f_1^{'}+f_2^{'}-f) \cdot \\ & \mathrm{LK}(f_1, f_2, f_1+f_2-f)\, \mathrm{LK}^*\left(f_1^{'}, f_2^{'}, f_1^{'}+f_2^{'}-f\right) \end{aligned} \tag{68}$$

where the integration limits were made explicit and correspond to the frequency interval occupied by the CUT, i.e.:

$$f \in [f_{m_{\mathrm{c}}} - B_{m_{\mathrm{c}}}/2, f_{m_{\mathrm{c}}} + B_{m_{\mathrm{c}}}/2] \tag{69}$$

The terms in Eq. (66) bearing the subscript 'X1' are EGN correction terms to the NLI produced by 'XPM' (Cross-Phase Modulation) or, according to a different taxonomy [50], due to 'XCI' (Cross-Channel Interference). Their expression is:

$$\begin{aligned} & \rho_{\mathrm{X1}}^{n_{\mathrm{ch}}}(f) = \tfrac{80}{81} R_{m_{\mathrm{c}}} R_{n_{\mathrm{ch}}} \\ & \int_{f_{m_{\mathrm{c}}}-B_{m_{\mathrm{c}}}/2}^{f_{m_{\mathrm{c}}}+B_{m_{\mathrm{c}}}/2} df_1 \int_{f_{n_{\mathrm{ch}}}-B_{n_{\mathrm{ch}}}/2}^{f_{n_{\mathrm{ch}}}+B_{n_{\mathrm{ch}}}/2} df_2 \int_{f_{n_{\mathrm{ch}}}-B_{n_{\mathrm{ch}}}/2}^{f_{n_{\mathrm{ch}}}+B_{n_{\mathrm{ch}}}/2} df_2^{'} \cdot \\ & |\tilde{s}_{m_{\mathrm{c}}}(f_1)|^2 \tilde{s}_{n_{\mathrm{ch}}}(f_2) \tilde{s}_{n_{\mathrm{ch}}}^*(f_2^{'}) \tilde{s}_{n_{\mathrm{ch}}}^*(f_1+f_2-f) \tilde{s}_{n_{\mathrm{ch}}}(f_1+f_2^{'}-f) \cdot \\ & \mathrm{LK}(f_1, f_2, f_1+f_2-f)\, \mathrm{LK}^*\left(f_1, f_2^{'}, f_1+f_2^{'}-f\right) \end{aligned} \tag{70}$$

where the integration limits were made explicit and correspond to either the frequency interval occupied by the CUT as written in Eq. (69), or by the generic $n_{\mathrm{ch}}$-th WDM channel:

$$f \in [f_{n_{\mathrm{ch}}} - B_{n_{\mathrm{ch}}}/2, f_{n_{\mathrm{ch}}} + B_{n_{\mathrm{ch}}}/2] \tag{71}$$

### *I. Gaussian-shaped constellations*

If a 'PM-Gaussian' constellation is used, then all the factors $\Phi$ and $\Psi$ in Eq. (66) are zero, according to Table I. This implies $G_{\mathrm{NLI}}^{\mathrm{corr}}(f) = 0$ or, equivalently, $G_{\mathrm{NLI}}^{\mathrm{EGN}}(f) = G_{\mathrm{NLI}}^{\mathrm{GN}}(f)$.

Given the current rapidly growing interest in Gaussian-shaped constellations (GSCs), this is an important result. To check it, we ran the sample test-set of Fig. 6a using an ideal GCS, taking as target MIs the values of GMI of the six QAM systems in Fig. 6a. The results are shown in Fig. 6b. Simulations agree very well with the GN/EGN curve. Therefore, if systems using GSCs earned widespread adoption, NLI modeling complexity would reduce to that of the GN-model, but EGN-model accuracy would be expected. This circumstance might help in the design and real-time handling of future physical-layer-aware networks based on GCSs. A specifically devoted paper was recently published on this topic, combining it with future ultra-high symbol rates [262]. As a caveat, GSCs are known to generate more NLPN than QAM constellations. For a discussion of the possible impact of this aspect on modeling, and for more information on how simulations involving GSCs can be performed, see [262].

### *J. GN-model closed-form approximate solutions*

In some cases, approximate closed-form solutions (ACFSs) can be found not just for the link factor, as shown in Sect. IV-F, but for the overall NLI PSD $G_{\mathrm{NLI}}(f)$, for either the GN or the EGN-model. Several ACFSs have been proposed, among which [50], [53], [56], [58], [60], [62], [65], [259]. Being approximate solutions, each one has specific limitations that must be taken into account.

For the EGN model, however, only one ACFS is currently available, consisting of an asymptotic form (in the number of spans) of the correction term $G_{\mathrm{NLI}}^{\mathrm{corr}}(f)$. It was proposed in [65] and then upgraded in [66], [259]. Though effective, it only works well for relatively long links.

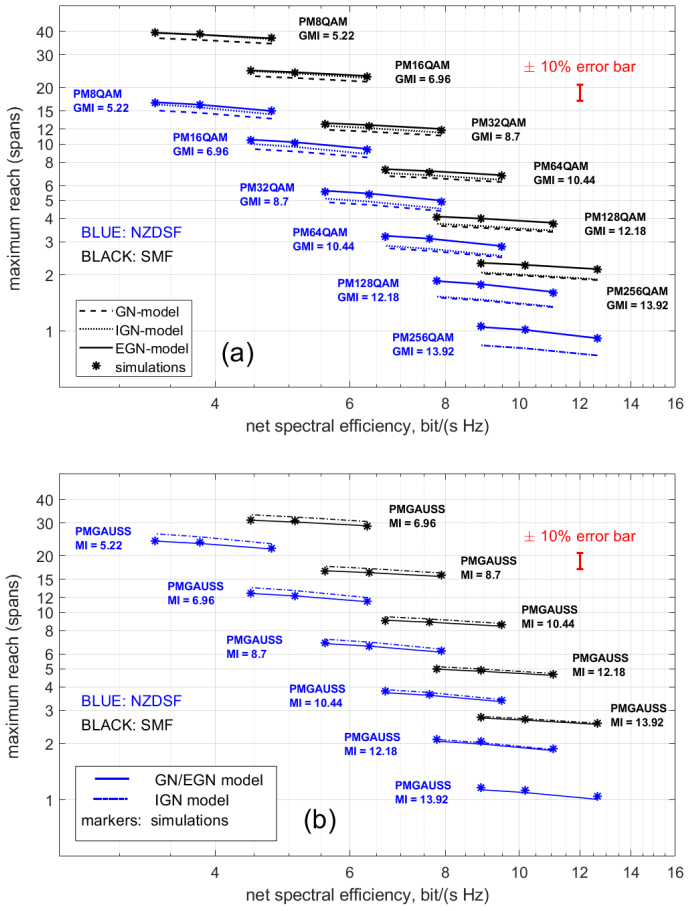


Fig. 6. System maximum reach (MR) in number of spans, vs. net spectral efficiency. Asterisks are simulation results, lines are analytical predictions using the GN-model, incoherent GN-model and EGN model (see legends). Results are in blue for NZDSF and in black for SMF. **System data:** The span length is 100km, EDFA noise figure 6 dB, symbol rate $R_s = 64$ GBaud, 15 channels, the tested channel is the center one. For each format the target GMI (for PM-QAM) or MI (for PM-Gaussian) is shown. The abscissa was calculated as: $\mathrm{GMI} \cdot R_s/\Delta f$, MI replacing GMI for PM-Gaussian. Three channel spacings $\Delta f$ were used for each format: 76.2, 87.5 and 100 GHz. The error bars indicate $\pm 10\%$ error interval. Fibers were: SMF, $D$=16.7 ps/(nm km), $\alpha$=0.21 dB/km, $\gamma$ =1.3 1/(W km); NZDSF, $D$=2.0 ps/(nm km), $\alpha$=0.22 dB/km, $\gamma$ =1.77 1/(W km).

Instead, several ACFSs have been found for the GN-model. A very accurate GN-model ACFS is available for transparent and uniform systems using an ideal Nyquist-WDM comb (zero roll-off and channel spacing equal to the symbol rate). It was originally derived in [50], and consists of Eqs. (7), (13) and (23) in that paper, combined.

An extension to non-Nyquist systems, for identical channels with uniform channel spacing, was also proposed in [50] (Eqs. (7),(15) and unnumbered formula after Eq. (23)). Its accuracy is quite good for narrow channel spacing but may degrade for large channel spacing.

An ACFS not requiring either transparency, uniformity or identical, equally-spaced channels, and therefore of great potential usefulness, was reported in [62]. Such formula is quite general, but it approximates the iGN-model, rather than GN, i.e., it assumes *incoherent NLI accumulation* (see Sect. IV-G). On the other hand, the results from the literature (see for instance [260], [261], [262]) as well as Fig. 6, show the iGN-model to be close to the GN-model in most practical situations, including that of GSC-based systems (Fig. 6b). Two more assumptions are made, which were needed in the derivation: the channels have approximately rectangular PSD (that is, very small roll-off) and span loss is at least 7 dB, with greater than 10 dB being the optimal condition. Also, this ACFS provides an estimate of $G_{\mathrm{NLI}}(f)$ at the *center* of any one of the WDM channels. To the purpose of system performance assessment, it is then necessary to assume that $G_{\mathrm{NLI}}(f)$ be flat over the channel of interest. This 'local white noise' assumption was studied in [50] and [62] and was shown to typically induce a small error. The formula is:

$$G_{\mathrm{NLI}}(f_{i_{\mathrm{ch}}}) = \tfrac{16}{27} \sum_{n_s=1}^{N_s} \gamma_{n_s}^2 L_{\mathrm{eff},n_s}^2 \cdot \prod_{p=1}^{n_s-1} \Gamma_p^3 e^{-6\alpha_p L_s^{(p)}} \cdot \prod_{p=n_s}^{N_s} \Gamma_p e^{-2\alpha_p L_s^{(p)}} \cdot \sum_{n_{\mathrm{ch}}=1}^{N_{\mathrm{ch}}} G_{\mathrm{S}}^2\left(f_{n_{\mathrm{ch}}}\right) G_{\mathrm{S}}\left(f_{i_{\mathrm{ch}}}\right) \cdot \left(2 - \delta_{n_{\mathrm{ch}} i_{\mathrm{ch}}}\right) \cdot \Theta_{n_{\mathrm{ch}} i_{\mathrm{ch}} n_s} \tag{72}$$

where: $G_{\mathrm{NLI}}(f_{i_{\mathrm{ch}}})$ is the NLI PSD at the center frequency $f_{i_{\mathrm{ch}}}$ of the $i_{\mathrm{ch}}$-th channel of the comb; $G_{\mathrm{S}}\left(f_{n_{\mathrm{ch}}}\right)$ and $G_{\mathrm{S}}\left(f_{i_{\mathrm{ch}}}\right)$ are the PSD's of the WDM data signal at the center frequency of the $n_{\mathrm{ch}}$-th and $i_{\mathrm{ch}}$-th channels, respectively; $\delta_{n_{\mathrm{ch}} i_{\mathrm{ch}}}$ is a Kronecker's delta, i.e., it is one if $n_{\mathrm{ch}}$=$i_{\mathrm{ch}}$ and zero otherwise; finally, $\Theta$ is:

$$\Theta_{n_{\mathrm{ch}} i_{\mathrm{ch}} n_s} \approx \frac{\operatorname{asinh}\left(\pi^2 [2\alpha_{n_s}]^{-1} |\beta_{2,n_s}| \left[f_{n_{\mathrm{ch}}} - f_{i_{\mathrm{ch}}} + B_{n_{\mathrm{ch}}}/2\right] B_{i_{\mathrm{ch}}}\right)}{4\pi (2\alpha_{n_s})^{-1} |\beta_{2,n_s}|} - \frac{\operatorname{asinh}\left(\pi^2 [2\alpha_{n_s}]^{-1} |\beta_{2,n_s}| \left[f_{n_{\mathrm{ch}}} - f_{i_{\mathrm{ch}}} - B_{n_{\mathrm{ch}}}/2\right] B_{i_{\mathrm{ch}}}\right)}{4\pi (2\alpha_{n_s})^{-1} |\beta_{2,n_s}|}, \quad n_{\mathrm{ch}} \neq i_{\mathrm{ch}} \tag{73}$$

$$\Theta_{i_{\mathrm{ch}} i_{\mathrm{ch}} n_s} \approx \frac{\operatorname{asinh}\left(\frac{\pi^2}{2} |\beta_{2,n_s}| [2\alpha_{n_s}]^{-1} B_{i_{\mathrm{ch}}}^2\right)}{2\pi |\beta_{2,n_s}| [2\alpha_{n_s}]^{-1}}, \quad n_{\mathrm{ch}} = i_{\mathrm{ch}} \tag{74}$$

where $B_{n_{\mathrm{ch}}}$ and $B_{i_{\mathrm{ch}}}$ are the bandwidth of the $n_{\mathrm{ch}}$-th and $i_{\mathrm{ch}}$-th channels, respectively. Note that the formula can be easily upgraded to support lumped frequency-dependent gain or loss, as well as allow to account for the drop-off and join-in of channels at any of the span starts.

Regarding accuracy, the typical absolute error between the predictions of Eqs. (72)–(74) vs. numerical integration of the iGN-model is typically 2%-3% of MR, making it a valuable tool for real-time system appraisal.

While this chapter was being finalized, two GN/iGN model ACFS have been proposed, which upgrade Eqs. (72)–(74) to also support dispersion slope through $\beta_3$, frequency-dependent loss and inter-channel stimulated Raman scattering [263], [264]. They represent substantial progress since they make it possible to analyze ultra-broadband systems, such as (C+L)-band ones. Due to their closed-form, they could potentially become effective tools for real-time physical-layer-aware management of fully-loaded reconfigurable optical networks.

*1) Closed-form NLI modeling and Raman amplification:* In previous sections, the GN and EGN general expressions Eqs. (44)-(46), were provided in such a form that they can support arbitrary distributed amplification, also frequency-dependent.

One overall ACFS is currently available for the GN-model assuming *ideal distributed* amplification, i.e., with $\alpha(z) = g(z)$ at each point along the link, and ideal Nyquist-WDM transmission. It is Eq. (24) in [50] but of course this represents a completely ideal reference scenario.

An exact closed-form was provided for the LK factor in Sect. IV-F, under simplifying assumptions (undepleted backward-propagating pump, frequency-flat gain). By means of various simplifications and approximations of the LK term and of the integration procedure, an ACFS for frequency-independent Raman amplification was proposed in [265] and extensively tested with good results. In [266], [267] the LK factor was approximated in suitable ways such that subsequent simplified numerical integration was possible, even in the most general case of depleted-pump and frequency-dependent Raman amplification.

It should also be pointed out that, as shown extensively in [260], provided that Raman amplification is backward-pumped and provides gain which is at least 6 dB lower than fiber loss, then its effect on NLI generation is small. In practice, it is possible to neglect it incurring only about 2%-3% MR estimation error. This result is significant, since in many practical cases Raman amplification complies with the above condition, as part of the span loss is compensated for by Raman amplification and part by an EDFA placed at the end of the span. This is called *hybrid Raman-EDFA amplification* and currently represents the typical solution employed in terrestrial (new or refurbished) links. In such cases, the effect of Raman on NLI generation can essentially be ignored.

## V. Time-domain perturbative model

In this section we explore a perturbation-based *time-domain* model [26], describing the way in which data symbols transmitted into the optical fiber are perturbed by the nonlinearity of the fiber. This approach is alternative to the frequency-domain FWM approach described in the previous section. While it leads to similar results as the EGN model, it also permits to predict statistical correlations between different temporal NLI symbols and to discern among different types of NLI. In particular, it allows to single-out and characterize its so-called PPRN components (Phase and Polarization Rotation Noise).

In what follows we ignore nonlinear perturbations generated by the presence of co-propagating ASE noise, and focus on SPM perturbations caused by the symbols transmitted over the channel of interest as well as on XPM and FWM perturbations caused by the data symbols transmitted over neighboring interfering WDM channels.

In the framework of first-order perturbation analysis, the received received symbols after ideal dispersion compensation can be written as

$$\underline{r}_n = \underline{a}_n + \Delta\underline{a}_n, \tag{75}$$

where the two-element column vectors $\underline{a}_n$ and $\underline{r}_n$ represent the transmitted and received polarization multiplexed data symbols of the channel of interest in the $n$-th time slot. The two-element data-vector $\Delta\underline{a}_n$ corresponds to the first-order nonlinear perturbation. The various signal-induced contributions to $\Delta\underline{a}_n$ can be categorized as resulting from SPM, XPM and FWM. Their time-domain representation is given by [26], [57]

$$\Delta\underline{a}_n^{\mathrm{SPM}} = \sum_{l,k,m} \underline{a}_{n+l}\underline{a}_{n+k}^{\dagger}\underline{a}_{n+m} S_{l,k,m} \ , \tag{76}$$

$$\Delta\underline{a}_n^{\mathrm{XPM}} = \sum_{l,k,m,j} \left(\underline{b}_{n+k,j}^{\dagger}\underline{b}_{n+m,j}\mathbf{I} + \underline{b}_{n+m,j}\underline{b}_{n+k,j}^{\dagger}\right) \times \underline{a}_{n+l} X_{l,k,m}(\Omega_j), \tag{77}$$

$$\Delta\underline{a}_n^{\mathrm{FWM}} = \sum_{\substack{l,k,m \\ j_1,j_2,j_3}} \underline{b}_{n+l,j_1}\underline{b}_{n+k,j_2}^{\dagger}\underline{b}_{n+m,j_1} F_{l,k,m}(\Omega_{j_1},\Omega_{j_2},\Omega_{j_3}) \ , \tag{78}$$

where $\underline{b}_{n,j}$ represents the two-element data-vector transmitted in the $n$-th time-slot over the $j$-th interfering WDM channel, having a frequency separation of $\Omega_j$ from the channel of interest. The FWM kernels $F_{l,k,m}(\Omega_{j_1},\Omega_{j_2},\Omega_{j_3})$ satisfy

$$F_{l,k,m}(\Omega_{j_1},\Omega_{j_2},\Omega_{j_3}) = i\gamma\frac{8}{9}\int\!\!\int_0^L f(z)s_0^*(z,t) \times s_{\Omega_{j_1}}(z,t-lT)s_{\Omega_{j_2}}^*(z,t-kT)s_{\Omega_{j_3}}(z,t-mT)\mathrm{d}z\mathrm{d}t, \tag{79}$$

where $\gamma$ is the nonlinear coefficient of the fiber, $L$ and $f(z)$ are the length and power profile of the link, $T$ is the symbol time duration, and where $s_\Omega(z,t)$ represents the dispersed waveform of the pulse transmitted over a WDM channel spaced by $\Omega$ from the channel of interest, when reaching point $z$ along the fiber. The SPM and XPM kernels are given by $S_{l,k,m} = F_{l,k,m}(0,0,0)$ and $X_{l,k,m}(\Omega) = F_{l,k,m}(0,\Omega,\Omega)$, respectively. We note that in cases where the channel spacing is sufficiently low there are additional XPM contributions that do not appear in Eq. (77) and that involve interactions between three data-vectors from the closest interfering channel or interactions between two data-vectors from the channel of interest and a single data-vector from the closest interfering channels.

*1) XPM time-varying ISI representation:* The effect of XPM on the received data symbols of the channel of interest can be described as time-varying ISI [51], [3]. This can be viewed by rewriting Eq. 77 as

$$\Delta\underline{a}_n^{\mathrm{XPM}} = \sum_h \mathbf{H}_l^{(n)}\underline{a}_{n+l}, \tag{80}$$

where the $2\times 2$ ISI matrices are given by

$$\mathbf{H}_l^{(n)} = \sum_{k,m,j} \left(\underline{b}_{k+n,j}^{\dagger}\underline{b}_{m+n,j}\mathbf{I} + \underline{b}_{m+n,j}\underline{b}_{k+n,j}^{\dagger}\right) X_{l,k,m}(\Omega_j) \tag{81}$$

Assuming that the data symbols transmitted over the interfering channels are unknown at the transmitter and receiver implies that the ISI matrices $\mathbf{H}_l^{(n)}$ are unknown as well. The summation over $k$ and $m$ in Eq. (81) implies that the set of data vectors $\underline{b}_{n,j}$ contributing to $\mathbf{H}_l^{(n)}$ changes with $n$ and hence the ISI matrices vary with time, as indicated by the superscript $(n)$. In highly dispersive systems a large number of data vectors $\underline{b}_{n,j}$ may participate in the summation of Eq. (81), implying that the ISI matrices $\mathbf{H}_l^{(n)}$ change slowly with $n$. The temporal correlation of these ISI matrices is further discussed in Sec. V-4.

*2) Nonlinear phase and polarization rotation:* The zeroth-order XPM contribution $\mathbf{H}_0^{(n)}\underline{a}_n$ possesses some interesting and unique properties that have important implications on the statistical properties of NLIN and on its mitigation [26], [57], [92]. As shown in Ref. [92] and in the Appendix of Ref. [4], its effect can be written as

$$\underline{a}_n + \mathbf{H}_0^{(n)}\underline{a}_n = \begin{pmatrix} e^{i\theta_n^{(x)}} & ih_n \\ ih_n^* & e^{i\theta_n^{(y)}} \end{pmatrix} \begin{pmatrix} a_n^{(x)} \\ a_n^{(y)} \end{pmatrix}, \tag{82}$$

where $a_n^{(x)}$ and $a_n^{(y)}$ are the two elements of the data vector $\underline{a}_n$ and where $\theta_n^{(x)}$, $\theta_n^{(y)}$ and $h_n$ are given by

$$\theta_n^{(x)} = \gamma\frac{8}{9}\sum_{k,m,j}\left(2b_{n+k,j}^{(x)*}b_{n+m,j}^{(x)} + b_{n+k,j}^{(y)*}b_{n+m,j}^{(y)}\right)X_{0,k,m}(\Omega_j) \tag{83}$$

$$\theta_n^{(y)} = \gamma\frac{8}{9}\sum_{k,m,j}\left(2b_{n+k,j}^{(y)*}b_{n+m,j}^{(y)} + b_{n+k,j}^{(x)*}b_{n+m,j}^{(x)}\right)X_{0,k,m}(\Omega_j) \tag{84}$$

$$h_n = \gamma\frac{8}{9}\sum_{k,m,j} b_{n+k,j}^{(y)*}b_{n+m,j}^{(x)}X_{0,k,m}(\Omega_j), \tag{85}$$

with $b_{n,j}^{(x)}$ and $b_{n,j}^{(y)}$ representing the two elements of $\underline{b}_{n,j}$.

The zeroth-order XPM contribution can therefore be viewed as inducing independent phase-noise in each polarization as well as polarization crosstalk, which is known also as cross polarization modulation (XpolM). Another important observation that arises from Eqs. (82)-(84) is that nonlinear phase-noise has strong dependence on the modulation format [26], [57]. This dependence results from the fact that the terms with $k = m$ in Eqs. (83) and (84) are proportional to $|b_{n+m,j}^{(x)}|^2$ and $|b_{n+m,j}^{(y)}|^2$. For constant-modulus formats these terms are fixed for any $n$, implying that they induce a constant phase-shift for all received data-symbols (i.e., only rotating the entire received constellation). On the other hand, when amplitude modulation is introduced to the data-symbols, $|b_{n+m,j}^{(x)}|^2$ and $|b_{n+m,j}^{(y)}|^2$ change with $n$, and $\theta_n^{(x)}$ and $\theta_n^{(y)}$ vary with time. This is why nonlinear phase-noise is relatively small in systems using QPSK but significantly larger in systems employing higher-order QAM formats [57]. Note however, that even for constant-modulus formats, polarization crosstalk (through $ih_n$) can be very important.

The effect of the zeroth-order XPM contribution on the dual-polarization signal can be also described as phase and polarization rotation noise. Using the power series expansion of matrix exponentials [5], [6], Eqs. (82)-(85) can be written as [92]

$$\underline{a}_n + \mathbf{H}_0^{(n)}\underline{a}_n = e^{i\varphi_n}e^{i\mathbf{\Phi}_n}\underline{a}_n, \tag{86}$$

where the term $\exp(i\varphi_n)$ induces phase-noise and is given by

$$\varphi_n = \gamma\frac{4}{3}\sum_{k,m,j}\underline{b}_{n+k,j}^{\dagger}\underline{b}_{n+m,j}X_{0,k,m}(\Omega_j), \tag{87}$$

and where the matrix exponential $\exp(i\mathbf{\Phi}_n)$ induces polarization rotation noise and is given by

$$\mathbf{\Phi}_n = \gamma\frac{8}{9}\sum_{k,m,j}\left(\underline{b}_{n+m,j}\underline{b}_{n+k,j}^{\dagger} - \frac{1}{2}\underline{b}_{n+k,j}^{\dagger}\underline{b}_{n+m,j}\mathbf{I}\right)X_{0,k,m}(\Omega_j). \tag{88}$$

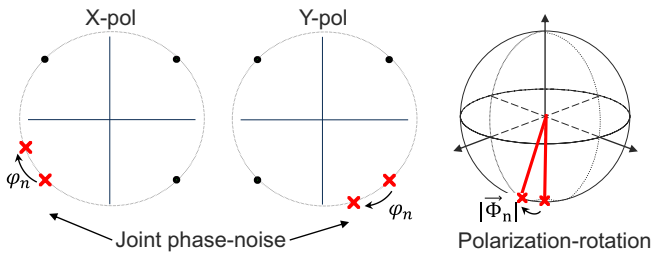


Fig. 7. Illustration of phase and polarization-rotation noise (PPRN) in QPSK transmission.

Figure 7 visualizes the two contributions of the zeroth-order XPM effect. The phase-noise part $e^{i\varphi_n}$ rotates the two polarizations together by a common angle $\varphi_n$ equal to the average of $\theta_n^{(x)}$ and $\theta_n^{(y)}$, whereas the polarization-state rotation $\exp(i\mathbf{\Phi}_n)$ causes the Stokes vector representing $\underline{a}_n$ on the Poincaré sphere [6] to rotate about $\vec{\Phi}_n$, the Stokes vector of $\mathbf{\Phi}_n$, at an angle equal to its length $|\vec{\Phi}_n|$. This motivates the term *Phase and Polarization-Rotation Noise* (PPRN) to denote the zeroth-order XPM contribution. Furthermore, the PPRN description captures the unitarity of the zeroth-order XPM effect which does not impair the norm of the transmitted dual-polarization data-vector $\underline{a}_n^{\dagger}\underline{a}_n$.

*3) Variance of PPRN and higher-order XPM terms:* In what follows we examine the relative importance of the various XPM contributions by examining their individual contribution to the overall XPM variance. The predictions are based on the calculations of Ref. [7] (also shown in the Appendix of Ref. [4]) which were shown to be in excellent agreement with split-step simulations. The calculations assume polarization-multiplexed transmission of statistically independent data symbols, isotropically symmetric in their phase space and are given here, for simplicity of notation, for perfect Nyquist pulses with zero roll-off. The dependence of XPM on the roll-off factor is typically small but may become significant for roll-off factors larger than ~0.2. We refer the interested reader to Appendix B of Ref. [7] for further information on how to modify the formulas below to account for non-zero roll-off factors.

The variance of XPM contributions can be written as $\mathrm{P}_0\mathrm{P}_{\mathrm{int}}^2\chi_{\mathrm{XPM}}$ where $\mathrm{P}_0$ and $\mathrm{P}_{\mathrm{int}}$ are the average launch power of the channel of interest and the interfering channel. Following the assumptions above, the various XPM terms are uncorrelated and the XPM nonlinear coefficient, $\chi_{\mathrm{XPM}}$, can be written as [7]

$$\begin{aligned}\chi_{\mathrm{XPM}} &= \sum_l \chi_{\mathrm{XPM}}^{(l)} \\ &= \sum_l \chi_{\mathrm{XPM},1}^{(l)} + \left(\frac{\langle|b|^4\rangle}{\langle|b|^2\rangle^2} - 2\right)\chi_{\mathrm{XPM},2}^{(l)}\end{aligned} \tag{89}$$

where $b$ represents a single data symbol transmitted over one of the polarizations of the interfering channel and where the angled brackets denote statistical averaging. The term $\langle|b|^4\rangle/\langle|b|^2\rangle^2$ accounts for the dependence of XPM on the fourth-order moment of the interfering data symbols.

The contribution of the $l$-th XPM term $\mathbf{H}_l^{(n)}\underline{a}_{n+l}$ to the overall XPM variance is therefore given by $P_0P_{\mathrm{INT}}^2\chi_{\mathrm{XPM}}^{(l)}$

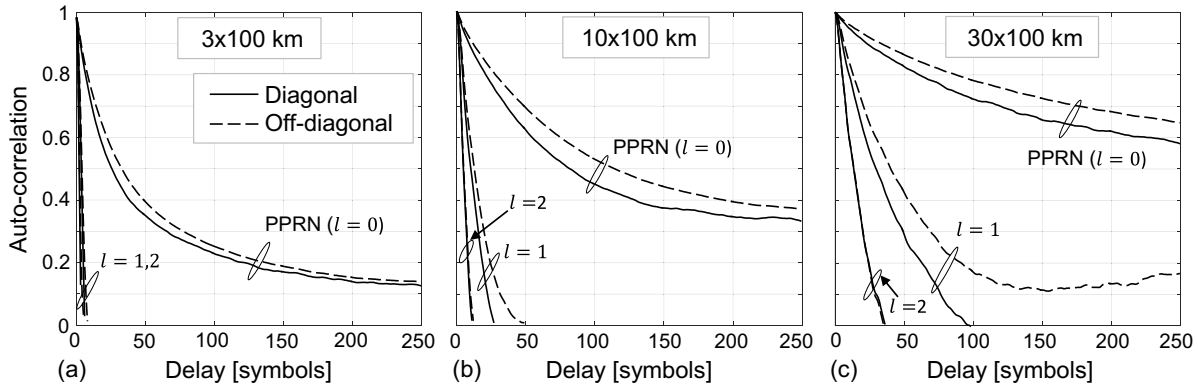


Fig. 8. Auto-correlation of the diagonal (solid) and off-diagonal (dashed) elements of $\mathbf{H}_l^{(n)}$ for $l = 0, 1, 2$ in 16-QAM fully loaded systems with 115 channels. The diagonal and off-diagonal elements of $\mathbf{H}_0^{(n)}$ induce nonlinear phase-noise and polarization crosstalk.

where the coefficients $\chi_{\mathrm{XPM},1}^{(l)}$ and $\chi_{\mathrm{XPM},2}^{(l)}$ are given by [7]

$$\chi_{\mathrm{XPM},1}^{(l)} = \frac{32}{27}\frac{\gamma^2}{T^2}\sum_j \int \rho_j(\omega_1,\omega_2,\omega_3)\rho_j^*(\omega_4,\omega_2,\omega_3)e^{i(\omega_1-\omega_4)lT}\frac{\mathrm{d}^4\omega}{(2\pi)^4},$$
$$\chi_{\mathrm{XPM},2}^{(l)} = \frac{80}{81}\frac{\gamma^2}{T}\sum_j \int \rho_j(\omega_1,\omega_2,\omega_3)\rho_j^*(\omega_4,\omega_5,\omega_6)e^{i(\omega_1-\omega_4)lT}\frac{\mathrm{d}^5\omega}{(2\pi)^5}, \quad (90)$$

with $\omega_6 = \omega_5 - \omega_2 + \omega_3$. The kernels $\rho_j(\omega_1,\omega_2,\omega_3)$ are given by

$$\rho_j(\omega_1,\omega_2,\omega_3) = \tilde{s}_0^*(\omega_1-\omega_2+\omega_3)\tilde{s}_0(\omega_1)\tilde{s}_j^*(\omega_2)\tilde{s}_j(\omega_3) \times \int_0^L f(z)e^{i(\omega_1-\omega_2-\Omega_j)(\omega_2-\omega_3)\beta''z}\mathrm{d}z, \quad (91)$$

where $\beta''$ is the dispersion coefficient of the fiber and where $\tilde{s}_0(\omega)$ and $\tilde{s}_j(\omega)$ represent the baseband spectral shape of the transmitted pulses from the channel of interest and the $j$-th interfering channel.

In Fig. 9 we examine the significance of the various XPM contributions in standard 16-QAM WDM systems with 100-km spans, 115 32-Gbaud channels and 37.5-GHz channel spacing. We plot the variance of the various XPM contributions $\mathbf{H}_l^{(n)}\underline{a}_{n+l}$ normalized by the variance of the zeroth-order XPM contribution $\mathbf{H}_0^{(n)}\underline{a}_n$. Evidently, the PPRN contribution has the most pronounced effect whereas the significance of the XPM terms $\mathbf{H}_l^{(n)}\underline{a}_{n+l}$ with $|l| > 0$ decreases monotonically with $|l|$.

*4) Temporal correlations:* An important observation regarding the ISI form of XPM is the fact that the ISI matrices $\mathbf{H}_l^{(n)}$ change slowly with $n$, as the matrix elements are the result of a summation over a significant number of interfering symbols, see Eq. 81. Following the assumptions of Sec. V-3, the temporal auto-correlation function of the diagonal elements of $\mathbf{H}_l^{(n)}$ is given by [7]

$$\mathrm{R}_{\mathrm{diag}}^{(l)}(s) = R_{\mathrm{diag},1}^{(l)}(s) + \left(\frac{\langle|b|^4\rangle}{\langle|b|^2\rangle^2} - 2\right)R_{\mathrm{diag},2}^{(l)}(s), \quad (92)$$

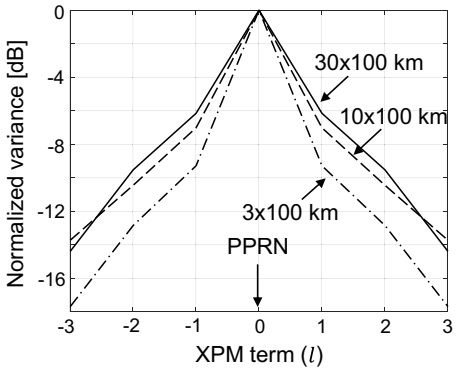


Fig. 9. Variance of the various XPM terms normalized by the variance of the zeroth-order XPM term (PPRN) for fully loaded 115-channel transmission using 16-QAM.

where

$$R_{\mathrm{diag},1}^{(l)}(s) = \frac{80}{81}\frac{\gamma^2}{T^2}\sum_j \int \rho_j(\omega_1,\omega_2,\omega_3)\rho_j^*(\omega_4,\omega_2,\omega_3) \times e^{i(\omega_1-\omega_4)lT}e^{i(\omega_2-\omega_3)sT}\frac{\mathrm{d}^4\omega}{(2\pi)^4},$$
$$R_{\mathrm{diag},2}^{(l)}(s) = \frac{80}{81}\frac{\gamma^2}{T}\sum_j \int \rho_j(\omega_1,\omega_2,\omega_3)\rho_j^*(\omega_4,\omega_5,\omega_6) \times e^{i(\omega_1-\omega_4)lT}e^{i(\omega_2-\omega_3)sT}\frac{\mathrm{d}^5\omega}{(2\pi)^5}, \quad (93)$$

with $\omega_6 = \omega_5 - \omega_2 + \omega_3$. The temporal auto-correlation function of the off-diagonal elements of $\mathbf{H}_l^{(n)}$ is given by [7]

$$\mathrm{R}_{\mathrm{off-diag}}^{(l)}(s) = \frac{1}{5}R_{\mathrm{diag},1}^{(l)}(s). \quad (94)$$

In Fig. 8 we show the auto-correlation function of the diagonal and off-diagonal elements of $\mathbf{H}_l^{(n)}$ for $l = 0, 1, 2$, considering 16-QAM transmission, 100-km spans, and standard fully loaded systems with 115 32-Gbaud channels and 37.5-GHz channel spacing. The correlations of the higher-order XPM terms quickly drop to zero, even in systems operating over a 3,000-km link. The correlations of nonlinear PPRN ($l = 0$), however, are relatively long, on the order

of tens of symbols. These results are aligned with numerical simulations [67], [8], [9], [10] and experimental measurements [11], [12] verifying the long temporal correlations of the phase-noise contribution of PPRN (diagonal elements with $l = 0$), as well as with the numerical [8], [9] and experimental [13], [14] verification of the long temporal correlations of the polarization crosstalk contribution of PPRN (off-diagonal elements with $l = 0$).

Regarding the specific system impact of PPRN and NLPN, see Sect. IX-B.

*5) Pulse collision theory:* The time-domain model further provides insights into how the various NLI contributions are formed. By examining the optical field of the transmitted signal as a collection of temporal pulses transmitted from multiple WDM channels, one can characterize and analyze the different nonlinear interactions between the transmitted pulses as they propagate through the optical fiber.

In particular, in the limits of first order perturbation analysis, the interactions can be classified as collisions between two pulses, three pulses, and four pulses; these collisions can be either complete, or incomplete. A rigorous and comprehensive analysis of the various types of collisions is provided in [92]. Each type of collision is shown to have its unique signature and the overall nature of NLI is determined by the relative significance of the various collisions in a given WDM transmission. The most important contributions to NLI are shown to follow from two-pulse and four-pulse collisions. Two-pulse collisions generate NLI in the form of PPRN whereas four-pulse collisions generate complex circular ISI noise. In addition, two-pulse collisions are shown to have strong dependence on the modulation format of the interfering WDM channels and to be most pronounced when the collision is complete, whereas four-pulse collisions are shown to be modulation format independent and to be strongest when the collision is incomplete. The theory further shows that in short optical links with lumped amplification, and links with perfect distributed amplification, NLI tends to be dominated by PPRN contributions generated by two-pulse collisions; in long links with lumped amplification, the importance of four-pulse collisions increases and NLI becomes more complex circular with reduced dependence on modulation format.

## VI. Spatially-Resolved Perturbative Models and Their Applications

The perturbative method discussed in Section III-A and IV shows the perturbation as a double integral in the frequency domain, weighted by the kernel function. Strictly speaking, each frequency of the signal is perturbed by all possible FWM combinations, and the kernel represents their efficiency.

This picture is a global input/output description of the whole optical link and it is of great interest for the system perspective as shown in several sections of this chapter. Nevertheless, it is of interest to relate such a global scale description to a local scale at a given coordinate. Such a relation highlights several physical connections that may help to understand the perturbative models and to search for novel applications of perturbation theory in nonlinear optical communications.

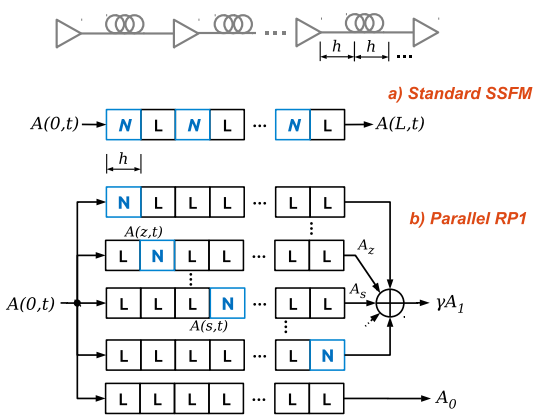


Fig. 10. Comparison between SSFM and first order perturbative solution of NLSE.

The local description of the perturbation is the dynamical model expressed by the differential equation (7). We find particularly interesting solving it by following the numerical split-step idea, as much as we did for the NLSE with the SSFM. The corresponding block diagram is shown in Fig. 10 [133], [22], [91]. For the sake of comparison, we also show the SSFM method in the same figure.

The RP1 solution is the sum of the unperturbed and the first-order perturbation. The unperturbed term, by definition, has been discretized by the concatenation of linear blocks only, here indicated by the letter L. Each linear block thus summarizes linear effects such as attenuation, GVD, etc, into an infinitesimal step $h$.

The perturbation is indeed the result of many contributions, each represented by a branch in the block diagram. In such paths we find nonlinear blocks which implement the input/output relation $\mathrm{N}(A) = -j\gamma |A|^2 Ah$. We note that, since the received perturbation is linear in $\gamma$, the electric field, according to RP1, while traveling along the link can cross only one nonlinear block from input to output. However, such a nonlinearity can appear anywhere, hence the reason for the final sum creating the perturbative term as a sort of multi-path interference.

The idea can be iterated, creating higher order RP schemes [134]. For instance, RP2 can be emulated by considering all possible combinations of two nonlinear blocks within the discretized distance grid. The SSFM algorithm can be viewed as an RP$\infty$ method implemented in a clever way where the nonlinear block is substituted by the SPM operator $N(A) = e^{-j\gamma|A|^2 h}$.

In the figure, we referred to RP1 as parallel RP1 because of its parallel nature, which is not the case of SSFM whose serial structure is unavoidable. Besides the physical intuition, such a parallel structure might find some advantages in numerical implementation when a high degree of parallelization is available [132].

We can exploit the RP1 diagram to derive a spatially-resolved GN model [58], [91]. Our target is the variance of the output perturbation $\gamma A_1$, and more generally its auto-correlation function. Since the output perturbation is the sum of many contributions, see Fig. 10, to get our target we need

the cross-correlation among any two generic paths forming the perturbation [154], [58]. Let us focus on two such paths, the ones where the nonlinearity appears after $s$ and $z$ km, respectively. See Fig. 10 for reference and for the main definitions of variables.

The common input to such paths is a linear digital modulated signal $A(z=0,t)$, with $t$ time. In case of interest, such a signal is actually a cyclo-stationary stochastic process, i.e., its statistical properties vary cyclically in time with a period equal to the symbol time. However, we believe that the signal can be safely treated as stationary. This claim finds its ground in the observation that we are planning to work with dispersion uncompensated links where the strong interference induced by dispersion likely removes higher order cyclo-stationary effects and because we mainly focus on sinc-like pulses such that the signal is almost stationary already at the fiber input.

In this framework, if we know the cross-correlation function of the fields outgoing the nonlinear blocks, here called $R_{zs}^{(\text{out})}(\tau)$, with $\tau$ the time lag, the cross-correlation of the received fields on such paths can be found by using basic linear system theory [135]. In detail, the cross-correlation between $A_z$ and $A_s$ at the output coordinate $L$ is related to $R_{zs}^{(\text{out})}(\tau)$ by:

$$\mathbb{E}\left[A_z(t+\tau)A_s^*(t)\right] = h_{zL}(\tau) \otimes h_{sL}^*(-\tau) \otimes R_{zs}^{(\text{out})}(\tau)$$

with $\otimes$ denoting convolution and $h_{zL}(\tau)$ the impulse response of the filter accounting for all linear effects from coordinate $z$ to output coordinate $L$. Such expression can be efficiently evaluated in frequency domain by working with power spectral densities.

With similar arguments, we can relate the auto-correlation function of the transmitted field $A(0,t)$ to the fields ingoing the nonlinear blocks at coordinates $z$ and $s$ , respectively:

$$R_{zs}^{(\text{in})} \triangleq \mathbb{E}\left[A(z,t+\tau)A^*(s,t)\right] = h_{0z}(\tau) \otimes h_{0s}^*(-\tau) \otimes R_{00}(\tau) \tag{95}$$

with $R_{00}(\tau) = \mathbb{E}\left[A(0,t+\tau)A^*(0,t)\right]$. If we are able to relate $R_{zs}^{(\text{out})}(\tau)$ to $R_{zs}^{(\text{in})}(\tau)$ we have therefore a relation between the system input/output auto-correlation functions and thus our target can be definitely solved by summing, i.e., integrating, all such contributions.

Such a relation is possible in the special, yet relevant, case of $A(z,t)$ with Gaussian statistics at any coordinate [116]. If we get rid of the constant phase shift term as in the eRP (see Section III-A), the relation takes a simple and elegant form [58]:

$$R_{zs}^{(\text{out})}(\tau) = 2\left|R_{zs}^{(\text{in})}(\tau)\right|^2 R_{zs}^{(\text{in})}(\tau) \ . \tag{96}$$

Such assumptions are exactly those of the GN model, whose spatially-resolved reference formula is thus the following:

$$\tilde{S}_{\text{GN}}(f) = \int_0^L \int_0^L \tilde{h}_{zL}(f)\tilde{h}_{sL}^*(f) \times \mathcal{F}\left\{R_{zs}^{(\text{out})}(\tau)\right\} \mathrm{d}z\mathrm{d}s \tag{97}$$

where $\tilde{S}_{\text{GN}}(f)$ is the PSD of the received perturbation, i.e., the Fourier transform of the auto-correlation function. The variance of the nonlinear interference given by the GN model is simply the integral of this PSD.

Such a solution has several similarities with the basic RP idea adopted for the electric field. Strictly speaking, we are propagating the signal correlation function from input to output by performing first a linear operation according to (95), then the same first-order nonlinear perturbation identical to the one experienced by the electric field except for a factor 2, then again linear effects up to the output. The main difference is that while the field perturbation is the sum of many perturbative infinitesimal contributions, here we have to account for all possible pairs of cross-correlations, thus increasing the complexity into a double integration. However, it has been shown that the double integral can be reduced to a single integration by exploiting the exponential behavior of attenuation along distance [58].

What about in the general case of non-Gaussian statistics for the propagating signal? The previous idea can still be used by properly modifying (96). Now the cross-correlation of the signals outgoing the nonlinear blocks cannot be related just to the same cross-correlation between the inputs, but we have to account for higher order statistics because of the nonlinear transformation [26]. Since nonlinearity is cubic, the cross-correlation involves products of six random variables. Such random variables are the digital symbols of the modulation format under use. Among all possible combinations, many of them have zero average for classical modulation formats with uniform distribution and rotationally symmetric constellations. If we let $a_k$ be the information symbol at time $k$, we are left with just three kinds of non-zero contributions to the cross-correlation: one involving $\left\{|a_k|^2 |a_n|^2 |a_l|^2\right\}$ with $k \neq n \neq l$, one with $\left\{|a_k|^4 |a_n|^2\right\}$ with $k \neq n$, and one for $\left\{|a_k|^6\right\}$. The first partition is the biggest one and generates terms like the one in (96) since indexes are different and symbols independent. Hence, such a partition is the one accounted by the GN model. The other two partitions generate the higher order contributions of the EGN model [63].

In summary, the cross-correlation $R_{zs}^{(\text{out})}(\tau)$ can be related to the shape of the supporting pulse distorted by linear effects up to coordinate $z$, $p(z,t) \triangleq h_{0z}(t) \otimes p(0,t)$, by [91]:

$$R_{zs}^{(\text{out})}(\tau) = \underbrace{\frac{\mu_2^3}{T^3} 2\left|Q(z,s,\tau)\right|^2 Q(z,s,\tau)}_{\text{GN model}} + \frac{\mu_2\kappa_{2;2}}{T^2}\left\{4F_4(z,s,\tau)+Q_4(z,s,\tau)\right\}Q(z,s,\tau)+\frac{\kappa_{3;3}}{T}Q_6(z,s,\tau) \tag{98}$$

with $\kappa_{i;i}$ the $i$-th cumulant of transmission symbol [91, Appendix B], $\mu_2 \triangleq \mathbb{E}\left[|a_k|^2\right]$, and:

$$\begin{aligned} Q(z,s,\tau) &\triangleq p(z,\tau) \otimes p^*(s,-\tau) \\ Q_4(z,s,\tau) &\triangleq p^2(z,\tau) \otimes \left(p^*(s,-\tau)\right)^2 \\ F_4(z,s,\tau) &\triangleq \left|p(z,\tau)\right|^2 \otimes \left|p(s,-\tau)\right|^2 \\ Q_6(z,s,\tau) &\triangleq \left\{\left|p(z,\tau)\right|^2 p(z,\tau)\right\} \otimes \left\{\left|p(s,-\tau)\right|^2 p^*(s,-\tau)\right\}. \end{aligned} \tag{99}$$

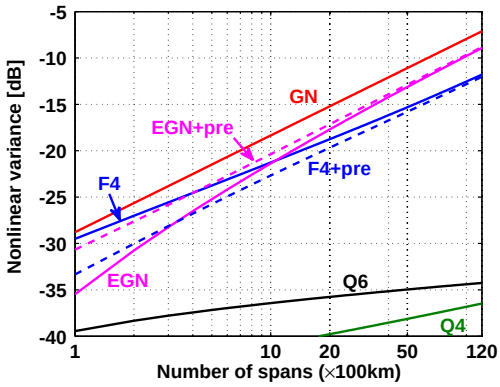


Fig. 11. Contributions to nonlinear interference variance (normalized to the cube of signal power $P$) versus number of spans $N_s$. 133 PDM-QPSK channels ($R = 32$ Gbaud, $\Delta f = 37.5$ GHz) transmitted into a $N_s \times 100$ km SMF dispersion uncompensated link. $F_4$ and $Q_4$ contributions are plotted in absolute value, since they are actually negative. The impact of a pre-compensation of 8500 ps/nm on some terms is shown by a dashed line. Note that GN model overestimates the EGN prediction. Originally from [91].

In particular, square $M$-QAM modulation with i.i.d. symbols has:

$$\kappa_{2;2} = -\frac{3}{5}\frac{M+1}{M-1}\mu_2^2, \qquad \kappa_{3;3} = \frac{36}{21}\frac{M^3-1}{(M-1)^3}\mu_2^2$$

The expression for $R_{zs}^{(\text{out})}(\tau)$ can be finally inserted in (97) and the PSD of the nonlinear interference can be evaluated by numerical integration. Please note that with Gaussian distributed symbols $\kappa_{i;i} = 0$, $i > 1$, so that (98) coincides with (96) since $R_{zs}^{(\text{in})}(\tau) = \frac{\mu_2}{T} Q(z, s, \tau)$ for a digital signal with supporting pulse $p(0, t)$.

The EGN model can be read as an additive correction to the GN model. Although the EGN is more accurate, it is worth noting that, except for very peculiar modulation formats [55], the GN model gives a conservative overestimation. The previous derivation was for single channel in single polarization. Extensions to PDM and WDM are available [91].

The relative importance of each term forming the EGN model is shown in Fig. 11 for a specific example detailed in the figure caption. We observe the mentioned GN overestimation compared to the EGN model. The most important EGN correction is the one provided by the fourth order correction $F_4$ [26]. Such a term is partially mitigated at short lengths by pre-distorting the signal with a pre-compensation fiber.

## VII. Multiplicative Models and Their Applications

We might wonder, at this point, if the models introduced in the previous sections definitively solve the problem of modeling the optical fiber channel. For instance, the representation of NLI as AWGN provides a good accuracy in terms of performance evaluation (the EGN model, in particular) with a reasonably low complexity (the GN model, in particular). However, it also entails that NLI cannot be mitigated, which is not entirely correct. For instance, deterministic intrachannel nonlinear effects can be exactly compensated for by digital backpropagation (DBP) (see Section XI-A). Moreover, as shown in Section V, some NLI terms due to XPM are not independent of the signal and should be more properly interpreted as time-varying ISI, rather than noise. In general, this means that the accuracy of AWGN-like models in terms of performance evaluation does not necessarily correspond to an accurate and detailed characterization of the statistical properties of NLI (e.g., non-Gaussian statistics, temporal correlation, and so on). These properties are not particularly relevant to determine the performance of systems employing conventional symbol-by-symbol detection, but might be the key to devise improved detection and nonlinearity mitigation strategies—e.g., able to mitigate also inter-channel nonlinearity and signal-noise interaction, or to perform DBP with a lower complexity. Research of alternative models for the optical fiber channel is, hence, still in progress. Strictly connected with this topic, is the computation of channel capacity (discussed in Section XII), which entails an accurate knowledge of the channel and the use of the best possible modulation and detection strategy.

A possible approach is provided by the LP model, briefly introduced in Section III-A, which falls within the broader class of *multiplicative models*. As opposed to *additive models*, which describe nonlinear effects by means of a noise-like additive term, multiplicative models describe nonlinear effects through a fading-like multiplicative term—a change of the channel characteristics that causes a distortion of the propagating signal. The latter approach is usually more complex than the former, but also closer to the underlying physics—the Kerr effect being a change of the refractive index induced by the propagating optical signal that causes a phase rotation of the signal itself.

The LP model was originally introduced in [38] (and generalized to all orders in [134]) to provide an analytical approximation of the signal at the output of a nonlinear dispersive fiber given the input signal. It has been later combined with the RP model to investigate signal-noise interaction, providing a joint description of both parametric gain and nonlinear phase noise effects [27]. It has been eventually applied to study interchannel nonlinearity in WDM systems, originating the frequency-resolved LP (FRLP) model [51], [59]. This last model and its applications will be briefly described in the sequel. The model has been developed for the single-polarization case, but the concept can be extended to dual polarization signals.

To investigate the impact of inter-channel nonlinearity in WDM systems, it is convenient to divide the overall WDM signal into the channel-of-interest (COI) component, still referred to as $A(z, t)$, and the interfering channels (IC) component $A_w(z, t)$—resulting from the combination of all the other WDM channels. Expanding the nonlinear term of the NLSE (1) into its SPM, XPM, and FWM components (according to the role played by the COI and IC components) [1], and retaining only the XPM one, the equation can be rewritten as

$$\frac{\partial A}{\partial z} = j\frac{\beta_2}{2}\frac{\partial^2 A}{\partial t^2} - 2j\gamma|A_w|^2 A \tag{100}$$

Equation (100) does not account for SPM, which can be compensated for by DBP (see Section XI-A), and for FWM and signal-noise interaction, whose impact in WDM systems

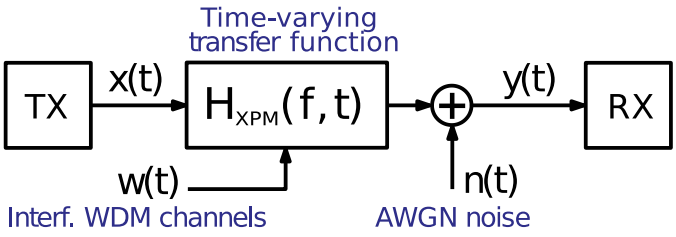


Fig. 12. Continuous-time FRLP model.

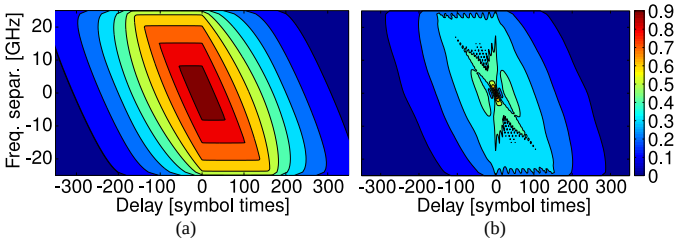


Fig. 13. XPM coherence: (a) IDA link; (b) 10x100km LA link.

is usually negligible compared to XPM (This might not be the case when considering WDM channels with a low symbol rate). The evolution of $A_w$ is, in turn, subject to nonlinear interference from $A$, such that we should also write an analogous equation for $A_w$, coupled to (100). However, in terms of the impact of $A_w$ on $A$, the nonlinear evolution of $A_w$ is a second-order effect. Thus, we will neglect it and assume that $A_w$ evolves linearly and independently of $A$. In this case, (100) is a linear Schrödinger equation with a space- and time-variant potential, whose solution can be formally expressed as

$$A(L,t) = \int_{-\infty}^{\infty} H_w(f,t)\tilde{A}(0,f)e^{j2\pi ft}df \qquad (101)$$

where $H_w(f,t)$ is a *time-variant* transfer function [257]. (In mathematics, the solution for this kind of equations is typically given in terms of the *Green's function* (or *propagator* in quantum mechanics), to which the time-variant transfer function is simply related by a Fourier transform.)

The transfer function can be obtained from the FRLP model and expressed as [51], [59]

$$H_w(f,t) = e^{-j\theta(f,t)} \qquad (102)$$

where the XPM term $\theta(f,t)$ depends on the input IC signal $A_w(0,t)$ and on the link characteristics. Some general expressions for $\theta(f,t)$ and its statistics are provided in [51], [59].

Eventually, including the effect of amplifier noise, the overall system (from the COI standpoint) can be schematically modeled as in Fig. 12 , in which $x(t) = A(0,t)$ is the transmitted signal; $y(t)$ the received signal; $H_w(f,t)$ the time-varying transfer function of the channel, which depends on the overall signal transmitted by the other channel users (the interfering channels) $w(t) = A_w(0,t)$; and $n(t)$ is an AWGN term that accounts for the accumulated optical amplifier noise.

A key observation here is that all the neglected effects (e.g., FWM and signal-noise interaction) can be eventually reintroduced by properly increasing the power spectral density of the AWGN term—for instance based on the calculations from the GN or EGN models or extracting the parameters from a numerical simulation. However, when modeling an effect as AWGN, we basically give up the possibility of mitigating it and make a worst-case assumption in terms of system performance. On the other hand, a more accurate modeling of the dominant XPM term may allow for its partial mitigation.

From the COI viewpoint, the model in Fig. 12 describes the nonlinear time-invariant optical fiber channel as a linear time-variant one. This seeming paradox is explained by the fact that channel nonlinearity is accounted for by the dependence of $H_w(f,t)$ on $w(t)$ [59]. However, assuming that $w(t)$ is unknown to both the transmitter and the receiver, such nonlinearity remains hidden, and the effect of $H_w(f,t)$ is simply perceived as a linear distortion. Moreover, since $w(t)$ depends on time, also the channel transfer function $H_w(f,t)$ depends on time.

The model in Fig. 12 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 [256]. This analogy may help to better understand the channel characteristics and behavior, as well as to devise improved transmission and detection strategies.

The coherence properties of the channel have been studied in [255] and are illustrated by the contour plots in Fig. 13, showing the correlation between the values $\theta(0,t)$ and $\theta(\Delta f, t+\tau)$ of the XPM term at two different times and frequencies inside the COI bandwidth, as a function of the delay $\tau$ and frequency separation $\Delta f$. (The channel is assumed to be stationary in time but not in frequency, as the impact of the XPM term depends on the frequency distance from the interfering channels and varies inside the COI bandwidth. The correlation is analyzed by holding one frequency fixed in the middle of the COI bandwidth (conventionally set to $f = 0$), and letting the other vary inside the COI bandwidth.) The considered scenarios refer to a Nyquist-WDM system with 50 GHz channel spacing and bandwidth, Gaussian symbols (a worst-case assumption in terms of XPM impact) and a 1000 km link of standard single-mode fiber with either ideal distributed amplification (IDA) (a) or 10x100 km lumped amplification (LA) (b). Only the XPM term generated by the couple of closest interfering channels (i.e., those located at $f = \pm[50]GHz$) is considered. The coherence is quite substantial in the IDA link, but significantly reduced in the LA link.

The FRLP model has been used to derive some closed-form expressions for the AIR with different modulation formats and a mismatched decoder optimized for the AWGN channel [59]. In this case, similar results could be obtained also with some additive models such as the RP or EGN models. But the FRLP model can be used also to devise more efficient detection strategies that exploit the coherence properties of the channel, and to derive some improved capacity bounds [67]. These issues will be discussed in greater detail in Section XII.

Equation (100) is also the starting point to derive other models and descriptions of interchannel nonlinearity. In [32], [253], Feynman path integrals and diagrammatic techniques

are used to derive the Green's function (propagator) of (100), from which the input-output channel statistics are obtained by assuming that the stochastic potential $|A_w|^2$ is a Gaussian process with short-ranged correlations in space and time—a more drastic approximation than the assumption of linear propagation of $A_w$ done in deriving (102). The resulting model has been used to compute some capacity lower bounds for the optical fiber channel [32], [253].

Another approach to solve (100) is the *Magnus expansion*, a general method to derive approximate exponential representations of the solution of linear differential equations with varying coefficients [250]. The Magnus expansion provides a power series expansion for the corresponding exponent and, for this reason, is sometimes referred to as time-dependent exponential perturbation theory. In optical fiber communications, the Magnus expansion has been already employed to model linear effects such as polarization-mode dispersion and polarization-dependent loss [251], while its application to fiber nonlinearity has been only slightly touched [252]. When applied to (100), the Magnus expansion is quite accurate even at first order—in many cases more than the corresponding first-order RP, eRP, and LP expansions [252]—and has the desirable additional property of preserving at any order some important features of the original equation, such as its unitarity. These characteristics make it an interesting subject for future research.

## VIII. Model-Specific Features

In this section we try to provide a list of features of the various fiber non-linearity models introduced so far, which make them attractive for certain applications. Carrying out a more comprehensive and comparative analysis of the models would be a difficult task and we refrain from doing it. Some of the aspects mentioned here are dealt with in later sections, which the reader should refer to.

**GN-model**: reasonable compromise between accuracy and complexity, very compact formula. **iGN-model**: for most systems, very good compromise between accuracy and complexity. **GN/iGN-models**: lend themselves to closed form solutions, some of which are extremely simple yet rather accurate, such as for ideal Nyquist WDM with completely identical spans, helping establish fundamental limits; others are quite flexible (any comb, any link) but still fully closed form, very good for management of physical-layer aware optical networks. **EGN-model**: complete and accurate model in frequency domain; computationally very complex, reduced version exists which is quite accurate and 1/3 as complex, but complexity is still high. Does not allow to accurately factor out PPRN, though it does allow to approximately factor out NLPN by assuming constant-envelope transmission (i.e. assuming PM-QPSK-like parameters). **Time-domain model**: Very similar to the EGN model in terms of accuracy and complexity when predicting interference variance; also predicts nonlinear phase and polarization rotation noises (PPRN); predicts temporal correlations; describes nonlinear interference as an inter-symbol interference (ISI) which can be used to examine and evaluate the performance of various ISI cancellation techniques; predicts the contribution of the various ISI terms. **Spatially-resolved model**: Alternative description of the EGN model; predicts temporal correlations of the optical nonlinear interference; focus on nonlinear spatial interactions along the link, such as nonlinear signal-noise interaction; complexity grows with system length. **FRLP model**: represents NLI as a multiplicative distortion rather than as an additive noise; describes the channel through a time-varying transfer function; allows to identify the NLPN component and calculate its correlation in time and frequency. These characteristics help devising improved detection and nonlinearity mitigation strategies. Closed-form solutions are available for simple system configurations (e.g., ideal Nyquist WDM).

## IX. Impact of specific effects on NL modeling

### A. Nonlinear signal ASE interaction

Amplified spontaneous emission (ASE) noise interacts with the nonlinear Kerr effect along propagation. The result is generally referred to as nonlinear signal-noise interaction (NSNI). NSNI can be classified as a FWM process. However, it finds its own space in the literature because of the distinctive nature of some of the photons joining the FWM interaction, which come from a wide-band Gaussian distributed signal such as ASE.

Initial studies on NSNI date back to investigations about the nonlinear interaction between a strong constant-wave (CW) signal and wide-band noise [17], [136]. In such a case only degenerate FWM yields significant NSNI, which motivated the use of small-signal analysis to analyze the effect. Results highlighted that an optical fiber can induce modulation instability (MI) depending on the dispersion sign. In particular, the white spectrum of noise may experience a gain (hence the name modulation instability) which is maximum at frequency $\sqrt{2\gamma P/\left|\beta_2\right|}$, with $P$ the power of the strong CW [17].

Modulation instability has been exploited to build optical parametric amplifiers (OPA) [137]. OPA shows distinguished new features compared to existing optical amplifiers such as adjustable gain spectra and center frequency, possibility to work in phase sensitive mode with ideally zero noise figure, possibility to perform phase conjugation and wavelength conversion [138]. However, OPA shows non trivial problems such as gain shape fluctuations and polarization dependence that limited their applications in practice [138].

Besides the positive implications of NSNI in building optical amplifications, NSNI originating along the transmission fiber is generally a problem for the system performance. Initial studies showed that NSNI can induce strong phase noise, usually referred to as Gordon and Mollenauer effect [29], whose impact has been shown to severely affect phase modulated signals such as differential phase shift keying (DPSK) transmissions [139]. A closed formula of the NSNI-induced phase noise variance and its main scaling properties was proposed in [29].

The phase noise induced by NSNI is a random variable whose exact probability density function was derived by K.-P. Ho in [37] in the absence of dispersive effects and before detection filters. The impact of dispersion is much more

complex, such that the exact PDF is unknown but closed formulas exist for the NSNI variance due to intra-channel pulse XPM [39]. Small-signal analyses of NSNI in dispersive links were carried out in [31], [140] and in [141], [142] where they were referred to as parametric gain (PG).

ASE noise can be included in the GN- and EGN-model. Since ASE is inserted in the link along propagation, the spatially-resolved model of Fig. 10 is well suited to search for modifications to GN-based formulas [145].

Since ASE is Gaussian distributed, its dominant contribution appears as an enhancement of the GN part of the EGN formula (98). However, ASE impacts even the fourth order nonlinear interference contribution proportional to $\kappa_{2;2}$.

One of the main implications of NSNI to GN-model variance is that the NSNI nonlinear interference variance does not scale with the cube of the signal power, but with the square. This is not surprising since the most dominant contribution to NSNI comes from FWM processes involving the beating of signal×signal×ASE [81], [149], [147], [145]. As a rule of thumb, numerical results showed that NSNI starts to impact the nonlinear interference variance at SNR smaller than 10 dB [28], [145].

Some formulas have been proposed in the literature to modify the GN-model formula and account for NSNI. Poggiolini et al. proposed to use the incoherent-GN formula by substituting the channel power at the input of fiber $k$ with $P + P_{\mathrm{ASE}}k$, $P$ being the transmitted power [28]. Serena [145, eq. (10)] proposed to correct the cross-correlation between the nonlinear interference brought by two different spans by an ASE dependent term. Lavery et al. [146] proposed to weight the NSNI accumulation coherence factor in a different way than the signal such that each ASE term contributes from its injection point in the link. Ghazisaeidi extended the theory of [26] by providing an efficient algorithm for NSNI variance computation [148].

NSNI is the fundamental limit to the performance in presence of nonlinear equalization [147], [146], [145], [148], a feature that motivated the search of effective models of NSNI with the aim of finding the capacity of optical communication links. If we perturb the entire link including digital back-propagation, we get the spatially resolved model of Fig. 14 [145]. We note that, without ASE, for any input/output branch there is a branch with opposite parameters, such that the overall nonlinear interference is zero. However, with ASE this symmetry is broken, such that a residual NSNI cannot be compensated for by a nonlinear equalizer. Lavery et al. [146] showed that the optimum is to equally split the nonlinear equalizer between transmitter and receiver side.

### B. System impact of nonlinear phase and polarization noise

Sect. V-2 provides the tools for accurate modeling of nonlinear phase and polarization noise (PPRN). These tools can then be used to assess the actual system impact of PPRN. As a relevant example, Fig. 15 shows the SNR gain that can be obtained by ideal removal of all inter-channel PPRN. The plot is significant, since it addresses a realistic C-band fully populated 115-channel system, over SMF, with 100 km

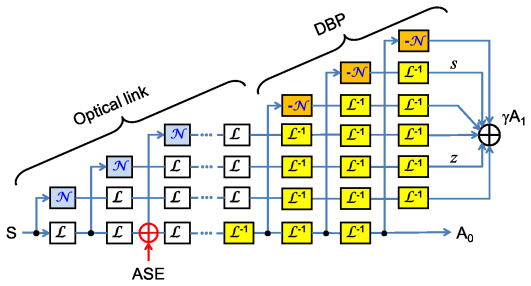


Fig. 14. first-order representation of an optical link including digital back propagation (DBP). Originally from [145]. The paths label with $s$ and $z$ do not cancel out in presence of the ASE source indicated in the figure.

span length, at 32 GBaud. The shown SNR gain assumes that launch power is optimized to take the maximum advantage out of inter-channel PPRN removal. Hence, the term 'peak-SNR gain' is used (see [4] for more details). Note that peak-SNR gain can also be interpreted, conversely, as the *impact* that PPRN has on the system. In other words, QPSK crossing the 0.5 dB line at 5 spans in Fig. 15 can be read as either a 0.5 dB potential gain if PPRN is removed, or a 0.5 dB SNR degradation due specifically to PPRN.

Standard CPE (Carrier Phase Estimation) algorithms in the receiver DSP can mitigate the effect of PPRN, and in particular of its phase-noise component (NLPN). However, these algorithms typically leverage the time-correlation that such disturbances exhibit, quantified in Fig. 8. Unfortunately, where mitigation would be most needed, such as in short links according to Fig. 15, PPRN has short correlation time (Fig. 8), so that mitigation by standard CPE algorithms is only partially effective. PPRN can be almost completely removed in long links, where it exhibits long correlation time. There, though, its impact is relatively small and the obtainable mitigation gain is limited.

This topic is however complex and Fig. 15 only looks at one scenario. Many system features, such as fiber dispersion, distributed amplification, modulation format, etc., affect PPRN. For a comprehensive analysis and extensive references, see [4]. Symbol rate too affects PPRN. Recently, the scaling of the impact of NLPN vs. *symbol rate* was studied for instance in [261], [262], using an approximate technique based on the EGN model. The indication is that, when the symbol rate is increased, the variance of NLPN decreases and its correlation-time increases. As a result, overall, its impact appears to decrease at higher symbol rates.

### C. Joint nonlinear and polarization effects, PMD / PDL

Although single mode fibers support only one propagating mode, the resulting field is the superposition of two polarizations [93]. Each polarization can be independently modulated and the resulting signal is generally referred to as polarization division multiplexing (PDM), also known as dual polarization (DP) or polarization multiplexing (PM). PDM is a simple technique to double the spectral efficiency and represents the starting point of more advanced techniques of spatial division multiplexing [94].

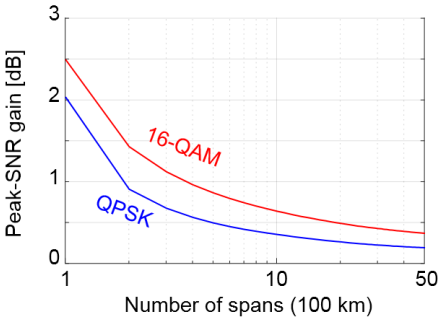


Fig. 15. Peak-SNR gain resulting from perfect inter-channel PPRN removal in fully loaded systems with 115 WDM channels at 32-GBaud. The detailed parameters used in this figure are given by Table 2 in [4]. The Peak-SNR gain is difference between the maximum SNR, with and without inter-channel PPRN. For more information on Peak-SNR gain, see [4] Sect. 4.

However, the two polarizations usually interact along propagation, thus generating polarization crosstalk at the receiver. The reason of such cross-coupling is related to perturbations to the ideal fiber structure, such as changes in the refractive index, non-circularity of the fiber, thermal stresses, variations in the core radius, irregularity at the core-cladding boundary, etc. Perturbations may be even deliberately introduced during fiber fabrication as a way to improve performance, as we will discuss later. All such stresses make the fiber anisotropic, i.e., they introduce birefringence.

As a result of birefringence, the polarizations may travel at different speed thus manifesting polarization mode dispersion (PMD) [110]. A main implication of PMD is the end-to-end group delay spread between polarizations, called differential group delay (DGD). The problem is of particular concern in presence of polarization coupling along transmission, because to undo the coupling at the receiver we must account for the memory introduced by DGD.

The accumulation of PMD along distance depends on the correlation length over which mode coupling occurs, typically of about 100 m for SMF [93]. This way, a long haul link operates in the strong coupling regime, whose random properties along distance eventually make the DGD a random variable, with Maxwellian statistics in the limit of correlation length approaching zero [97], [98]. Most importantly, the average DGD and the DGD standard deviation accumulate proportionally to the square root of distance [96], [99]. This circumstance eases equalization of PMD. On the contrary, GVD has a much smaller randomness such that it accumulates with a rate proportional to the distance.

For such reasons, short fiber correlation length is intentionally induced during fabrication by properly spinning the fiber [101]. Data sheets usually reports the PMD coefficient $\eta_{\text{PMD}}$. Typical values are 0.32 ps/$\sqrt{\text{km}}$ for pre-1991 fibers, 0.13 ps/$\sqrt{\text{km}}$ for fibers installed in the period from 1992 to 1998, and 0.05 ps/$\sqrt{\text{km}}$ for post-1999 fibers [100]. For a fiber of length $L$ The average DGD $\overline{\tau}$ is related to the PMD coefficient by $\overline{\tau} = \eta_{\text{PMD}}\sqrt{L}$, while the r.m.s. value of DGD is $\tau_{\text{rms}} = \sqrt{\frac{3\pi}{8}}\overline{\tau}$ [98].

PMD is generally a problem in high symbol rate systems and must be properly equalized at the receiver side. Adaptive equalization is mandatory since PMD varies in time on scales of the order of milliseconds. The problem is best solved with coherent detection where classical algorithms first developed for wireless communications have been successfully used. Some examples are the blind constant-modulus algorithm (CMA), the least mean square (LMS) algorithm and the data-aided least squares method [95].

Both birefringence and PMD play a role in the nonlinear regime. As mentioned in the introduction, the fast variations of birefringence with distance are usually averaged out as in the Manakov equation [99], [20]. In this scenario, the two polarization tributaries still nonlinearly interact along propagation through XPM, but they also interact through cross polarization modulation (XPolM) [19], [99], [109], [106], [107]. XPolM is a generalization of the XPM phase rotation in the complex plane into a rotation in the three-dimensional space described by the Poincaré sphere. Now it is the total power that is preserved: however, power can be exchanged between the polarization tributaries thus creating nonlinear crosstalk. Contrary to PMD temporal variations, XPolM is rapidly varying in time, with temporal scales of the order of the walk-off between interacting channels. As a consequence, XPolM scattering manifests as a Brownian motion over the Poincaré sphere [106], [107] and its equalization is extremely challenging. XPolM has been shown to be one of the dominant nonlinear impairments in dispersion-managed PDM systems [104], [105].

PMD joins the polarization and nonlinear Kerr effect interactions [108]. The analysis is simplified in the strong coupling regime where the reference model is the Manakov-PMD equation [20]. Numerical simulations and experiments [102], [103] showed that PMD helps improving the average performance of an optical link, provided that linear PMD is fully equalized at the receiver. The reason is that PMD is a dispersive effect, hence it decorrelates channels farther away than the PMD correlation bandwidth [82], thus mitigating the nonlinear interference accumulation along distance.

The numerical investigation of PMD in the nonlinear regime is particularly burdensome. To correctly emulate the slow birefringence temporal variations it is mandatory to independently test many fiber realizations to find rare events that set the outage probability induced by PMD, i.e., the probability that the Q-factor is below a given threshold. Attempts to extend the GN model to include PMD are reported in [82].

Birefringence may also induce polarization dependent loss (PDL), i.e., an energy loss preferential to one polarization. PDL is mainly present in optical devices such as reconfigurable add drop multiplexers (ROADM) or EDFA. PDL is expressed in international standards by [93]:

$$\rho_{\text{dB}} \triangleq 10\log_{10}\left(\frac{T_{\max}}{T_{\min}}\right)$$

where $T_{\max}$ and $T_{\min}$ are the maximum/minimum transmission power after a PDL element. Typical values of $\rho_{\text{dB}}$ are fractions of dB (e.g., 0.4 dB for a ROADM). The overall PDL cumulated along a link is a random variable, whose statistics can be

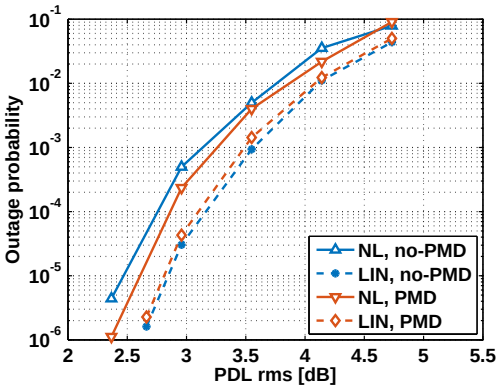


Fig. 16. Outage Probability vs PDL. An outage event occurs when Q-factor is smaller than 6.25 dB. Dashed lines: equivalent linear (LIN) model with the same average Q-factor as the in nonlinear (NL) propagation. PMD = 0.13 ps/$\sqrt{\text{km}}$. Originally from [118].

evaluated by following similar methods as for DGD [112], [111].

PDL induces a penalty which is enhanced in nonlinear regime by interaction with the nonlinear Kerr effect, thus increasing outage probability [114], [21], [113].

Fig. 16 shows an example of outage probability estimated by numerical simulation in a 15 channel 32 Gbaud QPSK system, traveling in a $35 \times 100$ km, $D = 4$ ps/nm/km, link at a power of 0.5 dBm and OSNR=16.1 dB/0.1 nm [118]. PMD was either 0 or 0.13 ps/$\sqrt{\text{km}}$. Dashed lines show the result by an equivalent linear model where the EDFA noise figure has been changed to get the same average Q-factor as in nonlinear regime. The difference between the solid and the dashed line is an indication that PMD/PDL interacted with nonlinearity along transmission.

## X. Applications

### A. Which performance metric?

The main outcome of the GN model is the received SNR. In the special, yet relevant, case of the additive white Gaussian noise (AWGN) channel with matched filter detection, the SNR can be converted into bit error rate (BER) by simple formulas, as shown in Appendix A.

Usually, BER follows a complementary error function (erfc)-like behavior versus SNR, which does not suit a system designer that usually works in dB scales. For this reason, the Q-factor has been introduced. The Q-factor is a one-to-one relation with the BER in a reference system tracking the erfc transformation (see Appendix A). This way, the Q-factor coincides with the SNR in the relevant case of QPSK, and with 2·SNR for BPSK. For several other modulation formats the relation is luckily almost linear with slope 1 in a dB/dB scale. Conversions graphs are available in Appendix A.

Whatever the choice, BER or Q-factor, they are normally estimated before FEC decoding. The rationale is to associate a threshold to the FEC code and claim the post-FEC BER to be 0 if the pre-FEC BER is smaller than the threshold. For instance, the second generation hard-decision FEC (HD-FEC) standardized by ITU-T G.975.1 had a Q-factor threshold of 8.5 dB with a FEC overhead of 6.7%.

Nowadays the trend is toward soft-decision FEC (SD-FEC) that can provide gains of about 1.5 dB with respect to hard-decision [120], or even more with iterative decoding. SD-FEC so far operate at overheads of ∼20% by exploiting soft decoding with advanced techniques such as low-density parity check codes (LDPC) or turbo product codes (TPC). However, it is worth noting that the higher complexity of SD-FEC may be an issue for some systems, thus HD-FEC with overhead $\lesssim 15\%$ may be preferred for those applications where low power consumption is mandatory [120].

The idea of a FEC threshold works fine with binary HD-FEC and bit-interleaving, and thus BER works great in this scenario. However, it is has been shown that it may fail with SD-FEC [121].

An alternative performance metric, more suited to SD-FEC, is the mutual information (MI). MI is a concept introduced in information theory to provide a measure of the "amount of information" obtained about one random variable through the knowledge of another random variable [122]. The concept of MI is strictly related to the concept of entropy, which provides a measure of unpredictability on average. For the continuous random variable $Y$ with probability density function $p_Y(y)$ the entropy (More correctly, differential entropy.) is defined as:

$$h(Y) \triangleq -\mathbb{E}\left[\log(p_Y(y))\right]$$

with $\mathbb{E}\left[.\right]$ expectation. MI between two random variables $X$ and $Y$ is thus defined as:

$$\text{MI}(X,Y) \triangleq h(Y) - h(Y|X) \tag{103}$$

where $h(Y|X)$ is the conditional entropy, i.e., the entropy of $Y$ conditioned to $X$ averaged over all possible values of $X$. Hence, the computation of (103) requires knowledge of the conditional probability $p_{Y|X}(y|x)$.

MI is a relationship between two random variables. For channels with memory we need to generalize the idea to stochastic processes, hence we should substitute MI with the information rate (IR) [122], [124]. Let $\mathbf{X}_N = (X_1, X_2, \ldots, X_n)$ and $\mathbf{Y}_N = (Y_1, Y_2, \ldots, Y_n)$ temporal sequences of the stochastic process $X(t)$ and $Y(t)$, respectively, from discrete time $t = 1$ to time $t = n$. The IR is defined as:

$$I(\mathbf{X}_N, \mathbf{Y}_N) \triangleq \lim_{n\to\infty} \frac{1}{n}\left(h(\mathbf{Y}_N) - h(\mathbf{Y}_N|\mathbf{X}_N)\right)$$

We are motivated to associate $Y_k$ to the input of soft-decoder and $X_k$ to the transmitted symbol. The GN model tells us that the AWGN channel is a good approximation of the true channel, thus MI can be evaluated from the memoryless model $Y_k = X_k + n_k$ with $n_k$ a complex Gaussian random variable. However, the true optical channel is different, not just in terms of different statistics but also because the optical channel exhibits memory, hence such straightforward approach cannot exploit all the useful information [143], [144]. Nevertheless, the theory of mismatched decoding [123] shows that such MI is a lower bound to the maximum IR [124] constrained to the modulation format under use. An example of MI for QPSK modulation in AWGN [130] is reported in Fig. 17.

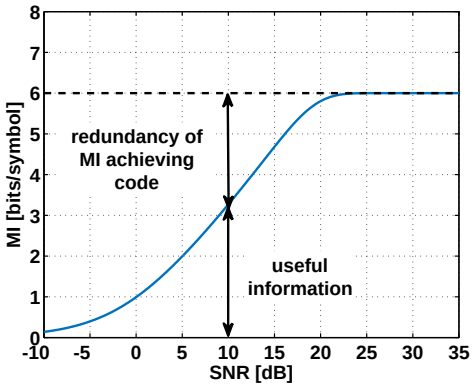


Fig. 17. MI of 64-QAM in AWGN. The error probability after FEC can be made arbitrary small by a code wasting in redundancy at least the bits/symbol indicated in the figure.

In summary, if the soft-decoder has just the information provided by the GN model, i.e., it believes that noise is additive and Gaussian and knows its variance, we say that the corresponding MI described above is achievable, i.e., there exists a code for which error-free transmission is possible at MI bits/symbol. The redundancy of such best-code is reported, for example, in Fig. 17 for the reference case of SNR=10 dB. Practical codes, able to exploit the information provided by the auxiliary channel, will need higher redundancies even if nowadays with LDPC and TPC the loss is very small.

We say that the AWGN channel approximation is an instance of an auxiliary channel of the true channel [124]. Other auxiliary channels are possible: the closer the auxiliary channel is to the true channel, the tighter is the lower bound to the maximum constrained IR. The capacity is the supremum among all possible modulation formats and symbols distributions.

MI is the relevant metric for symbol-wise receivers. However, optical communication systems often use bit-interleaved coded modulation (BICM) [125] whose non-iterative implementation [126] simplifies decoding. For such schemes, the generalized mutual information (GMI) is better suited [126], [121]. The main difference with MI is that GMI depends on the bit-symbol mapping. If $X = (B_1 B_2 \dots B_L)$ is the string of bits, $B_k \in (0,1)$, corresponding to symbol $X$, GMI is defined as [126], [127]:

$$\text{GMI} \triangleq \sum_{k=1}^{L} \text{MI}(B_k, Y)$$

It is always GMI $\leq$ MI. For instance, for BPSK it is GMI=MI because here bits≡symbols. QPSK is the composition of two BPSK in quadrature, hence with Gray coding and AWGN each bit can be detected independently from the other by using the real/imaginary axis as discriminating threshold. Hence, even for QPSK GMI=MI. In particular, $\text{MI}_{\text{QPSK}}(\text{SNR}) = 2{\cdot}\text{MI}_{\text{BPSK}}(\text{SNR}/2)$.

For higher order PSK and QAM modulation formats GMI $<$ MI. The reason is that in these cases the mentioned bit-independence in detection rule is broken, and thus we can extract more information by using a symbol wise detector rather than a bit-wise detector after bit interleaving.

For the AWGN channel the difference between GMI and MI is limited to fractions of bits, see Appendix A.

The debate about the most appropriate performance metric is still open. Schmalen et al. showed that with non-binary FEC MI is the right candidate [128]. Cho et al. showed that with probabilistic shaping normalized GMI yields a better correlation with the post-FEC [129], while Yoshida et al. [131] proposed asymmetric information (ASI) as a better predictor in nonlinear regime.

### *B. Maximum reach and optimum launch power*

According to the perturbative description of the NLSE, the received SNR can be expressed by [43], [119]:

$$\text{SNR} = \frac{P}{\sigma^2_{\text{ASE}} + \sigma^2_{\text{NL}}} \tag{104}$$

where $P$ is the signal power, $\sigma^2_{\text{ASE}}$ is the cumulative ASE power generated along the link while $\sigma^2_{\text{NL}}$ is the nonlinear interference variance. For a transparent $N$-span periodic link with optical amplifiers of noise figure $F$ and gain $G$ it is $\sigma^2_{\text{ASE}} = h\nu FGBN$, with $h$ Planck's constant, $\nu$ carrier frequency and $B$ receiver bandwidth. According to GN model theory, $\sigma^2_{\text{NL}} = \eta P^3$ where $\eta$ is the unit-power nonlinear interference coefficient discussed in the previous sections.

Several interesting implications for system design can be inferred from (104). First, since ASE dominates the SNR at small powers while nonlinear interference dominates at high powers, an optimal power exists (see Fig. 1). Such a power is generally referred to as *nonlinear threshold* $P_{\text{NLT}}$, and can be easily found by setting to zero $\frac{\text{dSNR}}{\text{d}P}$, with result:

$$P_{\text{NLT}} = \left(\frac{\sigma^2_{\text{ASE}}}{2\eta}\right)^{1/3} .$$

By inverting such relation we find that $\sigma^2_{\text{ASE}} = 2\eta P^3_{\text{NLT}}$, i.e., at the optimal launched power, also known as nonlinear threshold (NLT), ASE variance is twice nonlinear interference variance. This important result tells us that at best power the role of linear ASE noise is more important than the role of nonlinear distortions, such that it is better to put efforts in optimizing linear propagation rather than nonlinear propagation.

By substituting $P_{\text{NLT}}$ into (104) we get the corresponding maximum SNR:

$$\text{SNR}_{\text{NLT}} \triangleq \max(\text{SNR}) = \frac{P_{\text{NLT}}}{\frac{3}{2}\sigma^2_{\text{ASE}}} = \frac{1}{\left[27\eta\left(\frac{h\nu FGBN}{2}\right)^2\right]^{1/3}} .$$

Besides these key ingredients to set up a connection, it is interesting to have a look at the SNR penalty with respect to linear impairments. By factoring out in (104) the SNR impaired by linear effects only, $\text{SNR}_{\text{lin}} = \frac{P}{\sigma^2_{\text{ASE}}}$, we can define such a penalty SP by:

$$\text{SP} \triangleq \frac{\text{SNR}_{\text{lin}}}{\text{SNR}} = \frac{\sigma^2_{\text{ASE}} + \sigma^2_{\text{NL}}}{\sigma^2_{\text{ASE}}} .$$

The interesting implication is that at best power $P_{\text{NLT}}$ the SNR penalty it is always $3/2$, whatever the value of $\sigma^2_{\text{ASE}}$ and $\eta$, i.e., whatever the link for which the GN model assumptions work!

Such a $3/2$ factor is more conveniently expressed in dB by $10\log_{10} 3/2 \simeq 1.76$ dB. Therefore, if our system is working at power $P_{\text{NLT}}$ and we do not want to change such a power, at best we can improve the SNR by 1.76 dB with an ideal receiver able to exactly compensate for nonlinear impairments. Only by increasing the power we can improve the SNR more than 1.76 dB by using a better receiver.

For the system designer used to work in dB it is interesting to have a look at what happens by 3 dB changes in ASE and nonlinear interference power. We have:

$$\sigma^{2'}_{\text{ASE}} = \sigma^{2}_{\text{ASE}} + 3 \text{ dB} \quad \Rightarrow \quad \begin{cases} P'_{\text{NLT}} = P_{\text{NLT}} + 1 & \text{dB} \\ \text{SNR}'_{\text{NLT}} = \text{SNR}_{\text{NLT}} - 2 & \text{dB} \end{cases}$$

$$\eta' = \eta + 3 \text{ dB} \quad \Rightarrow \quad \begin{cases} P'_{\text{NLT}} = P_{\text{NLT}} - 1 & \text{dB} \\ \text{SNR}'_{\text{NLT}} = \text{SNR}_{\text{NLT}} - 1 \ . & \text{dB} \end{cases}$$

Interestingly, a 3 dB change in ASE or nonlinear interference induces an absolute change of 1 dB in nonlinear threshold.

The GN model can be used to infer the reach of the system at a given SNR, i.e., the maximum transmission distance ensuring the signal to noise ratio SNR at the receiver. Such SNR can be, for instance, the SNR threshold of the FEC code under use. An example of signal power contour levels and their related reach is shown in Fig. 2. To estimate the reach we need to know how the ASE power and the nonlinear interference $\eta$ scale with the number of spans $N$. The ASE power has a simple linear scaling $\sigma^2_{\text{ASE}} = h\nu FGBN$. Experiments, numerical simulations, and theoretical models suggest indeed that the nonlinear interference follows the scaling law [70], [47], [62]:

$$\eta = \eta_1 N^{1+\varepsilon}$$

where $\eta_1$ is the normalized nonlinear interference variance after 1 span, while $\varepsilon$ is the coherence accumulation factor accounting for the super-linear accumulation of nonlinear effects along distance. The presence of $\varepsilon$ is due to correlations among the different paths creating the first order perturbation (see Fig. 10). In highly dispersive links the strong dispersion makes the signal entering a given span almost uncorrelated with the signal entering a different span, hence we expect also the nonlinear interference to be almost uncorrelated span-by-span, such that the cumulative effect accounted by $\eta$ scales linearly with distance, i.e., $\varepsilon \approx 0$. On the contrary, the accumulation factor is $0 < \varepsilon < 1$, with the extreme case of $\varepsilon = 1$ in fully-compensated dispersion-managed links. In dispersion uncompensated links it is typically $\varepsilon < 0.2$ [47], [62]. $\epsilon$ would be almost zero if only cross-channel nonlinearities were present because the channel-walk-off decorrelates cross-nonlinear interference from span to span. A $\epsilon > 0$ is thus due to a significant single-channel nonlinearity. Therefore single channel nonlinearity grows faster with distance than cross-channel nonlinearity. This implies for instance that single channel nonlinear equalizers become more effective at large propagation distances.

In this framework the SNR is thus function of three variables, $\sigma^2_{\text{ASE}}$, $\eta_1$ and $\varepsilon$. To get the reach we thus need at least three measurements of power yielding the reference SNR at different distances. Based on the previous discussion, at the reach $N_0 \triangleq \max(N)$ the ASE variance is twice the nonlinear interference variance, hence the following identities hold:

$$\text{SNR} = \frac{P_0}{\frac{3}{2} h\nu FGBN_0} = \frac{P_0}{3\eta_1 N_0^{1+\varepsilon} P_0^3}$$

where $P_0$ is the signal power at the reach. The previous expression gives two equations in two variables $(P_0,\ N_0)$ from which we get our target, the reach [47]:

$$\text{reach} \triangleq N_0 = \frac{1}{\left[(3\text{SNR})^3 \eta_1 \left(\frac{h\nu FGB}{2}\right)^2\right]^{\frac{1}{3+\varepsilon}}}. \qquad \text{[spans]}$$

The reach is a function of the GN model key parameters $\eta_1,\ \varepsilon$ whose estimation is affected by errors. It is interesting to have a feeling of how reliable the formula of the reach is with variations of such parameters. Since it is preferable to work in dB scales, the following derivatives hold [47]:

$$\frac{\partial N_0^{\text{dB}}}{\partial \text{SNR}^{\text{dB}}} = -\frac{3}{3+\varepsilon}, \quad \frac{\partial N_0^{\text{dB}}}{\partial F^{\text{dB}}} = -\frac{2}{3+\varepsilon}, \quad \frac{\partial N_0^{\text{dB}}}{\partial \eta_1^{\text{dB}}} = -\frac{1}{3+\varepsilon}. \tag{105}$$

For instance, at $\varepsilon = 0$ an estimation error on $\eta_1$ of 1 dB translates into a reach error of 0.33 dB [62], [150].

Another key message from (105) is the following. To increase the reach we can play with i) the FEC by using a better code with a smaller SNR threshold, with ii) the optical link, for instance by using better amplifiers with a smaller noise figure $F$, or with iii) the nonlinearity, for instance by using a nonlinear equalizer to reduce $\eta_1$. Equation (105) tell us that such strategies are in order of effectiveness.

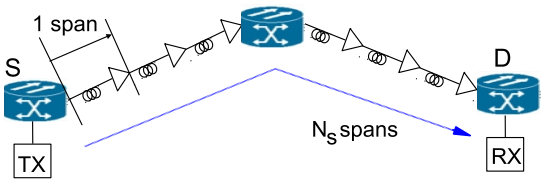


Fig. 18. Example of selected lightpath from source (S) to destination (D) transmitted on a fixed wavelength across 6 spans, with 2 hops and 3 span/hop.

### C. Modeling in networks

The GN model proved to be an invaluable analytical tool, able to explain the most important scaling laws and features of the new regime of highly dispersed transmission entailed by dispersion uncompensated links with coherent detection and electronic digital signal processing. However, the GN model also finds one of its most practical uses in route selection in modern optical networks with quality-of-transmission (QoT) guarantees.

The dominant optical network paradigm is that of wavelength-routed optical networks (WRON), where a transparent optical channel (a lightpath) is established from source to destination on a fixed wavelength across multiple fiber links and optical switching nodes (see Fig. 18 for an example). The most recent evolution of WRONs is that of elastic optical networks (EON) (see, e.g., [151], [152] and references therein), where bandwidth can be flexibly allocated to lightpaths with a granularity down to 6.25 GHz, thanks to

coherent detection and smart wavelength selective switches. Flexibility may also concern the modulation format and/or the forward error-correction (FEC) code of each lightpath at the transmitter in order to assure reaching the destination without intermediate regenerations. Electro-optical (EO) regenerations at intermediate nodes may however be necessary to both allow electrical multiplexing/demultiplexing onto the lightpath at intermediate nodes (*grooming*) and possibly to change the wavelength of the subsequent section of the lightpath in order to reduce wavelength blocking (WB). Reduction of WB through wavelength conversion has the beneficial effect of allowing an increase of network utilization (i.e., the fraction of used wavelengths in the network) and thus of network throughput, but at the moment is still considered a too-costly option because of the cost of EO regenerations and is used with great care.

Initial efforts in dimensioning EONs when taking into account nonlinear propagation effects assumed a worst-case full-load scenario where all wavelengths in every fiber link are populated, and nonlinear effects are at their maximum [153]. Such a full-load assumption has the great advantage of simplifying the routing, modulation and spectrum assignment (RMSA), since it decouples the propagation problem from the traffic-dependent linear and nonlinear interference. Hence every modulation format has an optimal launch power spectral density and a maximum error-free transmission distance, known as the *reach*, and a simple comparison of the reach with the requested source-destination distance is enough to assess whether a connection is feasible without intermediate regeneration or not. The other major advantage of the full-load assumption is that of making existing lightpaths insensitive to the establishment of new connections, a fact that makes RMSA decisions fast even in a dynamic traffic scenario where connections my be established and released on short time scales.

The first RMSA algorithms were based on look-up tables containing the reach for every possible modulation format. The reach tables were evaluated offline by long statistical computations based on the split-step Fourier propagation method. Semi-analytical methods based on a single span propagation and tuning with variable residual input span dispersion were also proposed to speed up reach computations [154].

However the fully-analytical formulas afforded by the GN model quickly found their way in the RMSA design because they drastically speed-up the calculation of the required received optical signal to noise ratio (OSNR) of a reference lightpath for a given modulation format, and thus the assessment of whether it is above its FEC threshold and the connection is thus feasible.

In particular, the incoherent GN model treats the nonlinear interference terms generated at each span as if they were independent additive Gaussian noise, exactly as the optical noise from the amplifiers.

The first proposed use of the incoherent GN model in an EON scenario appeared in [88], [155] where the locally-optimal globally-optimal Nyquist (LOGON) strategy was introduced. It was there first observed that, when using the optimal power spectral density at full load, the reach at any load is never larger than 3/2 the full-load reach. This still is the major theoretical justification for using the apparently resource-wasteful full-load assumption for nonlinear interference evaluation. The reach computation using the GN model for optical networking was explored in [156]. The LOGON strategy was analyzed in terms of blocking probability versus carried traffic in a theoretical setting [157] and recently for various network topologies [158], while in [159] the expected gain in the network throughput due to transceiver adaptation was studied in EONs using the full-load LOGON strategy.

Alternative, equivalent formulations of the GN model that make the single- and cross-channel contributions explicit for networking applications appeared in [160].

The enhanced GN (EGN) model, which takes into account the effect of the modulation format on nonlinear interference, can alternatively be used in place of the GN model [161] for applications where route establishment can take place on time scales of several minutes, thanks to the development of fast computing techniques (see e.g., [162]).

Recently, many papers have appeared that use either the GN or the EGN model within RMSA algorithms and try to remove the full-load assumption in order to make a better use of network resources.

One line of such developments [163], [164], [165], [166], [167], [168] explores the use of the GN analytical OSNR within a mixed nonlinear programming framework in order to obtain the absolute optimal RMSA in a static traffic scenario (one where a traffic matrix is offered to a green-field network and lightpaths are set up one by one to satisfy as many connections as possible) from complete knowledge of each network fiber's active lightpaths.

Since the problem is provably NP-hard [169], [170], such techniques can work in reasonable time only for impractically small networks. Heuristics have thus to be found in order to make them work in practical-sized networks. In this context, linearized GN formulas have been proposed in order to effectively use the GN model within a mixed integer linear programming framework [170], [171].

Another line of developments explores instead a statistical approach to the RMSA problem, i.e., a model-based approach for lightpath selection which is extremely fast to compute but has a bounded probability of being incorrect, and thus may sometimes require path re-computation. For instance, in [172] a QoT-based routing method was proposed where a lightpath is set-up based solely on wavelength load measurements along its selected physical path, and the probability that the newly set-up lightpath is unfeasible is bounded below a desired threshold.

Another instance of the statistical RMSA approach is one that makes use of machine learning techniques for route selection. Here the GN/EGN model can be used to synthesize positive and negative examples of feasible lightpaths, and such examples are used to train the weights of a parametric model, for instance a neural network [173] or a random forest [161].

In summary, the use of the GN/EGN model for network planning and optimization is gaining momentum and is likely to become the method of choice for RMSA in future EoNs.

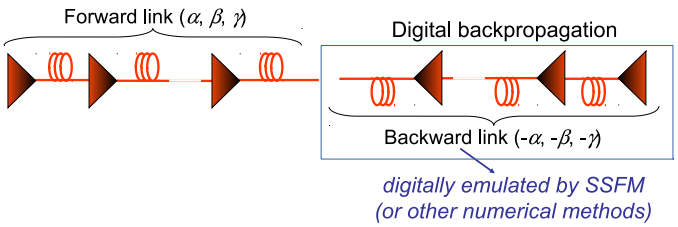


Fig. 19. Schematic representation of the DBP strategy

| | GVD compensation | DBP (per step) |
|---|---|---|
| Additions | $\frac{N}{N-N_m}(8\log_2 N+8)$ | $\frac{N}{N-N_m}(8\log_2 N+21+N_c)$ |
| Multiplications | $\frac{N}{N-N_m}(8\log_2 N+4)$ | $\frac{N}{N-N_m}(8\log_2 N+11+2N_c)$ |

TABLE II
NUMBER OF REAL OPERATIONS REQUIRED BY GVD COMPENSATION AND SSFM-BASED DBP (PER STEP)—$N_c=0$ FOR STANDARD SSFM, AND $N_c>0$ FOR FILTERED OR ENHANCED SSFM WITH $N_c$ COEFFICIENTS.

## XI. NONLINEARITY MITIGATION

This Section discusses the main approaches for nonlinearity mitigation based on DSP. The interested reader might also refer to [194] for an extensive review on this subject.

We point out that very recently artificial neural networks (ANNs) have also been advocated as a possible DSP approach for the mitigation of non-linear effects. We refrain from dealing with it here, but the interested reader can see for instance the recent prominent result [269].

We also point out that nonlinearity mitigation can be carried out by means of all-optical techniques as well. One example is for instance optical phase-conjugation [270]. All-optical mitigation techniques are a vast and interesting investigation field, but they are outside of the scope of this chapter.

### A. Digital back-propagation

One of the most promising and studied strategies to combat fiber nonlinearity is digital backpropagation (DBP). DBP is a channel inversion technique which aims at removing fiber propagation effects by digitally emulating the propagation of the received signal through a fictitious fiber link—equal to the actual fiber link but reversed in space and with opposite-sign parameters $-\alpha_P$, $-\beta_2$, and $-\gamma$ for each span of fiber [195], [196], [79]—as schematically depicted in Fig. 19. A simple analysis reveals that DBP can, in principle, exactly invert the propagation equation (1), hence removing all deterministic propagation effects due to the interplay between dispersion and nonlinearity. DBP can be implemented at the transmitter, at the receiver, or both. Several numerical studies and experimental demonstrations of DBP are available, reporting different gains in terms of SNR and reach, depending on the considered scenario [80], [71], [211], [210]. There is, anyway, a general consensus that gains up to 1 dB might be realistically achieved, and that DBP is a practical candidate for extending the reach of next-generation transponders.

In practice, DBP operates on a digitized version of the signal and is usually implemented by the SSFM algorithm (Section II) for its good characteristics in terms of performance and complexity. The sequence of received samples is divided into several overlapping blocks of length $N$, with $N_m$ overlapping samples. Each block is propagated through the whole link according to the algorithm in Fig. 3, with a total of $N_s$ steps. For a dual-polarization signal, the number of real additions and multiplications required by an SSFM-based implementation of DBP per each processed sample is reported in Table II [212]. The parameter $N_c$ accounts for alternative implementations of DBP (see later) and must be set to zero for a standard SSFM implementation; $N_m$ is of the order of channel memory induced by GVD (see Section II); and $N$ is optimized to minimize the overall computational complexity. For comparison, Table II shows the same figures also for GVD compensation based on a frequency-domain feed-forward equalizer. Each DBP step typically requires about 10–20% more operations than GVD compensation, such that the overall complexity of conventional DBP (based on standard SSFM and employing one step per span) easily exceeds a factor $\times 10$ the complexity of GVD compensation in terrestrial links [210].

Attaining better improvements with lower complexity is one of the big challenges to bring DBP from the lab to the field. A first solution is filtered DBP [198], [81], [197], in which the nonlinear step uses a lowpass filtered version of the signal power. The nonlinear step (5) is thus replaced by

$$A_{k,i}=A'_{k,i}e^{-\alpha_p h/2}e^{-j\gamma h_{\mathrm{eff}}\left[c_0|A'_{k,i}|^2+\sum_{\ell=1}^{N_c}c_i(|A'_{k-\ell,i}|^2+|A'_{k+\ell,i}|^2)\right]},\quad i=1,\ldots \tag{106}$$

where $c_o,\ldots,c_{N_C}$ are $N_c+1$ real coefficients of a tapped-delay-line low-pass filter. For $N_c=0$, the algorithm reduces to the standard SSFM. The idea is to overcome the inability of the standard SSFM to account for the continuously varying phase mismatch induced by GVD along each nonlinear propagation step—a problem that is particularly significant at large frequencies. Filtered DBP attempts to mitigate such a problem by low-pass filtering signal power during the nonlinear step in a way to limit the intensity fluctuations induced by the artificial phase matching condition of GVD [198]. Few taps are generally required in the additional filtering operation, such that the additional complexity per step (shown in Table II) is usually more than compensated for by the step elongation, with about one order of magnitude reduction of the overall complexity. A similar strategy is the one based on the enhanced SSFM algorithm, in which the nonlinear step is still replaced by (106), with the difference that the expression is derived from a perturbation analysis based on the FRLP model (Section VII), and the $N_c+1$ coefficients are optimized by a minimum mean square error criterion [213]. A single-step DBP achieving the same performance as a standard DBP with a 16 times higher complexity has been experimentally demonstrated [212].

Alternative DBP design methods showed that finely positioning the nonlinear step in a symmetric-SSFM implementation of DBP can enhance the algorithm performance [199]. Gonçalves et al. [205] showed that DBP can be aided by a memory polynomial model, a technique used in wireless communications to relax the requirements of Volterra equation. In dispersion managed links the strong correlation of the nonlinear interference from span to span can be efficiently

exploited to build a folded DBP where many spans are folded into a single span [206].

Low-complexity intrachannel nonlinearity mitigation techniques based on the perturbation analysis in [24] have been also proposed. In the special case of QPSK symbols, Tao et al. [202] showed how to exploit the constant modulus property of the constellation to build a multiplier-free intrachannel Volterra-based equalizer by substituting multiplications with summations. Tao et al. [200] proposed to reduce the number of multiplications in the perturbation evaluation by quantization of the kernel function. By using 50% of pre-compensation the number of quantized coefficients has been shown to be reduced down to few units still keeping significant gains [201]. Gao et al. [203] showed that with 50% of pre-compensation the symmetry of the dispersion map with respect to the midpoint of transmission link can be used to reduce the number of multiplications in a perturbation-based nonlinear pre-compensation. The reason is that this way the kernel coefficients are i) real-valued, thus halving the number of complex multiplications and ii) evaluated over half link length. A reduction factor of 6.8 has been experimentally demonstrated. Although the kernel description of the perturbation is a global input/output description of the link, it has been shown that implementing the perturbative nonlinear equalization by a multi-stage compensation cascade may relax complexity at a given accuracy [207]. Indeed, when dividing the link in $N$ stages, the equalization procedure is repeated $N$ times, once per stage. Nevertheless, since the total number of nonlinear interactions per stage decreases *quadratically* with $N$ due to the shorter memory of each stage, the overall complexity is reduced. A comprehensive analysis of the symmetries in Volterra-based nonlinear equalization that help practical implementation has been investigated in [204]. An analytical expression for the computation of the coefficients of perturbative nonlinear precompensation for Nyquist pulses was derived and experimentally verified in [208]. The gains achievable by DBP or perturbative nonlinear compensation were investigated and compared in a C+L band ultralonghaul transmission experiment in [209].

Besides computational issues, and despite the theoretical possibility of an exact channel inversion, both fundamental and practical limitations prevent DBP from completely removing nonlinear impairments. A first limitation arises in WDM systems, which are impaired by both intra- and inter-channel nonlinearity. In typical systems, each channel is separately detected and processed, such that only a *single-channel DBP* can be actually implemented, effectively compensating only for intra-channel nonlinearity. In principle, it is possible to implement a *multi-channel DBP* by jointly detecting and processing two or more WDM channels [81], [211]. As the number of backpropagated channel increases, a higher portion of inter-channel nonlinearity can be effectively mitigated. However, the required computational resources grow more than linearly with the number of backpropagated channels.

The peak-SNR gains that can be theoretically achieved by ideal single- and multi-channel DBP over typical dispersion-unmanaged WDM systems are shown in Fig. 20 as a function of the number of spans [4], for systems operating at 32 GBaud.

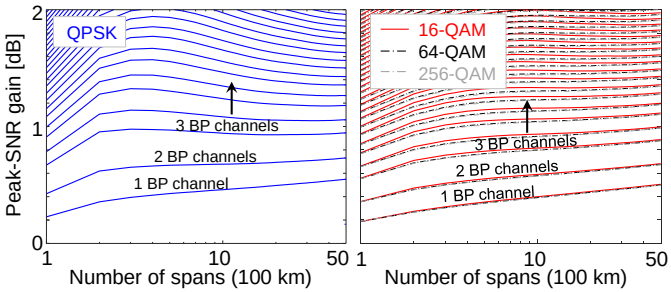


Fig. 20. Peak-SNR gains achievable by single- and multi-channel DBP over a fully-loaded (115 channels) dispersion-unmanaged WDM system over SMF, with channel symbol rate 32 GBaud. The system parameters are set as in [4].

The various solid curves correspond to the number of channels that are jointly backpropagated. The potential gain of single- and 3-channel DBP is limited to about 0.5 dB and 1 dB, respectively. Beyond 3-channel DBP, each additional backpropagated channel provides an incremental gain of no more than $\sim$0.1 dB, possibly not enough to justify the corresponding increase of the required computational resources. Slightly higher gains are achievable when considering systems with inter-channel phase noise and polarization rotation compensation [4]. Moreover, in optically routed networks, signals at different wavelengths may follow different paths in the network, such that an effective multi-channel DBP is no longer possible.

The possibility to obtain a perfect channel inversion may be limited also by the presence of PMD [82]. In principle, an exact knowledge of the PMD evolution both in time and along the fiber link would still allow for a perfect channel inversion without any performance degradation. In practice, such a knowledge is unavailable both at the transmitter and receiver, and DBP is usually performed by ignoring the presence of PMD (which is separately compensated for by an adaptive linear equalizer after DBP). In typical WDM systems, the impact of PMD is negligible for single-channel DBP, but becomes relevant when increasing the number of backpropagated channels, practically limiting the achievable SNR gains [211]. Some possible countermeasures, based on the insertion of PMD sections among the DBP steps, have been investigated in [215], [214], [216], considering different strategies to configure the PMD sections.

Eventually, the nonlinear compensation achievable by DBP is fundamentally limited by signal-noise interaction, for which a channel inversion strategy turns out to be ineffective. This is easily understood by noting that the ASE noise injected by each in-line amplifier interacts with the propagating signal along a different portion of the optical link, such that DBP reduces the signal-noise interaction due to the first amplifiers, but enhances that due to the last amplifiers. The impact of signal-noise interaction on DBP is generally negligible compared to that of inter-channel nonlinearity and of PMD. It is, therefore, of little practical relevance in real systems. However, it becomes essential when investigating the ultimate limitations of optical fiber systems—for instance when considering single-channel scenarios or, equivalently, in the case of an ideal multi-channel DBP covering the whole WDM spectrum. This kind of

analysis is complicated by the unavailability of an exact model of the optical fiber channel accounting for signal-noise interaction, though some approximated models, based on perturbation theory, are indeed available, e.g., [140], [33], [141], [142], [27], [145], [148], which would be required to optimize the detection strategy, as discussed more in detail in Sect. XII. In fact, DBP is not sufficient to implement an optimum detector, and alternative strategies have been proposed to replace or complement DBP [218], [217], [219]. One possible approach is that of extending the DBP concept to include the uncertainty due to amplifier ASE noise when backpropagating the received signal. This procedure, known as stochastic DBP (SDBP), consists in backpropagating distributions rather than signals, eventually obtaining an estimate of the posterior distribution of the transmitted symbols given the received signal, which is required to implement an optimum detector [217]. The specific SDBP algorithm proposed in [217] is based on a particle representation of the distributions. It is too complex for a practical implementation, but offers a benchmark for simpler implementations.

Most DBP demonstrations assume that the required system parameters (e.g., launch powers and fiber parameters for each span of fiber) are known at the receiver. Often, this is not the case. Moreover, some parameters may change over time due, for instance, to network reconfigurations. It is therefore essential to devise adaptive strategies for the configuration and dynamic control of DBP parameters [220], [221]. While there are no particularly stringent requirements in terms of convergence and adaptation speed for standard DBP algorithms, this issues becomes critical when considering PMD-aware DBP algorithms.

### B. Symbol Rate Optimization

One of the features of nonlinear fiber propagation is that the amount of generated NLI power depends on the symbol rate of the WDM channels. Specifically, it was shown simulatively [84], [85] and then experimentally [83] that, by distributing the same total data throughput across the same optical bandwidth, over a *larger* number of WDM channels at a *smaller* symbol rate, the NLI power impairing reception would *decrease*. Initial theoretical studies concurred [86], [104]. These findings were then investigated by means of the EGN-model [268]. Further investigation was carried out in [4], using both the time-domain and the pulse-collision models, which also clarified the role and impact of NLPN in this phenomenon. Interestingly, SRO was predicted by the EGN model results to actually slightly increase its effectiveness when the overall system WDM bandwidth is widened [268], contrary, for instance, to DBP. Due to all these findings, the idea of reducing the symbol rate per carrier in order to improve system performance has been considered as a potential NLI mitigation technique, under the acronym SRO (symbol rate optimization).

SRO theory predicts that the optimum symbol rate is, in uniform links of $N_s$ spans, each of length $L_s$ [268]:

$$R_{\rm opt} = \sqrt{2/(\pi\,|\beta_2|\,L_{\rm s}\,N_{\rm s})} \qquad (107)$$

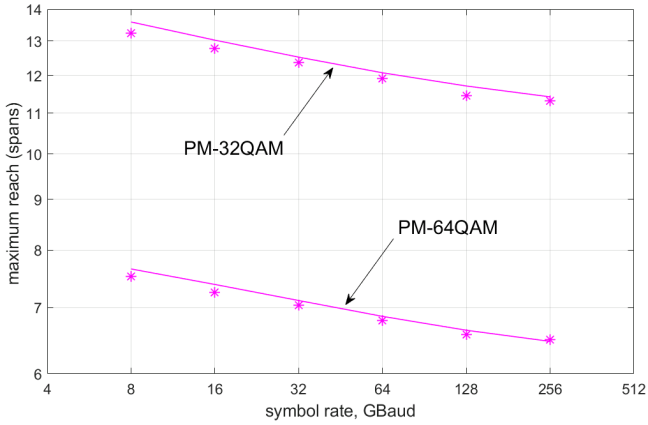


Fig. 21. Maximum reach vs. per-channel symbol rate. Solid lines: calculations using the EGN-model. Markers: simulations. Span-length 100 km, SMF, EDFA noise figure 6 dB, roll-off 0.2. Target performance: GMI 8.7 and 10.44 bits/symb, for PM-32QAM and PM-64QAM, respectively. The total optical bandwidth is 1.83 THz for all rates. For symbol rates from 8 to 256 GBaud, the number of channels ranges from 192 to 6. Carrier spacing: $\Delta f = 76.2\frac{R_s}{64}$, where $R_s$ is the symbol rate in GHz.

Values typically range between 1 and 10 GBaud. These rates are quite low as compared to industry standards. In addition, the industrial trend, for techno-economical reasons, has historically been that of constantly ramping up symbol rates, with a transition between 32 to 64 GBaud currently taking place. This circumstance clearly clashes with the SRO prescription of operating at low symbol rates.

One way to make an increase in *total* symbol rate per transponder compatible with a decrease of symbol rate *per carrier* is that of resorting to *subcarrier multiplexing*. Essentially, high-symbol-rate channels are created as a collection of DAC-generated electrical subcarriers, which operate at the symbol-rate that is optimum from the viewpoint of NLI mitigation. For instance, a 64-GBaud channel could consist of sixteen 4-GBaud subcarriers.

No closed-form formula currently exists for the prediction of the potential MR gain. Estimating it requires the numerical evaluation of the EGN model. Incidentally, SRO is one of those contexts where the GN-model is inadequate, as suggested by the strong violation of criterion (1), which prescribes high symbol rate for GN-model accuracy (see Sect. IV-G1). Direct EGN (or time-domain model) calculations or Monte-Carlo simulations are required.

Fig. 21 provides a pictorial appreciation of the predicted MR gain for two high-cardinality systems, either PM-64QAM or PM-32QAM. The relevant system data is reported in the figure caption. Both the solid line (EGN-model) and the markers (simulations) concur in predicting about 16%-18% MR increase between the figure extremes, i.e., from 256 GBaud down to 8 GBaud. The latter is close to the theoretical optimum, which according to Eq. (107) is about 5 GBaud for PM-32QAM and about 6.5 GBaud for PM-64QAM. Note that in order to clearly show an optimum, Fig. 21 would have to be pushed down to about 2 GBaud. This was not possible

because the CPU time of the simulations as well as the their memory occupation is inversely proportional to the symbol rate. For the almost 2 THz total optical bandwidth considered in Fig. 21, 8 GBaud was the practical limit. It was done instead in [268], Figs. (1) and (4), because the total bandwidth was 500 GHz. There, the optimum is clearly visible.

The MR percentage gain due to SRO is actually larger for longer links and for smaller constellations [268],[4], with PM-QPSK being the format benefiting the most, with potential MR gains on the order of 25%-30% across the same symbol rate range as in Fig. 21. Also, there are important aspects as to the role played by NLPN in the context of SRO. For more details on this, see [4].

In addition, when GSCs are used, the gain due to SRO essentially vanishes, as predicted by the GN-model (which is equivalent to the EGN-model for GCSs). Part or all of the gain could be gained back for GSCs if NLPN mitigation was possible at low symbol rates. This is currently unclear and the subject of ongoing research [4].

Despite all the limitations, SRO could be a possible way to partially mitigate NLI, at least for QAM-based systems, with relatively low complexity. In [268] it was shown by EGN model calculations that for full C-band ultra-long-haul systems, at 32 Gbaud, its MR gain could be comparable to DBP.

It should however be mentioned that negative findings regarding SRO have been published too. In particular [87] found very limited gain from SRO, which also tended to vanish as the number of WDM channels was increased, in contrast with SRO theory. Overall, a conclusive specific experiment proving the effectiveness of SRO over a fully-populated C-band system is not yet available. Therefore, SRO should still be considered a research topic, needing confirmation. We should also mention that presently there are commercial systems that use multi-subcarrier transmission. Whether its use grants actual performance gain in these commercial systems, and why, is currently undisclosed information.

### C. Nonlinear Fourier transform

The nonlinear Fourier transform (NFT)—also known in the mathematical community as inverse scattering transform—is a mathematical tool to solve a certain class of nonlinear partial differential equations—the so called *integrable* equations. The NFT was introduced in [222] for the Kortweg–de Vries equation and later applied to many other equations, including the NLSE [223] and the Manakov equation [224]. A general treatise on the NFT can be found in [225], while a more specific review of the NFT for the NLSE and its application to optical fiber communications can be found in [226], [227].

The NFT can be regarded as the generalization of the ordinary Fourier transform to nonlinear systems [225]. The basic idea is that of representing the propagating waveform through a proper set of spectral data—the scattering data or *nonlinear spectrum*—whose evolution along the fiber link is simple and linear. As with linear systems and the ordinary Fourier transform, the initial value problem is therefore solved by performing three main steps: :

1) Computation of the input nonlinear spectrum by a direct NFT. For the NLSE, this operation consists in solving the Zakharov–Shabat spectral problem [223]. In general, the Zakharov–Shabat operator has both a discrete and a continuous spectrum. The former consists of a set of eigenvalues in the complex plane and the related norming constants, and is associated with *soliton* components. The latter is defined over the real line and is analogue to the ordinary Fourier transform, to which it converges in the low-power limit.
2) Propagation of the nonlinear spectrum. For the NLSE, this is equivalent to the propagation of the linear spectrum in a linear dispersive fiber.
3) Reconstruction of the output waveform from its nonlinear spectrum by an inverse NFT (INFT). For the NLSE, this is classically done by solving the Gelfand–Levitan–Marchenko integral equation [223], [225].

Many integrable equations, including the NLSE, have *soliton* solutions, which play an important role in the NFT theory. Solitons maintain their shapes (or return periodically to it) during propagation thanks to a cancellation of nonlinear and dispersive effects. This property makes them attractive for communications. Indeed, soliton-based communications have been extensively explored during the eighties and nineties (see [228] and references therein). However, they never made their way to commercial systems because of their low spectral efficiency (about 0.2 bit/s/Hz) and the detrimental impact of the Gordon–Haus effect and inter-soliton interaction.

Soliton communication is just the simplest way in which NFT-related concepts can be used to design nonlinearity-tolerant communication systems. A more general approach, originally named *eigenvalue communication* [229], is that of avoiding nonlinear interference by encoding information on the spectral data of the Zakharov-Shabat operator. After going almost unnoticed for a couple of decades, this approach has been recently resurrected—with different names and flavors such as *eigenvalue communication* [234], *nonlinear frequency–division multiplexing (NFDM)* [226] and *nonlinear inverse synthesis* [230]—thanks to the impressive progress in DSP technology, which makes conceivable, if not yet feasible, a real-time implementation of the NFT.

All the NFT-based approaches proposed so far can be generally represented by one of the four schemes depicted in Fig. 22 (the first three are discussed also in [227]). In the standard NFDM scheme of Fig. 22(a), information is encoded and decoded in the nonlinear frequency domain, using INFT and NFT to convert the nonlinear spectrum into a time-domain waveform and the other way around, respectively [229], [226], [230], [231], [237]. On the other hand, in the scheme of Fig. 22(b), information is encoded and decoded in time domain, as in conventional systems, while NFT and INFT are used to implement DBP [232]. Fig. 22(c) represents a hybrid approach in which information is encoded in time domain and decoded in nonlinear frequency domain [234], [233]. Finally, Fig. 22(d) shows the complementary hybrid approach in which information is encoded in nonlinear frequency domain and decoded in time domain [236], [235]. While the first two

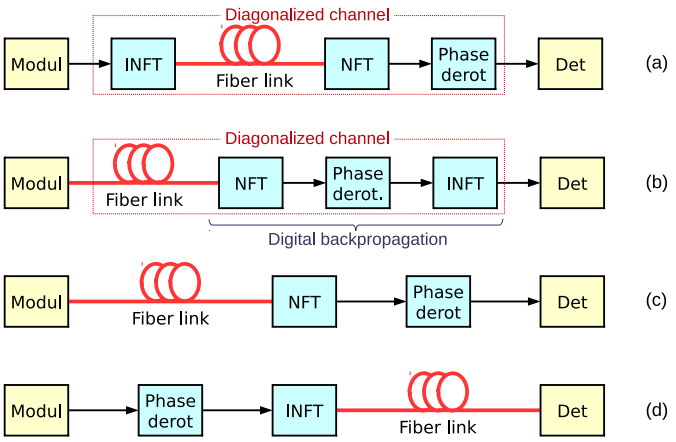


Fig. 22. Different ways to use NFT in optical communications: (a) Modulation and detection in nonlinear frequency domain; (b) modulation and detection in time domain, with NFT-based DBP; (c) modulation in time domain and detection in nonlinear frequency domain; (d) modulation in nonlinear frequency domain and detection in time domain.

approaches do achieve a diagonalization of the channel, such that symbol-by-symbol detection can be employed, the last two do not, in fact requiring some kind of sequence detection. Quite interestingly, time-domain detection of NFDM signals (fourth approach) seems to achieve a superior performance compared to the conventional NFDM scheme (first approach), though with some additional complexity, probably due to a better matching between the employed detection metrics and the actual channel statistics [235].

NFDM schemes can be further classified according to which part of the nonlinear spectrum is used to encode information: the discrete part [229], [234], [226], [236], the continuous part [230], [226], [237], [235], or both [238]. Furthermore, the NFT theory can be applied also to the Manakov equation, such that the proposed schemes can be extended to consider polarization-multiplexed signals [242], [243].

As for the ordinary Fourier transform, the NFT and INFT can be computed analytically only in a few simple cases. In general, one has to resort to numerical methods. An overview of available numerical methods can be found in [226], [227]. While the complexity of the most classical methods usually scales at least quadratically with the number of processed samples, super fast algorithms with log–linear complexity (as the FFT) have been also developed [239]. Research of low-complexity methods with general validity is still ongoing and will play an important role in determining the feasibility of NFT-based techniques.

The NFT theory and the techniques mentioned in this section are based on *vanishing boundary conditions* for the optical signal, which is assumed to decay sufficiently fast as $t \to \pm\infty$. In a real transmission system, this condition can be emulated by operating in burst mode and inserting a sufficiently long guard time between bursts [230]. The NFT theory has been developed also under different boundary conditions. A typical case, extensively studied in the literature, is that of *periodic boundary conditions*. This approach may offer some advantages compared to the case of vanishing boundary conditions [227], but its application to optical communications has been much less explored [240].

Research about NFT-based techniques is still at an early stage and there is not a general consensus about their potentials, perspectives, and suitability for optical communications. Research is mostly following a trial-and-error approach, and the lack of a unified theory makes hard to understand which strategy (among those depicted in Fig. 22) is the most promising, which part of the spectrum should be modulated, how information should be encoded on it, and what kind of boundary conditions should be considered. Several experimental demonstrations of the schemes in Fig. 22 can be found. Yet, so far, there is no clear evidence of a performance improvement compared to conventional systems. In fact, some critical issues which seriously hamper the use of NFT operations at high power [241] and the achievement of high spectral efficiencies [237] have been highlighted. Performance needs to be significantly improved and complexity reduced before NFT-based techniques can make their way to real systems. Moreover, some propagation effects, which are not included in the NLSE (1) but can be easily described and coped with in time domain (e.g., attenuation, higher-order dispersion, PMD), break the integrability of the system and are extremely hard to model in the nonlinear frequency domain. Their impact on NFDM systems needs to be better investigated, as it might be potentially more detrimental than the impact of nonlinear effects on conventional systems. Yet, the promise of a tool able to harness fiber nonlinearity keep research alive.

### *D. Other nonlinear compensation techniques*

While DBP is certainly one of the most studied nonlinear compensation techniques, other techniques have been also proposed in the literature over the past years, ranging from the quite practical and simple approach of SRO to the more visionary techniques based on the NFT, both reviewed in this Section. Nonlinearity compensation based on inverse Volterra series transfer function [204] has been also studied. Given the equivalence between the RP method and the Volterra series approach, discussed in Section III, this technique can be included in the broader category of perturbation-based channel inversion techniques for alternative DBP implementations.

According to the time-domain perturbative model and the FRLP model, introduced in Sections V and VII, respectively, interchannel nonlinearity causes the emergence of phase noise, polarization rotations, and ISI. These effects can be partly mitigated by some classical algorithms typically employed to counteract them, when generated by causes other than fiber nonlinearity. These algorithms can hence be considered in all respects as nonlinearity mitigation strategies. Some examples are reported in [244], [67], [8], [10].

Another class of nonlinearity mitigation techniques aims at reducing the amount of generated NLI by a proper combination of coding and modulation. Possible examples are the design of nearly constant-envelope modulation formats [245]; the ad hoc probabilistic shaping of QAM constellations over one [246] or several time slots [247]; and the use of twin waves [248] or conjugate data repetition [249] to partly cancel NLI.

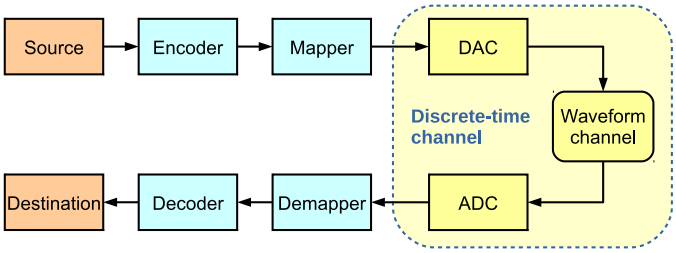


Fig. 23. Schematic representation of a digital communication system

## XII. Capacity of the nonlinear channel

### A. Capacity and spectral efficiency

One of the key results of Claude Shannon's pioneering work was demonstrating that a reliable communication over a noisy channel is possible, provided that the information rate is less than a characteristic quantity determined by the statistical properties of the channel, which he named *channel capacity* [77]. Other than laying the foundations of what is now called *information theory* and establishing such a general result, Shannon also derived a specific closed-form expression for the capacity of an AWGN channel. As we shall see, this expression is widely employed also in the context of optical fibers, though its validity and interpretation need some care. Shannon's work has been since extended and generalized both to account for a broader class of channels and to obtain closed-form capacity expressions for other specific channels (see [181] and references therein). In this context, the optical fiber channel is a major challenge due to its peculiar nonlinear behavior. In fact, as we shall see in this section, research to determine channel capacity is still ongoing and the problem, so far, remains essentially open [182].

We start by considering the generic digital communication system schematically depicted in Fig. 23. Firstly, the channel encoder takes a stream of information bits from the source and adds some redundancy to it (e.g., parity check bits) to be used for error correcting purposes. Then, the encoded bits are mapped onto a sequence of information symbols belonging to the input modulation alphabet $\mathcal{X}$. Since the physical channel is usually a *waveform channel* (e.g., an optical fiber link), a DAC converts the sequence of symbols into a waveform (e.g., by linearly modulating a sequence of pulses of given shape) that propagates through the channel. The opposite operations are then performed at the receiver: the output waveform (possibly distorted and corrupted by noise) is converted to a sequence of symbols belonging to the output alphabet $\mathcal{Y}$ (e.g., by filtering and sampling the waveform). The detector converts the output symbols into a sequence of bits, which are finally decoded (correcting possible errors) and sent to destination. The combination of DAC, waveform channel, and ADC constitutes the *discrete-time channel*, which is usually the subject of the information theoretical analysis. In practical cases, the input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$ are determined by the quantization levels available in the DAC and ADC and are, hence, finite. Often, when no constraints about the DAC and ADC characteristics are given, the whole field of complex numbers is considered for both alphabets. The discrete-time channel is characterized by a family of conditional distributions $p(\mathbf{y}_N|\mathbf{x}_N)$, for $N = 1, 2, \dots$ , where $\mathbf{x}_N = (x_1, x_2, \dots, x_N)$, with $x_i \in \mathcal{X}$, and $\mathbf{y}_N = (y_1, y_2, \dots, y_N)$, with $y_i \in \mathcal{Y}$, are length-$N$ realizations of the input and output processes $X = X_1, X_2, \dots$, and $Y = Y_1, Y_2, \dots$, respectively.

A typical problem in information theory is that of determining channel capacity—operationally defined as the maximum bit rate at which information can be reliably (i.e., at an arbitrarily low error probability) transmitted through the channel with the best possible combination of coding and modulation—without actually trying such (infinitely many) possible combinations. Given the statistics of the input process, i.e., the input distributions $p(\mathbf{x}_N)$ for $N = 1, 2, \dots$, the maximum bit rate at which information can be reliably transmitted through the channel is given by the information rate [183]

$$I(X;Y) = \lim_{N\to\infty} \frac{1}{N} I(\mathbf{X}_N; \mathbf{Y}_N) \tag{108}$$

where

$$I(\mathbf{X}_N; \mathbf{Y}_N) = E\left\{ \log_2 \frac{p(\mathbf{Y}_N|\mathbf{X}_N)}{p(\mathbf{Y}_N)} \right\} \tag{109}$$

is the mutual information between $\mathbf{X}_N$ and $\mathbf{Y}_N$, $E\{\cdot\}$ being the expectation operator. The information rate (108) depends both on the input statistics—which are determined by the channel user by selecting a combination of code and modulation—and on the channel itself. Eventually, the capacity of the channel is obtained by selecting the input statistics for which the information rate is maximized

$$\mathcal{C} = \lim_{N\to\infty} \frac{1}{N} \sup_{\mathbf{X}_N} I(\mathbf{X}_N; \mathbf{Y}_N) \tag{110}$$

where the supremum $\sup_{\mathbf{X}_N}$ is taken with respect to all input distributions $p(\mathbf{x}_N)$ satisfying a specific constraint (usually on the average power) [183]. The capacity is, hence, a specific property of the channel which determines a fundamental limit for transferring information.

Both quantities in (108) and (110) refer to the discrete-time channel and are measured in bit/symbol (or bit/channel). The corresponding information rate and capacity of the waveform channel are expressed in bit/s and are obtained by multiplying (108) and (110) times the baudrate $R$ at which symbols are mapped to and demapped from the input and output waveforms. A further normalization by the *occupied bandwidth* (whenever a suitable definition of it is available) yields the *spectral efficiency* (SE) in bit/s/Hz.

In general, only the capacity of some specific channels can be evaluated analytically, the AWGN channel being perhaps the most notable example [77]. A discrete-time AWGN channel is characterized by the input-output relation $y_k = x_k + n_k$, where the noise samples $n_k$ are realizations of i.i.d. variables with a zero-mean proper complex Gaussian distribution [184]—often referred to as circularly-symmetric complex Gaussian (CSCG)—with variance $P_n$. The analysis in this case is greatly simplified by the absence of memory, which allows to compute (109) by considering only a single pair of input and output symbols ($N = 1$). Given a constraint on the average signal power $P_s$, the capacity of the AWGN channel

is given by the well known expression $\mathcal{C} = \log_2(1 + P_s/P_n)$ and is achievable by i.i.d. CSCG input symbols [77]. As the capacity achieving input distribution and the corresponding output distribution are zero-mean, this capacity can also be expressed in terms of the input variance $\sigma_x^2 = P_s$, output variance $\sigma_y^2 = P_s + P_n$, and covariance $\sigma_{xy} = P_s$ as

$$\mathcal{C} = \log_2(1 + P_s/P_n) = \log_2\left(\frac{\sigma_x^2\sigma_y^2}{\sigma_x^2\sigma_y^2 - |\sigma_{xy}|^2}\right) \qquad (111)$$

In this case, it is apparent that, since the noise power $P_n$ is fixed and independent of the signal, the channel capacity grows unbounded with signal power $P_s$. This means that the channel capacity is limited only if the available resources (power) are limited (e.g., due to economic or technological constraints).

When turning our attention to the optical fiber channel, the picture is more complicated. At low signal power, the nonlinear term of the NLSE (1) is practically negligible and the channel essentially behaves like an AWGN channel, (Linear propagation impairments, such as chromatic dispersion and polarization mode dispersion, are characterized by a slowly varying (or constant) unitary transfer matrix. Hence, in the linear regime, their presence has no impact on channel capacity as they can be fully compensated by a linear equalizer.) with a capacity that grows logarithmically with launch power according to (111). However, at higher signal power, the nonlinear term of the NLSE is no longer negligible and must be accounted for in the evaluation of channel capacity. This is an extremely difficult task. First of all, an exact and mathematically tractable expression for $p(\mathbf{y}_N|\mathbf{x}_N)$ is unavailable, making the analytical evaluation of (108) unfeasible. Moreover, its numerical estimation, which must be performed in a $4N$-dimensional space, is practically limited to very small values of $N$. On the other hand, the convergence of (108) and (110) to their actual limit can be expected when $N$ is of the order of the channel memory, which can be many hundreds of symbols in typical dispersion-unmanaged systems. For the same reasons, also the optimization of $p(\mathbf{x}_N)$ in (110) cannot be performed analytically and, when resorting to numerical approaches, it is limited to small values of $N$. A further issues arises when considering the discretization of the waveform channel. In fact, due to nonlinearity, the propagating signal is subject to spectral broadening and signal-noise interaction during propagation. Therefore, the structure and bandwidth of an optimum demodulator (providing a sufficient statistic) is also unknown. Moreover, since the signal bandwidth changes during propagation, there is not a general consensus on how the spectral efficiency should be defined [185].

Typical optical systems employ WDM, in which many transmitter/receiver pairs share the same fiber to transmit and receive signals at different wavelengths. This multi-user scenario requires some extra care to be modeled and analyzed. A typical and reasonable assumption is that each user has access only to its allotted portion of the WDM spectrum, and that all the signals are independently generated and detected. In this situation, the capacity analysis is performed from a single-user perspective, i.e., by focusing on a single WDM channel (user), typically referred to as the COI, and considering all the other channels as a source of possible interference. Moreover, a specific behavioral model for the interfering channels, i.e., a rule that relates the input distribution on the interfering channels to the COI input distribution, should be specified [186]. The most common (and fair) assumption is that all the users transmit with same input distribution and power. Different choices are possible, corresponding to different network scenarios and leading to completely different results in terms of capacity [186]. Eventually, in the WDM scenario, spectral efficiency is usually defined by considering that the bandwidth occupied by each channel equals channel spacing.

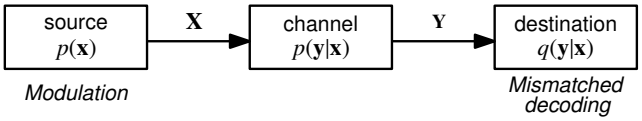


Fig. 24. Discrete-time channel with mismatched decoding.

All the difficulties highlighted in this subsection make the exact evaluation of the capacity of the optical fiber channel extremely hard. In cases like this, a typical approach is that of resorting to the computation of capacity upper and lower bounds, trying to identify a possibly narrow range of values in which the actual (unknown) capacity lies. This will be discussed in the next subsections.

### *B. Achievable information rate with mismatched decoding*

When the actual channel law is unknown, such as in the case of the optical fiber channel, it is useful to consider the more general case, depicted in Fig. 24, of a discrete-time channel with a detector that makes maximum-a-posteriori-probability (MAP) decisions based on a mismatched channel law $q(\mathbf{y}_N|\mathbf{x}_N) \neq p(\mathbf{y}_N|\mathbf{x}_N)$. Following [187], we define the *achievable information rate* (AIR) with mismatched decoding metric $q(\mathbf{y}_N|\mathbf{x}_N)$ as (In [187], this quantity is actually referred to as the *auxiliary-channel lower bound* to the information rate.)

$$I_q(X;Y) \triangleq \lim_{N\to\infty} \frac{1}{N} E\left\{\log \frac{q(\mathbf{y}_N|\mathbf{x}_N)}{\int p(\mathbf{x}_N)q(\mathbf{y}_N|\mathbf{x}_N)d\mathbf{x}_N}\right\} \qquad (112)$$

With respect to the average mutual information rate (108), (112) is obtained by replacing the actual channel law $p(\mathbf{y}_N|\mathbf{x}_N)$ with an arbitrary mismatched law $q(\mathbf{y}_N|\mathbf{x}_N)$, while the expectation is still taken with respect to the actual true distribution $p(\mathbf{y}_N|\mathbf{x}_N)p(\mathbf{x}_N)$ induced by the input distribution and the actual channel law.

The AIR (112) has some interesting properties which hold for any true and auxiliary channel and make it suitable for a practical use in optical communications: i) It is a lower bound to the information rate and, therefore, to channel capacity $I_q(X;Y) \leq I(X;Y) \leq \mathcal{C}$; ii) its maximization over any possible detection law (obtained for $q(\mathbf{y}_N|\mathbf{x}_N) = p(\mathbf{y}_N|\mathbf{x}_N)$) leads to the actual information rate; iii) its further maximization over the input distribution $p(\mathbf{x}_N)$ leads to channel capacity; iv) it is achievable over the true channel with source probability $p(\mathbf{x}_N)$ and a MAP detector matched to $q(\mathbf{y}_N|\mathbf{x}_N)$; v) it can be computed through numerical simulations, without an explicit

knowledge of the true channel law $p(\mathbf{y}_N|\mathbf{x}_N)$, provided that $q(\mathbf{y}_N|\mathbf{x}_N)$ can be computed.

In practice, the system is designed by selecting a modulation format, which determines the input distribution $p(\mathbf{x}_N)$, and an approximated (auxiliary) channel model (e.g., one of those discussed in Section III), which determines the mismatched channel law $q(\mathbf{y}_N|\mathbf{x}_N)$ and, hence, the detection metrics. The AIR (112) for this configuration is then computed through numerical simulations. Possibly, $p(\mathbf{x}_N)$ and/or $q(\mathbf{y}_N|\mathbf{x}_N)$ can be numerically optimized by using the AIR as a performance metric to be maximized.

While an exact analytical evaluation of (112) is still unfeasible—as the joint distribution $p(\mathbf{x}_K, \mathbf{y}_K)$ of the true channel, with respect to which the expectation in (112) must be computed, is generally unknown—an accurate numerical estimate can be efficiently obtained by relying on the asymptotic equipartition property [122] and following the procedure described in [187]:

1) Draw a long input sequence $\mathbf{x}_N = (x_1, \ldots, x_N)$ of samples from the selected input distribution $p(\mathbf{x}_N)$.
2) Compute the corresponding output sequence $\mathbf{y}_N = (x_1, \ldots, x_N)$ by using the SSFM to simulate the true channel.
3) Compute the selected conditional distribution $q(\mathbf{y}_N|\mathbf{x}_N)$.
4) Compute the corresponding output distribution $q(\mathbf{y}_N) = \int p(\mathbf{x}_N) q(\mathbf{y}_N|\mathbf{x}_N) d\mathbf{x}_N$.
5) Estimate the AIR as

$$\hat{I}_q(X;Y) = \frac{1}{N} \log \frac{q(\mathbf{y}_N|\mathbf{x}_N)}{q(\mathbf{y}_N)}$$

For finite-state source and auxiliary channel models, the actual computation of $q(\mathbf{y}_N|\mathbf{x}_N)$ and $q(\mathbf{y}_N)$ can be efficiently carried out by the sum-product algorithm on a suitably defined factor graph [187]. The procedure can be extended to the more general case of a continuous state space by resorting to particle methods [188].

### C. The nonlinear 'Shannon' limit (NSL) lower bound

As already mentioned in the previous subsection, at low signal powers the optical fiber channel behaves like an AWGN channel; in fact, modern optical systems can achieve information rates close to channel capacity (111) when working in such a linear regime [41]. However, contrarily to the AWGN case, when the signal power increases, the ensuing nonlinearities impair the existing systems to the point that they cease to work. This naturally poses questions about the impact of nonlinearities on the capacity of a fiber-optic channel and the possible existence of an ultimate limit to channel capacity.

A variety of studies and facts hint at the possibility that a finite limit to the fiber-optic channel capacity does exist, so much that a name for it was coined—the *nonlinear Shannon limit* (NSL) [41]. To the best our knowledge, a finite capacity limit for a fiber-optic channel was predicted for the first time in [30]. Similar results were later obtained by using different models and either analytical approximations or numerical methods [32], [34], [68], [46], [26]. Such a convergence of results has induced the belief that the actual channel capacity is very close to the NSL or, at least, follows the same trend. However, an analysis of the literature reveals that the NSL, as defined in [32], [41], but also as computed in many other publications [34], [68], [26], is just an instance of the AIR (112), computed with specific input distributions (usually i.i.d. symbols with CSCG or uniform-ring distribution) and mismatched channel laws (usually memoryless Gaussian). Therefore, it is only a lower bound to channel capacity, whose tightness is not known a priori.

A special case of (112)—referred to as the Gaussian AIR in the sequel—is obtained when considering i.i.d. CSCG input variables and a detector matched to an AWGN channel with the same input-output covariance matrix of the real channel

$$I_G(X;Y) = \log_2 \left( \frac{\sigma_x^2 \sigma_y^2}{\sigma_x^2 \sigma_y^2 - |\sigma_{xy}|^2} \right) \tag{113}$$

This result shows that the Gaussian AIR can be computed for *any* channel by simply estimating—either analytically, through an approximated channel model, or numerically, through accurate but time consuming SSFM simulations—a covariance matrix. Remarkably, (113) equals the expression of the AWGN channel capacity (111). This means that the same expression can be used (and, in fact, it is commonly used) for any channel, but with different meanings: it gives the true channel capacity for AWGN channels, while it provides only a lower bound (achievable by a conventional detector optimized for the AWGN channel) for generic channels [192]. In fact, the information theoretical limits computed in [30], [32], [46], [26], [51] are all obtained by using (113), though computing the covariance matrix with different approaches. A good agreement between those results can be generally observed.

When considering the GN model in Section ?, (113) reduces to the simple expression

$$I_G(X;Y) = \log_2 \left( 1 + \frac{\mathrm{P_s}}{P_{\mathrm{ASE}} + P_{\mathrm{NLI}}} \right) \tag{114}$$

where $P_{\mathrm{ASE}}$ is the power of the accumulated ASE noise over the signal bandwidth $W$, and $P_{\mathrm{NLI}}$ is the nonlinear interference power obtained by integrating (34). Note that in this case, since the input process is Gaussian by assumption, the EGN model gives exactly the same result, as the correction term vanishes.

An alternative closed-form expression is obtained with the FRLP model in Section VII. In this case, (113) reduces to

$$I_G(X;Y) = \log_2 \left[ 1 + \frac{P_s e^{-\overline{\sigma_\theta^2}}}{P_{\mathrm{ASE}} + P_s (1 - e^{-\overline{\sigma_\theta^2}})} \right] \tag{115}$$

where $\overline{\sigma_\theta^2}$ is the effective variance of the XPM term $\theta(f,t)$ over the signal bandwidth [59]. A very good agreement between (115) and the numerical computation of (113) based on SSFM simulations has been demonstrated [59]. A similar expression, with the same functional dependence on signal power $P_s$, can be obtained through the Feynman's path integral approach [32]. Moreover, when $\sigma_\theta^2 \ll 1$—i.e., in most cases of practical interest—(115) reduces to (114), with $\sigma_\theta^2$ replacing $P_{\mathrm{NLI}}$. In this case, the scaling of (114) and (115) with system

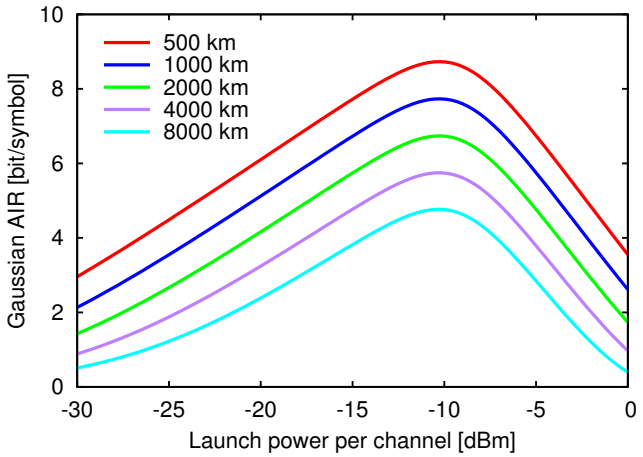


Fig. 25. Gaussian AIR for a single-polarization fully loaded WDM system (81 channels) with 50 GHz channel spacing on a dispersion-unmanaged IDA link.

parameters becomes the same and equals the one obtained in [26] with the RP method.

As an example, we compute the Gaussian AIR (115) for a single-polarization fully-loaded Nyquist-WDM system (81 channels) with 50 GHz channel spacing, employing a standard transmission fiber with attenuation, dispersion, and nonlinearity coefficients $\alpha = [0.2]dB/km$, $D = [17]ps/nm/km$, and $\gamma = [1.27 \cdot 10^{-3}]W^{-1}m^{-1}$, respectively. We consider both ideal distributed amplification (IDA), with spontaneous emission coefficient $\eta_s = 1$, and the more practical case of lumped amplification (LA) with $\eta_s = 1.6$ (a noise figure of about 5 dB) and different dispersion maps and span lengths. The system includes single-channel ideal DBP to remove intra-channel nonlinearity. The difference between the single- and dual-polarization cases is usually small and will be investigated in the next subsection. Fig. 25 shows the Gaussian AIR as a function of the signal launch power (per channel) $P_s$, for dispersion-unmanaged IDA links with different length. All the curves reach a maximum at the same optimum power ($[\sim -10]dBm$ per channel). Doubling the distance, the AIR decreases of about $[1]bit/symbol$.

The more practical case of an LA link is considered in Fig. 26, which shows the maximum Gaussian AIR (at optimum launch power) for different link lengths. In particular, Fig. 26a considers the case of a dispersion-unmanaged link and shows the dependence of the maximum AIR on the amplifier spacing (span length). Longer spacing means higher attenuation; hence, more ASE noise, higher optimum power, higher nonlinear interference, and, eventually, lower maximum AIR. The limit of zero amplifier spacing corresponds to the case of an IDA link (though with a higher noise figure and, hence, a lower maximum AIR than Fig. 25). On the other hand, Fig. 26b considers the case of a dispersion-managed link with an amplifier spacing of 100 km and shows the dependence of the maximum AIR on the residual dispersion per span. In-line dispersion compensation turns out to be detrimental in terms of AIR, as it reduces the walk-off between channels causing a coherent accumulation of the nonlinear interference generated in different fiber spans. However, since nonlinear interaction mostly takes place at the beginning of each fiber span, a small residual dispersion per span (about 20-30%) is sufficient to avoid such a coherent accumulation and practically approach the dispersion-uncompensated case.

### *D. Improved lower bounds*

The Gaussian AIRs shown in Fig. 25 and 26 are capacity lower bounds achievable by a detection strategy that does not account for residual channel memory (after DBP) and for the peculiar characteristics of nonlinear interference. The possible exploitation of the long time coherence of nonlinear interchannel interference (see Section VII) to compute tighter capacity lower bounds is suggested in [51] and then demonstrated in [189], with further improvements obtained in [67], [190].

Fig. 27 shows the AIR for a dispersion-unmanaged WDM system, considering the same configuration of Fig. 25, an IDA link of 1000 km, and different detection strategies. Both the single- and dual-polarization cases are shown with solid and dashed lines, respectively. Since these curves are obtained by full SSFM simulations, only five WDM channels are considered rather than a fully loaded system. Before estimating the Gaussian AIR (113), different nonlinear mitigation strategies are considered: single-channel DBP as in Fig. 25 (the AIR is slightly higher because of considering only five channels), single-channel DBP followed by adaptive least-square equalization (LSE), and multi-channel DBP performed on the whole received WDM spectrum. It turns out that the maximum AIR can be improved by at least 1 bit/symbol by including LSE to mitigate inter-channel nonlinearity (the gain increases with the number of WDM channels [67]).

As shown in Section VII, the XPM term generated by nonlinear interference is highly coherent (both in time and frequency) for IDA links, but much less coherent for links with lumped amplification. Thus, lower AIR gains are expected over the latter. More in general, the coherence decreases as the portion of the link along which nonlinear interaction takes place decreases, such that the AIR gain achievable by LSE (or similar techniques) compared to the Gaussian AIR over links with lumped amplification decreases with amplifier spacing. Fig. 28 shows the AIR gain (difference between the maximum AIRs obtained with and without LSE) as a function of the amplifier spacing for the same single-polarization system of Fig. 27 over a 1000 km link with lumped amplification (but unitary spontaneous emission factor). The limit of zero amplifier spacing corresponds to the case of an IDA link.

### *E. Beyond the nonlinear 'Shannon' Limit*

So far, only some capacity *lower bounds* have been presented, all sharing the same typical dependence on optical power—they all reach a maximum at some optimum power and then decrease again. We might be tempted to conclude that even the true channel capacity follows the same trend and has a finite maximum. However, as we already argued, there is no proof of the existence of such a *nonlinear Shannon limit*. In fact, the tightest available *upper bound* is the one proposed in [185], stating that the capacity of the optical fiber

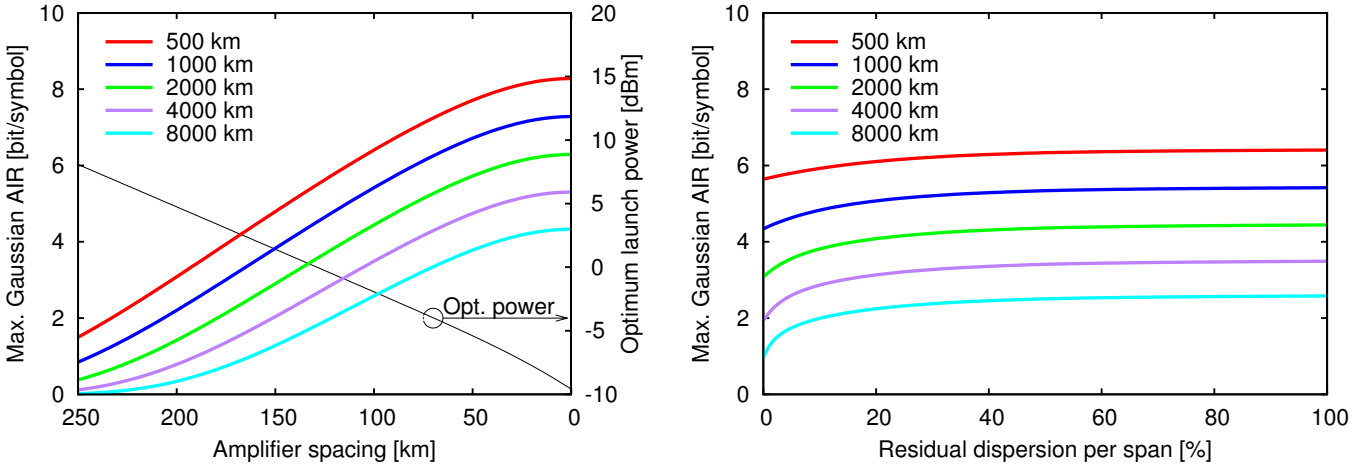


Fig. 26. Maximum Gaussian AIR for a single-polarization fully loaded (81 channels) WDM system with 50 GHz channel spacing on an LA link as a function of: (a) Amplifier spacing (without in-line compensation); (b) in-line residual dispersion (and a fixed amplifier spacing of 100 km).

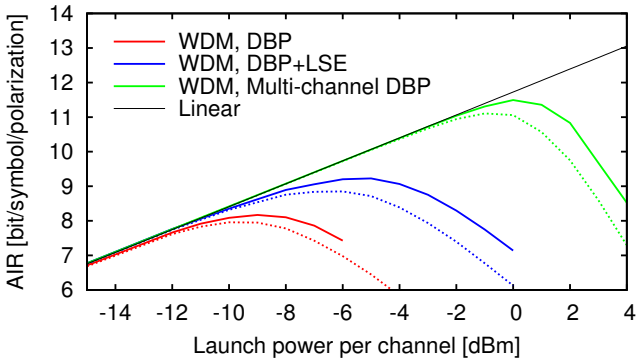


Fig. 27. AIR for a single- or dual-polarization 5-channel WDM system employing different detection strategies.

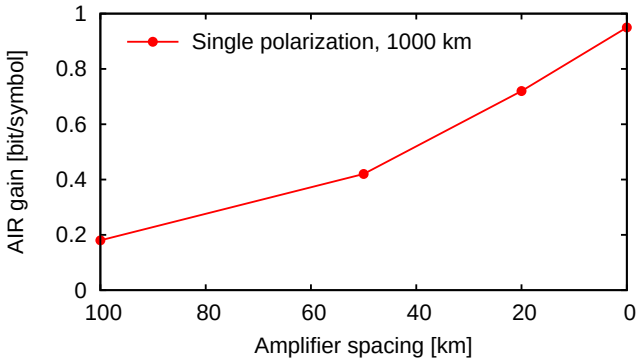


Fig. 28. AIR gain as a function of the amplifier spacing for a 5-channel single-polarization WDM system over a 1000 km link with lumped amplification.

channel is upper bounded by $\log_2(1+\mathrm{SNR})$—the capacity of an AWGN channel with same total accumulated noise. This has important theoretical and practical implications. In fact, it leaves the capacity problem essentially open, as the distance between the available lower and upper bounds diverges at high power, and stimulates the research to build communication systems that operate in the highly nonlinear with increased spectral efficiency. Nevertheless, despite a significant research effort, only small improvements with respect to the Gaussian AIR (the conventional NSL) have been demonstrated so far, which leaves no much room for optimism. Yet, some theoretical results suggest that more significant improvements might be possible, and that channel capacity might even increase unbounded with signal power.

One such result is the non-decreasing capacity lower bound [191], which states that the capacity of a discrete-time channel cannot decrease with power. This entails that the typical behavior of the capacity lower bounds discussed before—their decrease after some optimum power—is not a characteristic of the true channel capacity and should be more properly ascribed to the use of non-optimized input distributions. For instance, the capacity bounds obtained with a Gaussian input distribution (e.g., the Gaussian AIR) are very tight at low power, but vanish at high power where, instead, satellite distributions [191] or ripple distributions [254] can be used to obtain non-decreasing lower bounds. This is a clear indication that, in a strongly nonlinear regime, conventional modulation formats are highly suboptimal, and that significant gains might be expected from the optimization of the input distribution.

Other important results in this sense are the demonstrations of the infinite asymptotic capacity (for $P \to \infty$) of some simplified channels related to the optical fiber channel. This is the case, for instance, of the zero-dispersion fiber channel [36], of the memoryless FWM model [186], and of the RP model [192]. In all those cases, the Gaussian AIR has a finite maximum, but the (per symbol) capacity grows unbounded with power. This is still very far from a practical scheme that can guarantee such an unlimited capacity over a realistic fiber channel. Moreover, none of the previous results account for spectral broadening induced by fiber nonlinearity, such that a finite spectral efficiency limit can be still expected [45].

The asymptotic capacity of the optical fiber channel in the presence of signal-noise interaction is eventually investigated in [193], which suggests an alternative approach to address the capacity problem in this case.

## XIII. Conclusion

The field of the investigation of the generation, characterization, and assessment of the impact of nonlinear fiber effects has been extremely active over the last 10-15 years. Undoubtedly, remarkably useful results and practical tools have been obtained and put at the community's disposal. Their use has become quite pervasive, both in the transmission and in the optical networking sectors.

Investigation is still ongoing. What is probably at the forefront of current research is the devising of mitigation techniques to suppress the nonlinear disturbance affecting signals. To this end, the current (and future) results obtained in modeling provide an indispensable tool. The next few years will certainly see further progress in mitigation, also because the DSP power that can be used in transceivers is still ramping up exponentially, so that sophisticated techniques that seemed to be unrealistically complex not long ago, are gradually becoming viable. This might lead to further substantial increase in the performance of optical transmission systems and networks.

## Appendix A
## Conversion among SNR, Q-factor, MI and GMI for the AWGN channel

The BER is related to the Q-factor by the following definition:

$$\text{Q-factor} \triangleq 20\log_{10}\left(\sqrt{2}\text{erfc}^{-1}(2\text{BER})\right) \qquad \text{[dB]}$$

with $\text{erfc}^{-1}$ the inverse of the complementary error function. The main advantage of the Q-factor is that it is linearly, or almost linearly, related to SNR with slope 1 in a dB/dB scale for several modulation formats, thus helping rule of thumb design.

The relation between BER and SNR is modulation format dependent. For the AWGN channel with square $M-$QAM and by neglecting more than one bit error per symbol error we have:

$$\text{BER} \simeq \frac{2}{\log_2 M}\left(1-\frac{1}{\sqrt{M}}\right)\text{erfc}\left(\sqrt{\frac{3}{2(M-1)}\text{SNR}}\right), \qquad \text{square QAM} \tag{116}$$

The formula can be used even for BPSK with $M=4$ but by multiplying the SNR by 2, since QPSK, aka 4-QAM, has the same minimum symbol distance as BPSK at twice SNR.

In the relevant case of a signal modulated with symbol rate $R$ Gbd at 1550 nm and traveling in a $N$-span periodic link with EDFAs of gain $G_{\text{dB}}$ and noise figure $F_{\text{dB}}$, the SNR in dB is related to signal power $P_{\text{dBm}}$ by:

$$\text{SNR}_{\text{dB}} = P_{\text{dBm}} + 58 - 10\log_{10} N - G_{\text{dB}} - F_{\text{dB}} - 10\log_{10}\frac{R_{\text{Gbd}}}{12.5}$$

If we neglect the last term accounting for $R$ we get the optical SNR (OSNR) that could be measured on an optical spectrum analyzer (OSA) over a bandwidth of 12.5 GHz (0.1 nm).

Fig. 29 shows the relation between SNR, BER, MI and GMI for the AWGN channel for the most popular modulation formats. We reported the difference between GMI and MI to better highlight the numerical values. The figure is for single polarization: for PDM without polarization crosstalk just multiply the MI and GMI values by 2.

The SNR mismatch between MI and GMI curves is a fraction of dB at practical code overheads. For instance, error free transmission of single polarization 64-QAM in AWGN is possible by using FEC with overhead of 50% at SNR=9 dB without BICM (i.e, at MI=3 bits/symbol), or at SNR=9.44 dB with BICM (i.e., at GMI=3 bits/symbol). For smaller overhead the difference is smaller.

## Appendix B
## Notation and Conventions

- $z$: is the longitudinal spatial coordinate, along the link [km].
- $\alpha$: is the fiber *field* loss coefficient (km$^{-1}$), such that the signal *power* is attenuated as $\exp(-2\alpha z)$.
- $\alpha_{\text{p}}$: is the fiber power loss coefficient (km$^{-1}$), with $\alpha_{\text{p}} \triangleq 2\alpha$.
- $\beta_2$: is the dispersion coefficient (ps$^2$·km$^{-1}$). The relationship between $\beta_2$ and the widely used dispersion parameter $D$ in ps/(nm·km) is: $D = -\left(2\pi c/\lambda^2\right)\beta_2$, with $c$ the speed of light in km/s and $\lambda$ the light wavelength in nm.
- $\gamma$: is the fiber nonlinearity coefficient (W$^{-1}$·km$^{-1}$).
- $L_s$: is the span length (km).
- $L_{\text{eff}}$: is the span effective length defined as: $[1-\exp(-2\alpha L_s)]/2\alpha$ (km).
- $N_s$: is the total number of spans in a link, sometimes written $N_{\text{span}}$ when necessary for clarity.
- $G_{\text{s}}(f)$: is the PSD of the overall WDM transmitted signal (W/Hz).
- $G_{\text{NLI}}(f)$: is the PSD of NLI noise (W/Hz).
- $N_{\text{ch}}$: is the total number of channels present in the WDM comb.
- $P_n$: is the launch power of the $n$-th channel in the WDM comb (W). The power of a single channel is also sometimes written $P_{\text{ch}}$ when necessary for clarity.
- $R_n$: is the symbol rate of the $n$-th channel (TBaud). The symbol rate of a single channel is also written $R$, or $R_{\text{ch}}$ when necessary for clarity.
- $T_n = R_n^{-1}$: is the symbol time of the $n$-th channel (ps).
- $\Delta f$: is the channel spacing, used for systems where it is uniform (THz).
- $s_n(t)$: is the pulse used by the $n$-th channel, in time domain. Its Fourier transform is $s_n(f)$. The pulse is assumed to be normalized (Note that according to such normalization a channel with an ideal rectangular spectrum and bandwidth $R_n$ would have the flat-top value of its Fourier transform $s_n(f)$ equal to $R_n^{-1}$.) so that the integral of its absolute value squared is $T_n$. If any predistortion or dispersion pre-compensation is applied at the transmitter, this should be taken into account in $s_n(t)$ and $s_n(f)$.
- $B_n$: is the full bandwidth of the $n$-th channel (THz). If the channel is Nyquist then $B_n = R_n$.
- $f_n$: is the center frequency of the $n$-th channel (THz).
- $a_{x,n}^k$, $a_{y,n}^k$: are random variables corresponding to the symbols sent on the $n$-th channel at the $k$-th signal-

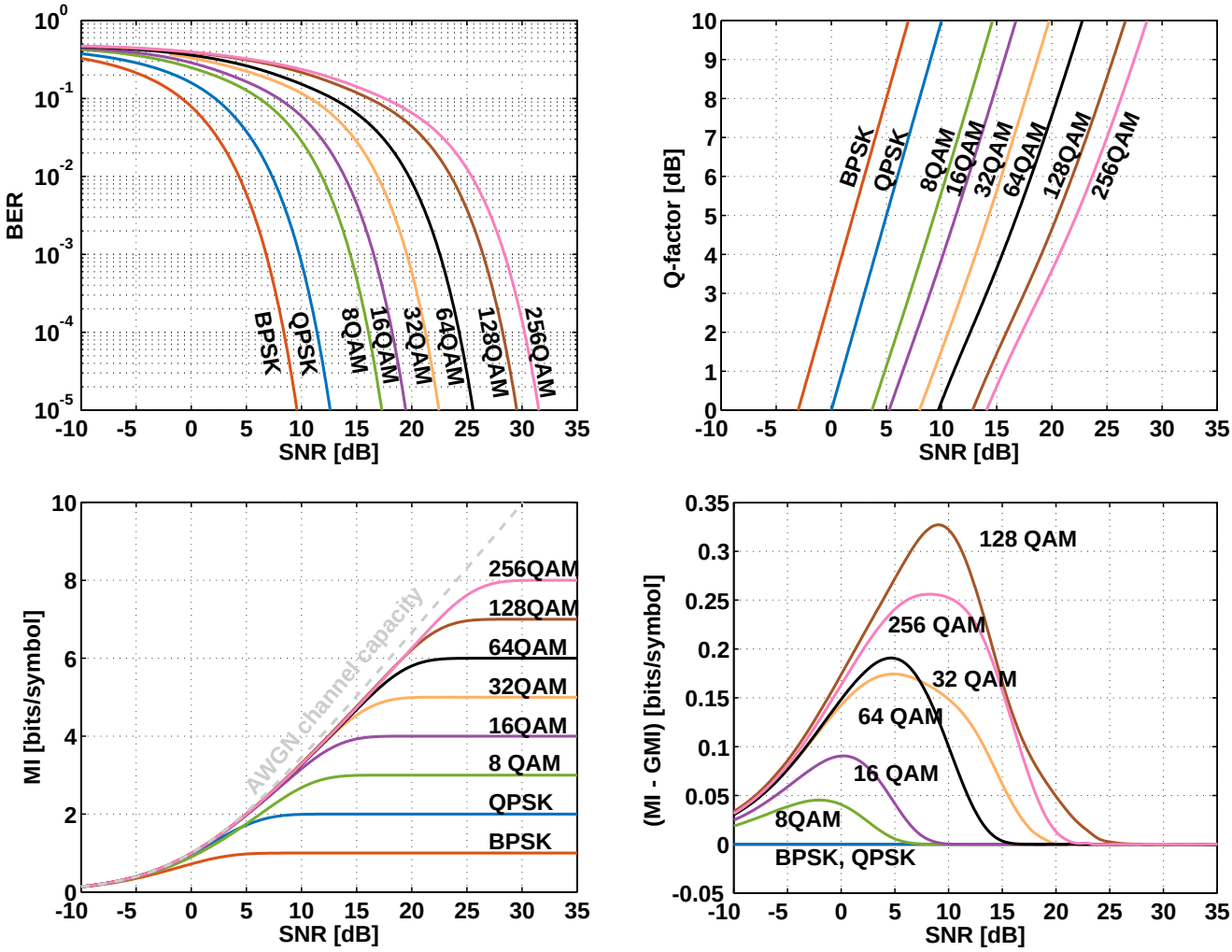


Fig. 29. Bit error rate (BER), Q-factor, mutual information (MI), and generalized mutual information (GMI) for the additive white Gaussian noise channel. 8 QAM is a rectangular QAM, 32 and 128 QAM are cross-QAM.

ing time, on either the polarization $\hat{x}$ or $\hat{y}$; note that $|a^k_{x,n}|^2$, $|a^k_{y,n}|^2$ must have dimensions of power (W). See also Eq. (63).

- $\mathbb{E}[.]$: is the expected value of a random variable.
- $F_{l,k,m}(\Omega_1, \Omega_2, \Omega_3)$ are kernels describing the NLI contribution resulting from FWN interaction between frequency tones $\Omega_1$, $\Omega_2$, $\Omega_3$ . The SPM and XPM kernels are given by $S_{l,k,m} = F_{l,k,m}(0,0,0)$ and $X_{l,k,m}(\Omega) = F_{l,k,m}(0,\Omega,\Omega)$ , respectively.
- $g_\Omega(z,t)$ represents the dispersed waveform of the pulse transmitted over a WDM channel spaced by $\Omega$ from the channel of interest, when reaching point $z$ along the fiber. $\tilde{g}_j()$ represents the baseband spectral shape of the transmitted pulses from the $j$-th interfering WDM channel.
- $\underline{a}_n$ and $\underline{r}_n$ represent the transmitted and received polarization multiplexed data symbols of the channel of interest in the $n$-th time slot. $\underline{b}_{n,j}$ represents the two-element data-vector transmitted in the $n$-th time-slot over the $j$-th interfering WDM channel.
- $\mathbf{H}_l^{(n)}$ are 2x2 matrices representing the $l$-th nonlinear ISI term in the $n$-th time slot.
- $h_{zs}(t)$: is the impulse response of a filter collecting all linear effects accumulated from coordinate $z$ to coordinate $s$.
- $R_{zs}(\tau)$: is the cross-correlation function between the electric field at coordinate $z$ and coordinate $s$, i.e., $R_{zs}(\tau) = \mathbb{E}[A(z,t+\tau)A^*(s,t)]$
- $\kappa_{n;n}$: is $n$-th-order cumulant of data symbols $a$. They are also written as $\kappa_{2;2} = -\mu_2^2\Phi$ and $\kappa_{3;3} = -\mu_2^3\Psi$, with $\mu_2 = \mathbb{E}[|a|^2]$.
- $F$: is the noise figure (dB) of an EDFA.
- $P_{\text{NLT}}$: is the nonlinear threshold, i.e., the optimal launched power maximizing the SNR at the receiver.
- $\eta$: is the normalized nonlinear interference variance (1/mW$^2$), i.e., $\sigma^2_{\text{NLI}} = \eta P^3$. For a single-span it is indicated by $\eta_1$.
- $\epsilon$: is the coherence accumulation factor of nonlinear interference along distance. It is $0 \leq \epsilon \leq 1$ where $\epsilon = 0$ and $\epsilon = 1$ indicate incoherent and fully-coherent accumulation, respectively.
- $N_0$: is the system reach (spans) of an homogeneous

optical link, i.e., the maximum distance for which the SNR is equal to a given threshold. The reach in km is $N_0 L_s$.

- $\mathcal{L}$: is a linear operator accounting for dispersive effects, i.e., $\mathcal{L} = j(\beta_2/2)(\partial^2/\partial t^2)$.
- $\mathcal{N}$: is a nonlinear operator accounting for the nonlinear Kerr effect and fiber attenuation, i.e., $\mathcal{N} = -\alpha_{\rm p}/2 - j\gamma |A|^2$. Fiber attenuation, although a linear effect, is included in the nonlinear operator to make the SSFM more efficient.
- $h(X)$; is the differential entropy of the random variable $X$. It gives a measure of the average rate at which information is produced by $X$. The differential entropy is also called continuous entropy to emphasize that focus on a continuous random variable $X$, contrary to the entropy that refers to a discrete random variable.

## APPENDIX C
## LIST OF ACRONYMS

| | |
|---|---|
| ACFS | approximate closed-form solution |
| ADC | analog to digital converter |
| AIR | achievable information rate |
| AWGN | additive white Gaussian noise |
| ASE | amplified spontaneous-emission |
| ASI | asymmetric information |
| BER | bit error-rate |
| BICM | Bit-interleaved coded modulation |
| CD | chromatic dispersion |
| COI | channel of interest |
| CPE | carrier phase estimator |
| CSCG | circularly symmetric complex Gaussian |
| CUT | channel under test |
| DAC | digital to analog converter |
| DGD | differential group delay |
| DBP | digital backward propagation |
| DCU | dispersion-compensating unit |
| DM | dispersion-managed |
| DP | dual-polarization |
| DSP | digital signal processing |
| EDFA | erbium-doped fiber amplifier |
| EGN-model | enhanced Gaussian-noise model |
| EON | elastic optical network |
| eRP | enhanced regular perturbation |
| FEC | forward error-correcting code |
| FFT | fast Fourier transform |
| FRLP | frequency-resolved logarithmic perturbation |
| FWM | four-wave mixing |
| GN-model | Gaussian-noise model |
| GNRF | GN-model reference formula |
| GMI | Generalized mutual information |
| GVD | group velocity dispersion |
| IAA | incoherent accumulation assumption |
| IC | interfering channel |
| IDA | ideal distributed amplification |
| IFFT | inverse fast Fourier transform |
| iGN-model | incoherent Gaussian-noise model |
| IGNRF | GN-model reference formula |
| INFT | inverse nonlinear Fourier transform |
| ISI | inter-symbol interference |
| LA | lumped amplification |
| LDPC | low-density parity check code |
| LOGO | local-optimization, global optimization |
| LOGON | LOGO with Nyquist-WDM |
| LP | logarithmic perturbation |
| LSE | least-square equalization |
| LWN | locally-white noise |
| MAP | maximum a-posteriori probability |
| ME | Manakov equation |
| MCI | multi-channel interference |
| MI | mutual information |
| MR | maximum (system) reach |
| NFDM | nonlinear frequency-division multiplexing |
| NFT | nonlinear Fourier transform |
| NLC | nonlinearity compensation |
| NLI | nonlinear interference |
| NLPN | nonlinear phase noise |
| NLSE | nonlinear Schroedinger equation |
| NLT | nonlinear threshold |
| NSL | nonlinear Shannon limit |
| NSNI | nonlinear signal-noise interaction |
| NZDSF | non-zero dispersion-shifted fiber |
| OFDM | orthogonal frequency-division multiplexing |
| OPA | optical parametric amplifier |
| OSNR | optical signal-to-noise ratio |
| PDF | probability density function |
| PDL | polarization dependent loss |
| PDM | polarization division multiplexing |
| PM | polarization-multiplexed |
| PMD | polarization mode dispersion |
| PPRN | phase and polarization rotation noise |
| PSCF | pure-silica-core fiber |
| PSD | power spectral density |
| QAM | quadrature amplitude modulation |
| QOT | quality of transmission |
| QPSK | quadrature phase-shift keying |
| RMSA | routing modulation and spectrum assignment |
| RP | regular perturbation |
| RPR | random process |
| RV | random variable |
| Rx | receiver |
| SA | statistical approach |
| SCI | self-channel interference |
| SDBP | stochastic digital backward propagation |
| SE | spectral efficiency |
| SI | statistical independence |
| SMF | (standard) single-mode fiber |
| SNR | signal-to-noise ratio |
| SPM | self phase modulation |
| SRO | symbol rate optimization |
| SSFM | split-step Fourier method |
| TD | time-domain |
| Tx | transmitter |
| ULH | ultra-long haul |
| UPA | undepleted pump assumption |
| UT | uncompensated transmission |

| | |
|---|---|
| VS | Volterra series |
| WDM | wavelength-division multiplexing |
| WRON | wavelength-routed optical network |
| XCI | cross-channel interference |
| XPM | cross phase modulation |
| XPolM | cross polarization modulation |

## REFERENCES


[1] G. P. Agrawal, *Non-Linear Fiber Optics*, $5^{\text{th}}$ edition, Academic Press, 2012. ISBN-13: 978-0123970237, ISBN-10: 0123970237.

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

[3] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, 'Time varying ISI model for nonlinear interference noise,' Optical Fiber Communication Conference (OFC), paper W2A.62 (2014).

[4] R. Dar and P. J. Winzer 'Nonlinear interference mitigation: methods and potential Gain,' *J. Lightwave Technology*, vol. 35, pp. 903–930 (2017).

[5] R. C. Jones, 'A new calculus for the treatment of optical systems. VII. Properties of the N-matrices,' *J. Optical Society of America*, vol. 38, pp. 671–685 (1948).

[6] J. P. Gordon and H. Kogelnik, 'PMD fundamentals: Polarization mode dispersion in optical fibers,' *National Academy of Sciences*, vol. 97, pp. 4541–4550 (2000).

[7] O. Golani, R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, 'Modeling the bit-error-rate performance of nonlinear fiber-optic systems,' *J. Lightwave Technology*, vol. 34, pp. 3482–3489 (2016).

[8] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "Inter-channel nonlinear interference noise in WDM systems: Modeling and mitigation," J. Lightwave Technology, vol. 33, pp. 1044–1053 (2015).

[9] R. Dar, O. Geller, M. Feder, A. Mecozzi, M. Shtaif, 'Mitigation of inter-channel nonlinear interference in WDM systems,' in *Proc. of European Conference on Optical Communication (ECOC)*, paper P.5.6, Cannes (FR), Sept. 2014.

[10] 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. Lightwave Technology, vol. 33, pp. 4975–4984 (2015).

[11] C. Pan, H. Bülow, W. Idler, L. Schmalen, and F. R. Kschischang, 'Optical nonlinear phase noise compensation for 9 x 32-Gbaud PolDM-16 QAM transmission using a code-aided Expectation-Maximization algorithm,' J. Lightwave Technology, vol. 33, pp. 3679–3686 (2015).

[12] C. Schmidt-Langhorst, R. Elschner, F. Frey, R. Emmerich, and C. Schubert, 'Experimental analysis of nonlinear interference noise in heterogeneous flex-grid WDM transmission,' European Conference on Optical Communication (ECOC), paper Tu.1.4.3 (2015).

[13] L. Li, Z. Tao, L. Liu, W. Yan, S. Oda, T. Hoshida, and J. C Rasmussen, 'Nonlinear polarization-crosstalk canceller for dual-polarization digital coherent receivers,' Optical Fiber Communication Conference (OFC), paper OWE3 (2010).

[14] A. Ghazisaeidi, M. Salsi, J. Renaudier, O. Bertran-Pardo, H. Mardoyan, and G. Charlet, 'Performance analysis of decision-aided nonlinear cross-polarization mitigation algorithm,' European Conference on Optical Communication (ECOC), paper We-3 (2012).

[15] C. R. Menyuk, 'Nonlinear pulse propagation in birefringent optical fibers,' *IEEE J. Quantum Electron.*, vol. 23, pp. 174-176, Feb. 1987.

[16] C. R. Menyuk, 'Pulse propagation in an elliptically birefringent Kerr medium,' *IEEE J. Quantum Electron.*, vol. 25, no. 12, pp. 2674-2682, Dec. 1989.

[17] G. P. Agrawal, *Fiber-Optic Communications Systems*, 3rd ed. New York: Wiley, 2002.

[18] P. K. A. Wai, C. R. Menyuk, and H. H. Chen, 'Stability of solitons in randomly varying birefringent fibers,' *Opt. Lett.*, vol. 16, no. 16, pp. 1231-1233, 1991.

[19] S. G. Evangelides Jr., L. F. Mollenauer, J. P. Gordon, and N. S. Bergano, 'Polarization multiplexing with solitons,' *J. Lightwave Technol.*, vol. 10, no. 1, pp. 28-35, Jan. 1992.

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

[21] N. Rossi, P. Serena and A. Bononi, 'Polarization-dependent loss impact on coherent optical systems in presence of fiber nonlinearity,' *IEEE Photon. Technol. Lett.*, vol. 26, no. 4, pp. 334-337, Feb. $15^{\text{th}}$, 2014.

[22] A. Vannucci, P. Serena, and A. Bononi, 'The RP method: a new tool for the iterative solution of the nonlinear Schrodinger equation,' *J. Lightwave Technol.*, vol. 20, no. 7, pp. 1102-1112, July 2002.

[23] K.V. Peddanarappagari and M. Brandt-Pearce 'Volterra series transfer function of single-mode fibers,' *J. of Lightwave. Technol.*, vol. 15, no. 12, pp. 2232-2241, Dec. 1997.

[24] A. Mecozzi, C. Balslev Clausen, and M. Shtaif, 'Analysis of intrachannel nonlinear effects in highly dispersed optical pulse transmission,' *IEEE Photon. Technol. Lett.*, vol. 12, no. 4, pp. 392-394, Apr. 2000.

[25] E. E. Narimanov and P. P. Mitra, 'The channel capacity of a fiber optics communication system: perturbation theory,' *J. Lightwave Technol.*, vol. 20, no. 3, pp. 530-537, Mar. 2002.

[26] A. Mecozzi and R.-J. Essiambre, 'Nonlinear Shannon limit in pseudolinear coherent systems,' *J. Lightwave Technol.*, vol. 30, no. 12, pp. 2011-2024, June $15^{\text{th}}$ 2012.

[27] 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, no. 16, pp. 3358-3369, Aug. 2009.

[28] P. Poggiolini, A. Carena, Y. Jiang, G. Bosco, V. Curri, and F. Forghieri, 'Impact of low-OSNR operation on the performance of advanced coherent optical transmission systems,' in *Proc. of ECOC 2014*, Cannes (FR), Sept. 2014. Available with corrections on www.arXiv.org, paper arXiv:1407.2223.

[29] J. P. Gordon and L. F. Mollenauer, 'Phase noise in photonic communications systems using linear amplifiers,' *Opt. Lett.*, vol. 15, no. 23, pp. 1351-1353, 1990.

[30] A. Splett, C. Kurzke, and K. Petermann, 'Ultimate transmission capacity of amplified optical fiber communication systems taking into account fiber nonlinearities,' *in Proc. ECOC 1993*, vol. 2, pp. 41-44, Montreux (CH), Sept. 1993.

[31] R. Hui, K. R. Demarest, and C. T. Allen, 'Cross-phase modulation in multispan WDM optical fiber systems,' *J. Lightwave. Technol.*, vol. 17, no. 6, pp. 1018-1026, Jun. 1999.

[32] P. P. Mitra and J. B. Stark, 'Nonlinear limits to the information capacity of optical fiber communications,' *Nature*, vol. 411, no. 6841, pp. 1027-1030, Jun. 2001.

[33] 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, no. 3, pp. 389–400, Mar. 2002.

[34] Jau Tang, 'The channel capacity of a multispan DWDM system employing dispersive nonlinear optical fibers and an ideal coherent optical receiver,' *J. Lightwave Technol.*, vol. 20, no. 7, pp. 1095-1101, July 2002.

[35] H. Louchet, A. Hodzic, and Klaus Petermann, 'Analytical model for the performance evaluation of DWDM transmission systems,' *IEEE Phot. Technol. Lett.*, vol. 15, no. 9, pp. 1219-1221, Sept. 2003.

[36] 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, pp. 203901-1 to 203901-4, Nov. 2003.

[37] K.-P. Ho, 'Error probability of DPSK signals with cross-phase modulation induced nonlinear phase noise,' *IEEE J. Sel. Topics Quantum Electron.*, vol. 10, no. 2, pp. 421-427, Mar./Apr. 2004.

[38] E. Ciaramella and E. Forestieri, 'Analytical approximation of nonlinear distortions,' *IEEE Photon. Technol. Lett.*, vol. 17, no. 1, pp. 91-93, Jan. 2005.

[39] K.-P. Ho and H.-C. Wang, 'Effect of dispersion on nonlinear phase noise,' *Opt. Lett.*, vol. 31, no. 14, pp. 2109-2111, July 2006.

[40] M. Nazarathy, J. Khurgin, R. Weidenfeld, Y. Meiman, Pak Cho, R. Noe, I. Shpantzer, and V. Karagodsky 'Phased-array cancellation of nonlinear FWM in coherent OFDM dispersive multi-span links,' *Optics Express*, vol. 16, pp. 15778-15810, Sept. 2008.

[41] A. D. Ellis, J. Zhao, and D. Cotter, 'Approaching the non-linear Shannon limit,' *J. Lightwave Technol.*, vol. 28, no. 4, pp. 423-433, Feb. 2010.

[42] X. Chen and W. Shieh, 'Closed-form expressions for nonlinear transmission performance of densely spaced coherent optical OFDM systems,' *Optics Express*, vol. 18, pp. 19039-19054, Aug. 2010.

[43] G. Bosco, A. Carena, R. Cigliutti, V. Curri, P. Poggiolini, and F. Forghieri, 'Performance prediction for WDM PM-QPSK transmission over uncompensated links,' in *Proc. OFC 2011*, paper OThO7, Los Angeles (CA), Mar. 2011.

[44] P. Poggiolini, A. Carena, V. Curri, G. Bosco, and F. Forghieri, 'Analytical modeling of non-Linear propagation in uncompensated optical transmission links,' *IEEE Photon. Technol. Lett.*, vol. 23, no. 11, pp. 742-744, June 2011.

[45] M. I. Yousefi and F. R. Kschischang, 'On the per-sample capacity of nondispersive optical fibers,' *IEEE Trans. Inf. Theory*, vol. 57, no. 11, pp. 7522–7541, Nov. 2011.

[46] G. Bosco, P. Poggiolini, A. Carena, V. Curri, and F. Forghieri, 'Analytical results on channel capacity in uncompensated optical links with coherent detection,' *Optics Express*, vol. 19, no. 26, pp. B438-B449, Dec. 2011.

[47] A. Bononi, P. Serena, N. Rossi, E. Grellier, and F. Vacondio, 'Modeling nonlinearity in coherent transmissions with dominant intrachannel-four-wave-mixing,' *Optics Express*, vol. 20, pp. 7777-7791, 26 March 2012.

[48] A. Carena, V. Curri, G. Bosco, P. Poggiolini, and F. Forghieri, 'Modeling of the impact of non-linear propagation effects in uncompensated optical coherent transmission links,' *J. of Lightwave. Technol.*, vol. 30, no. 10, pp. 1524-1539, May 2012.

[49] P. Poggiolini, G. Bosco, A. Carena, V. Curri, Y. Jiang, and F. Forghieri, 'A detailed analytical derivation of the GN model of non-linear interference in coherent optical transmission systems,' posted on arXiv, www.arxiv.org, paper identifier 1209.0394. First posted Sept. 2012.

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

[51] M. Secondini and E. Forestieri, 'Analytical fiber-optic channel model in the presence of cross-phase modulations,' *IEEE Photon. Technol. Lett.*, vol. 24, no. 22, pp. 2016-2019, Nov. $15^{\text{th}}$ 2012.

[52] P. Johannisson and M. Karlsson, 'Perturbation analysis of nonlinear propagation in a strongly dispersive optical communication system,' *J. Lightwave Technol.*, vol. 31, no. 8, pp. 1273-1282, Apr. $15^{\text{th}}$ 2013.

[53] S. J. Savory, 'Approximations for the nonlinear self-channel interference of channels with rectangular spectra,' *IEEE Phot. Technol. Lett.*, vol. 25, no. 10, pp. 961-964, May $15^{\text{th}}$ 2013.

[54] A. Carena, G. Bosco, V. Curri, P. Poggiolini, and F. Forghieri, 'Impact of the transmitted signal initial dispersion transient on the accuracy of the GN-model of non-linear propagation,' in *Proc. of ECOC 2013*, paper Th.1.D.4, London (UK), Sept. 2013.

[55] P. Serena and A. Bononi, 'On the accuracy of the Gaussian nonlinear model for dispersion-unmanaged coherent links,' in *Proc. of ECOC 2013*, paper Th.1.D.3, London (UK), Sept. 2013.

[56] P. Johannisson and E. Agrell, 'Modeling of nonlinear signal distortion in fiber-optical networks,' posted on arXiv.org, paper arXiv:1309.4000, Sept. 2013.

[57] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, 'Properties of nonlinear noise in long, dispersion-uncompensated fiber links,' *Optics Express*, vol. 21, no. 22, pp. 25685-25699, Nov. 2013.

[58] P. Serena and A. Bononi, 'An alternative approach to the Gaussian noise model and its system implications,' *J. Lightwave Technol.*, vol. 31, no. 22, pp. 3489-3499, Nov. $15^{\text{th}}$ 2013.

[59] M. Secondini, E. Forestieri and G. Prati, 'Achievable information rate in nonlinear WDM fiber-optic systems with arbitrary modulation formats and dispersion maps,' *J. Lightwawe Technol.*, vol. 31, no. 23, pp. 3839-3852, Dec. $1^{\text{st}}$, 2013.

[60] A. Bononi, O. Beucher, P. Serena 'Single- and cross-channel nonlinear interference in the gaussian noise model with rectangular spectra,' *Optics Express*, vol. 21, no. 26, pp. 32254-32268, Dec. 2013.

[61] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, 'Accumulation of nonlinear interference noise in fiber-optic systems,' *Optics Express*, vol. 22, no. 12, pp. 14199-14211, June 2014.

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

[63] A. Carena, G. Bosco, V. Curri, Y. Jiang, P. Poggiolini and F. Forghieri, 'EGN model of non-linear fiber propagation,' *Optics Express*, vol. 22, no. 13, pp. 16335-16362, June 2014. Extended appendices with full formulas derivations can be found in the version available on *www.arXiv.org* .

[64] A. Carena, G. Bosco, V. Curri, Y. Jiang, P. Poggiolini, F. Forghieri 'On the Accuracy of the GN-Model and on Analytical Correction Terms to Improve It,' paper arXiv:1401.6946, www.arXiv.org, June 6th, 2014.

[65] P. Poggiolini, G. Bosco, A. Carena, V. Curri, Y. Jiang, and F. Forghieri, 'A simple and effective closed-form GN model correction formula accounting for signal non-Gaussian distribution,' *J. of Lightw. Technol.*, vol. 33, no. 2, pp. 459-473, Jan. 2015.

[66] P. Poggiolini, Y. Jiang, A. Carena, and F. Forghieri, 'A Simple and Accurate Closed-Form EGN Model Formula,' paper arXiv:1503.04132, Mar. 2015.

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

[68] 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, no. 4, pp. 662-701, Feb. 2011.

[69] E. Torrengo, R. Cigliutti, G. Bosco, A. Carena, V. Curri, P. Poggiolini, A. Nespola, D. Zeolla, and F. Forghieri 'Experimental validation of an analytical model for nonlinear propagation in uncompensated optical links,' *Optics Express*, vol. 19, no. 26, pp. B790-B798, Dec. 2011.

[70] F. Vacondio, O. Rival, C. Simonneau, E. Grellier, A. Bononi, L. Lorcy, J.-C. Antona and S. Bigo, 'On nonlinear distortions of highly dispersive optical coherent systems,' *Optics Express*, Vol. 20, No. 2, pp. 1022-1032, Jan. 2012.

[71] J.-X. Cai, H. Zhang, H. G. Batshon, M. Mazurczyk, O. V. Sinkin, D. G. Foursa, A. N. Pilipetskii, G. Mohs, and Neal S. Bergano, '200 Gb/s and dual wavelength 400 Gb/s transmission over transpacific distance at 6.0 b/s/Hz spectral efficiency,' *J. Lightwave Technol.* vol. 32, no. 4, pp. 832-839, Feb. 2014.

[72] J.-X. Cai, O. V. Sinkin, H. Zhang, H. G. Batshon, M. Mazurczyk, D. G. Foursa, A. Pilipetskii, and G. Mohs, 'Nonlinearity compensation benefit in high capacity ultra-long haul transmission systems,' in *Proc. of ECOC 2013*, paper We.4.D.2, London (UK), Sept. 2013.

[73] A. J. Stark, Y.-T. Hsueh, T. F. Detwiler, M. M. Filer, S. Tibuleac and S.E. Ralph, 'System performance prediction with the Gaussian noise model in 100G PDM-QPSK coherent optical networks,' *J. Lightwave. Technol.*, vol. 31, no. 21, pp. 3352-3360, Nov. 2013.

[74] A. Nespola, S. Straullu, A. Carena, G. Bosco, R. Cigliutti, V. Curri, P. Poggiolini, M. Hirano, Y. Yamamoto, T. Sasaki, J. Bauwelinck, K. Verheyen, and F. Forghieri, 'GN-model validation over seven fiber types in uncompensated PM-16QAM Nyquist-WDM links,' *IEEE Photon. Technol. Lett.*, vol. 26, no. 2, pp. 206-209, Jan. 2014.

[75] J.-X. Cai, H. G. Batshon, H. Zhang, M. Mazurczyk, O. V. Sinkin, D. G. Foursa, and A. N. Pilipetskii, 'Transmission performance of coded modulation formats in a wide range of spectral efficiencies,' *in Proc. of OFC 2014*, paper M2C.3, San Francisco (CA), Mar. 2014.

[76] J. Pan, P. Isautier, M. Filer, S. Tibuleac and S. E. Ralph, 'Gaussian noise model aided in-band crosstalk analysis in ROADM-enabled DWDM networks,' *in Proc. of OFC 2014*, paper Th1I.1, San Francisco (CA), Mar. 2014.

[77] C. E. Shannon, 'A mathematical theory of communication,' in *Bell Syst. Tech. J.*, vol. 27, pp. 379-423 1948.

[78] S. Savory, 'Digital filters for coherent optical receivers,' *Opt. Express*, vol. 16, no. 2, pp. 804-817, 2008.

[79] E. Ip, J.M. Kahn, 'Compensation of Dispersion and Nonlinear Impairments Using Digital Backpropagation,' *J. Lightw. Technol.*, vol. 26, no. 20, Oct. 2008, pp. 3416-3425.

[80] S. J. Savory, G. Gavioli, E. Torrengo, P. Poggiolini, 'Impact of Interchannel Nonlinearities on a Split-Step Intrachannel Nonlinear Equalizer,' *IEEE Phot. Technol. Lett.*, vol. 22, no. 10, pp. 673-675, May 15th 2010.

[81] D. Rafique, J. Zhao, D. Ellis, 'Digital back-propagation for spectrally efficient WDM 112 Gbit/s PM m-ary QAM transmission,' *Optics Express*, vol. 19, no. 6, pp. 5219-5224, Mar. 2011.

[82] G. Gao, X. Chen, W. Shieh, 'Influence of PMD on nonlinearity compensation using digital back propagation,' *Optics Express*, vol.20, pp.14406-14418, 2012.

[83] M. Qiu, Q. Zhuge, X. Xu, M. Chagnon, M. Morsy-Osman, and David V. Plant 'Subcarrier multiplexing using DACs for fiber nonlinearity mitigation in coherent optical communication systems,' in *Proc. OFC 2014*, paper Tu3J.2, San Francisco (CA), Mar. 2014.

[84] W. Shieh and Y. Tang, 'Ultrahigh-speed signal transmission over nonlinear and dispersive fiber optic channel: the multicarrier advantage,' *IEEE Photonics J.*, vol. 2, no. 3, pp. 276-283, June 2010.

[85] L. B. Du and A. J. Lowery, 'Optimizing the subcarrier granularity of coherent optical communications systems,' *Optics Expr.*, vol. 19, no. 9, pp. 8079, Apr. 2011.

[86] A. Bononi, N. Rossi and P. Serena, 'Performance dependence on channel baud-rate of coherent single-carrier WDM systems,' in *Proc. of ECOC 2013*, paper Th.1.D.5, London (UK), Sept. 2013.

[87] A. Carbó, et al. 'Experimental Analysis of Non Linear Tolerance Dependency of Multicarrier Modulations versus Number of WDM Channels', in *Proc. OFC 2016*, paper Tu3A.6, Anaheim (CA) , Mar. 2016.

[88] P. Poggiolini, G. Bosco, A. Carena, R. Cigliutti, V. Curri, F. Forghieri, R. Pastorelli, and S. Piciaccia, 'The LOGON strategy for low-complexity control plane implementation in new-generation flexible networks,' in *Proc. of OFC 2013*, paper OW1H.3, Los Angeles (CA), Mar. 2013.

[89] R. Pastorelli, S. Piciaccia, G. Galimberti, E. Self, M. Brunella, G. Calabretta, F. Forghieri, D. Siracusa, A. Zanardi, E. Salvadori, G. Bosco, A. Carena, V. Curri, and P. Poggiolini 'Optical control plane based

on an analytical model of non-linear transmission effects in a self-optimized network,' in *Proc. of ECOC 2013*, paper We.3.E.4, London (UK), Sept. 2013.

[90] R. Pastorelli, G. Bosco, A. Nespola, S. Piciaccia, and F. Forghieri, 'Network planning strategies for next-generation flexible optical networks,' in *Proc. OFC 2014*, paper M2B.1, San Francisco (CA), Mar. 2014.

[91] P. Serena, A. Bononi, 'A Time-Domain Extended Gaussian Noise Model,' *J. Lightw. Technol.*, vol. 33, no. 7, pp. 1459-1472, Apr. 2015.

[92] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, 'Pulse collision picture of inter-channel nonlinear interference noise in fiber-optic communications,' J. Lightwave Technology, vol. 34, pp. 593–607 (2016).

[93] J. N. Damask, *Polarization Optics in Telecommunications*. Springer, 2005.

[94] I. P. Kaminow, T. Li, and A. E. Willner, *Optical Fiber Telecommunications Volume VIB: Systems and Networks*. Academic Press, 2013.

[95] S. Haykin, *Adaptive Filter Theory* , 3rd ed. Prentice Hall, Inc, 1996.

[96] C. D. Poole, 'Statistical treatment of polarization dispersion in single-mode fiber.,' *Opt. Lett.*, vol. 13, no. 8, p. 687, 1988.

[97] G. J. Foschini and C. D. Poole, 'Statistical Theory of Polarization Dispersion in Single Mode Fibers,' *J. Lightw. Technol.*, vol. 9, no. 11, pp. 1439–1456, 1991.

[98] M. Karlsson, 'Probability density functions of the differential group delay in optical fiber communication systems,' *J. Lightw. Technol.*, vol. 19, no. 3, pp. 324–331, 2001.

[99] P. K. A. Wai and C. R. Menyuk, 'Polarization mode dispersion, decorrelation, and diffusion in optical fibers with randomly varying birefringence,' *J. Lightw. Technol.*, vol. 14, no. 2, pp. 148–157, 1996.

[100] D. Breuer, H. Tessmann, A. Gladisch, H. M. Foisel, G. Neumann, H. Reiner, and H. Cremer, 'Measurements of PMD in the installed fiber plant of Deutsche Telekom,' *in Proc. LEOS 2003*, 2003.

[101] D. Nolan, X. Chen, and M.-J. Li, 'Fibers with low polarization-mode dispersion,' *J. Lightw. Technol.*, vol. 22, no. 4, pp. 1066–1077, Apr. 2004.

[102] O. Bertran-Pardo, J. Renaudier, G. Charlet, P. Tran, H. Mardoyan, M. Bertolini, M. Salsi, and S. Bigo, "Demonstration of the benefits brought by PMD in polarization-multiplexed systems," in *Proc. of ECOC 2010*, paper Th.10.E.4, Turin (Italy), Sept. 2010.

[103] P. Serena, N. Rossi, O. Bertran-Pardo, J. Renaudier, A. Vannucci, and A. Bononi, "Intra- Versus Inter-Channel PMD in Linearly Compensated Coherent PDM-PSK Nonlinear Transmissions," *J. Lightw. Technol.*, vol. 29, no. 11, pp. 1691–1700, 2011.

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

[105] C. Xie, 'WDM coherent PDM-QPSK systems with and without inline optical dispersion compensation,' *Opt. Express*, vol. 17, no. 6, pp. 4815–4823, 2009.

[106] M. Karlsson and H. Sunnerud, 'Effects of Nonlinearities on PMD-Induced System Impairments,' *J. Lightw. Technol.*, vol. 24, no. 11, pp. 4127–4137, Nov. 2006.

[107] M. Winter, C. -a. Bunge, D. Setti, and K. Petermann, 'A Statistical Treatment of Cross-Polarization Modulation in DWDM Systems,' *J. Lightw. Technol.*, vol. 27, no. 17, pp. 3739–3751, Sep. 2009.

[108] M. Boroditsky et al., 'Effect of nonlinearities on PMD,' *J. Lightw. Technol.*, vol. 24, pp. 4100–4107, Nov. 2006.

[109] B. C. Collings and L. Boivin, 'Nonlinear polarization evolution induced by cross-phase modulation and its impact on transmission systems,' *IEEE Photon. Technol. Lett.*, vol. 12, pp. 1582–1584, Nov. 2000.

[110] J. P. Gordon and H. Kogelnik, 'PMD fundamentals: Polarization mode dispersion in optical fibers,' in *Proc. National Academy of Sciences of the United States of America*, vol. 97, no. 9, pp. 4541–4550, Apr. 2000.

[111] M. Karlsson and M. Petersson, 'Quaternion Approach to PMD and PDL Phenomena in Optical Fiber Systems,' *J. Lightw. Technol.*, vol. 22, no. 4, pp. 1137–1146, 2004.

[112] A. Mecozzi and M. Shtaif, 'Signal-to-noise-ratio degradation caused by polarization-dependent loss and the effect of dynamic gain equalization,' *J. Lightw. Technol.*, vol. 22, no. 8, pp. 1856–1871, 2004.

[113] O. Vassilieva, S. Oda, H. T., J. C. Rasmussen, and M. Sekiya, 'Experimental Investigation of the Statistics of the Interplay between Nonlinear and PDL Effects in Polarization Multiplexed Systems,' in Proc. OFC 2013, 2013, vol. OM3B.6, pp. 6–8.

[114] Z. Tao, L. Dou, H. T., and J. C. Rasmussen, 'A Fast Method to Simulate the PDL Impact on Dual-Polarization Coherent Systems,' *IEEE Photon. Technol. Lett.*, vol. 21, no. 24, pp. 1882–1884, 2009.

[115] [1] D. Zwillinger, *Handbook of Differential Equations*, 3rd ed. New York, NY, USA: Academic Press, 1997.

[116] M. Schetzen, *The Volterra and Wiener theories of nonlinear systems*. New York, NY, USA: Wiley, 1980.

[117] M. J. Ablowitz and G. Biondini, 'Multiscale pulse dynamics in communication systems with strong dispersion management,' Opt. Lett., vol. 23, no. 21, pp. 1668–1670, 1998.

[118] N. Rossi, P. Serena and A. Bononi, "Stratified-Sampling Estimation of PDL-Induced Outage Probability in Nonlinear Coherent Systems," *J. Lightw. Technol.*, vol. 32, no. 24, pp. 4905–4911, 2014.

[119] E. Grellier and A. Bononi, 'Quality parameter for coherent transmissions with Gaussian-distributed nonlinear noise,' *Opt. Express*, vol. 19, no. 13, pp. 12781–12788, 2011.

[120] D. A. Morero, M. A. Castrillon, A. Aguirre, M. R. Hueda, and O. E. Agazzi, 'Design tradeoffs and challenges in practical coherent optical transceiver implementations,' *J. Lightw. Technol.*, vol. 34, no. 1, pp. 121–136, 2016.

[121] A. Alvarado, E. Agrell, D. Lavery, R. Maher, and P. Bayvel, 'Replacing the soft-decision FEC limit paradigm in the design of optical communication systems,' *J. Lightw. Technol.*, vol. 34, no. 2, pp. 707–721, 2016.

[122] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley, 2005.

[123] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai, 'On information rates for mismatched decoders,' *IEEE Trans. Inform. Theory*, vol. 40, no. 6, pp. 1953–1967, Nov. 1994.

[124] D. M. Arnold, H. A. Loeliger, P. O. Vontobel, A. Kavcic, and W. Zeng, 'Simulation-Based Computation of Information Rates for Channels With Memory,' *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3498–3508, Aug. 2006.

[125] A. Guillén i Fàbregas, A. Martinez, and G. Caire, 'Bit-interleaved coded modulation,' *Found. Trends Commun. Inf. Theory*, vol. 5, no. 1–2, pp. 1– 153, 2008.

[126] G. Caire, G. Taricco, and E. Biglieri, 'Bit-Interleaved Coded Modulation,' *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 927–946, 1998.

[127] M. Karlsson and E. Agrell, 'Multidimensional Modulation and Coding in Optical Transport,' *J. Lightw. Technol.*, vol. 35, no. 4, pp. 876–884, 2017.

[128] L. Schmalen, A. Alvarado, and R. Rios-Müller, 'Performance Prediction of Nonbinary Forward Error Correction in Optical Transmission Experiments,' *J. Lightw. Technol.*, vol. 35, no. 4, pp. 1015--1027, 2017.

[129] J. Cho, L. Schmalen, P. J. Winzer, N. B. Labs, and D.- Stuttgart, 'Normalized Generalized Mutual Information as a Forward Error Correction Threshold for Probabilistically Shaped QAM,' in *Proc. ECOC 2017, Goteborg, Sweden*, 2017, vol. M.2.D.2.

[130] R. E. Blahut, *Principles and Practice of Information Theory*, Addison Wesley, 1988.

[131] T. Yoshida, M. Karlsson, and E. Agrell, 'Performance Metrics for Systems with Soft-Decision FEC and Probabilistic Shaping,' *IEEE Photon. Technol. Lett.*, vol. xxxx, no. xxxx, pp. xxxx, 2018.

[132] A. O. Korotkevich and P. M. Lushnikov, 'Proof-of-concept implementation of the massively parallel algorithm for simulation of dispersion-managed WDM optical fiber systems.,' *Opt. Lett.*, vol. 36, no. 10, pp. 1851–1853, May 2011.

[133] P. M. Lushnikov, 'Fully parallel algorithm for simulating dispersion-managed wavelength-division-multiplexed optical fiber systems,' *Opt. Lett.*, vol. 27, no. 11, p. 939, 2002.

[134] E. Forestieri and M. Secondini, *Solving the Nonlinear Schrödinger Equation*, in: E. Forestieri (eds) Optical Communication Theory and Techniques. Springer, Boston, MA, 2005.

[135] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. New York, NY, USA: McGraw-Hill, 1991.

[136] V. E. Zakharov and L. A. Ostrovsky, 'Modulation instability: The beginning,' *Physica D*, vol. 238, no. 5, pp. 540–548, Mar. 2009.

[137] J. Hansryd, P. A. Andrekson, M. Westlund, J. Li, and P. O. Hedekvist, 'Fiber-based optical parametric amplifiers and their applications,' *IEEE J. Sel. Top. Quantum Electron.*, vol. 8, no. 3, pp. 506–520, 2002.

[138] M. E. Marhic, P. A. Andrekson, P. Petropoulos, S. Radic, C. Peucheret, and M. Jazayerifar, 'Fiber optical parametric amplifiers in optical communication systems,' *Laser Photon. Rev.*, vol. 9, no. 1, pp. 50–74, 2015.

[139] H. Kim and A. H. Gnauck, 'Experimental investigation of the performance limitation of DPSK systems due to nonlinear phase noise,' *IEEE Photon. Technol. Lett.*, vol. 15, no. 2, pp. 320–322, 2003.

[140] 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, no. 4, pp. 535–537, Apr. 1997.

[141] B. Xu and M. Brandt-Pearce, 'Optical fiber parametric-gain-induced noise coloring and amplification by modulated signals,' *J. Opt. Soc. Am. B*, vol. 21, no. 3, p. 499–513, Mar. 2004.

[142] P. Serena, A. Orlandini, and A. Bononi, "Parametric-Gain Approach to the Analysis of Single-Channel DPSK/DQPSK Systems With Nonlinear

Phase Noise," *J. Lightw. Technol.*, vol. 24, no. 5, pp. 2026–2037, May 2006.

[143] I. B. Djordjevic, B. Vasic, M. Ivkovic, and I. Gabitov, 'Achievable information rates for high-speed long-haul optical transmission,' *J. Lightw. Technol.*, vol. 23, no. 11, pp. 3755–3763, 2005.

[144] G. Colavolpe, T. Foggi, A. Modenini, and A. Piemontese, "Faster-than-Nyquist and beyond: how to improve spectral efficiency by accepting interference," Opt. Exp., vol. 19, pp. 26600-26609, 2011.

[145] P. Serena, "Nonlinear Signal–Noise Interaction in Optical Links With Nonlinear Equalization," *J. Lightw. Technol.*, vol. 34, no. 6, pp. 1476–1483, Mar. 2016.

[146] D. Lavery, D. Ives, G. Liga, A. Alvarado, S. J. Savory, and P. Bayvel, 'The Benefit of Split Nonlinearity Compensation for Single-Channel Optical Fiber Communications,' *IEEE Photon. Technol. Lett.*, vol. 28, no. 17, pp. 1803–1806, 2016.

[147] A. D. Ellis, M. E. McCarthy, M. A. Z. Al-Khateeb, and S. Sygletos, 'Capacity limits of systems employing multiple optical phase conjugators,' *Opt. Express*, vol. 23, no. 16, p. 20381, 2015.

[148] A. Ghazisaeidi, "A Theory of Nonlinear Interactions Between Signal and Amplified Spontaneous Emission Noise in Coherent Wavelength Division Multiplexed Systems," *J. Lightw. Technol.* vol. 35, no. 23, pp. 5150-5175, 2017.

[149] L. Beygi, N. Irukulapati, E. Agrell, P. Johannisson, M. Karlsson, H. Wymeersch, P. Serena, and A. Bononi, 'On nonlinearly-induced noise in single-channel optical links with digital backpropagation,' *Opt. Exp.*, vol. 21, pp. 26376–26386, Nov. 2013.

[150] V. Curri, A. Carena, and A. Arduino, 'Design Strategies and Merit of System Parameters for Uniform Uncompensated Links Supporting Nyquist-WDM Transmission,' *J. Lightw. Technol.*, vol. 33, no. 18, pp. 3921–3932, 2015.

[151] I. Tomkos, S. Azodolmolky, J. Sol?-Pareta, D. Careglio, E. Palkopoulou, 'A tutorial on the flexible optical networking paradigm: State of the art, trends, and research challenges,' Proc. IEEE, vol. 102, no. 9, pp. 1317 - 1337, Sept. 2014.

[152] G. Zhang, M. De Leenheer, A. Morea, and B. Mukherjee, "A Survey on OFDM-Based Elastic Core Optical Networking," IEEE Commun. Surv. & Tutorials, vol. 15, pp. 65-87, 1st quarter 2013.

[153] S., Angelou M., Tomkos I., Morea A., Pointurier Y., Sol??-Pareta J. "A Comparative Study of Impairments Aware Optical Networks Planning Tools". In: Tomkos I., Bouras C.J., Ellinas G., Demestichas P., Sinha P. (Eds) Broadband Communications, Networks, and Systems. BROADNETS 2010. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 66. Springer.

[154] E. Seve, P. Ramantanis, J.-C. Antona, E. Grellier, O. Rival, F. Vacondio and S. Bigo "Semi-Analytical Model for the Performance Estimation of 100Gb/s PDM-QPSK Optical Transmission Systems without Inline Dispersion Compensation and Mixed Fiber Types," in Proc. European Conf. on Opt. Commun. 2013, paper Th.1.D.2.

[155] R. Pastorelli, G. Bosco, S. Piciaccia, and F. Forghieri, 'Network planning strategies for next-generation flexible optical networks,' J. Opt. Commun. Netw., vol. 7, no. 3, pp. A511–A525, Mar. 2015.

[156] E. Palkopoulou, G. Bosco, A. Carena, D. Klonidis, P. Poggiolini, and I. Tomkos, ' Nyquist-WDM-based flexible optical networks: Exploring physical layer design parameters, ' J. Lightw. Technol., vol. 31, no. 14, pp. 2332–2339, 2013.

[157] S. J. Savory, "Congestion Aware Routing in Nonlinear Elastic Optical Networks," IEEE Photon. Technol. Lett., vol. 26, no. 10, pp. 1057-1060, 2015.

[158] V. Curri, M. Cantono, R. Gaudino, 'Elastic All-Optical Networks: A New Paradigm Enabled by the Physical Layer. How to Optimize Network Performances?,' *J. Lightwave Technol.*, vol. 35, pp. 1211-1221 (2017).

[159] D. J. Ives, A. Alvarado, and S. J. Savory, "Throughput Gains From Adaptive Transceivers in Nonlinear Elastic Optical Networks," J. Lightw. Technol. vol. 35, no. 6, pp. 1280-1289 (2017).

[160] P. Johannisson and E. Agrell, 'Modeling of Nonlinear Signal Distortion in Fiber-Optic Networks,' J. Lightw. Technol., vol. 32, no. 23, pp. 3942–3950, Dec. 2014.

[161] C. Rottondi, L. Barletta, A. Giusti and M. Tornatore, "A Machine Learning Method for Quality of Transmission Estimation of Unestablished Lightpaths," J. Opt. Commun. and Netw., 2018, to appear.

[162] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, Nonlinear Interference Noise Wizard, available at: http://nlinwizard.eng.tau.ac.il.

[163] L. Yan, E. Agrell, H. Wymeersch, P. Johannisson, R. Di Taranto, and M. Brandt-Pearce, 'Link-level resource allocation for flexible-grid nonlinear fiber-optic communication systems,' IEEE Photon. Technol. Lett., vol. 27, no. 12, pp. 1250–1253, Jun. 2015.

[164] L. Yan, E. Agrell, H. Wymeersch, and M. Brandt-Pearce, 'Resource allocation for flexible-grid optical networks with nonlinear channel model,' J. Opt. Commun. Netw., vol. 7, no. 11, pp. B101–B108, 2015.

[165] J. Zhao, H. Wymeersch, and E. Agrell, 'Nonlinear Impairment-Aware Static Resource Allocation in Elastic Optical Networks,' IEEE/OSA J. Lightw. Technol., vol. 32, no. 22, pp. 4554-4564, Nov. 2015.

[166] M. N. Dharmaweera, J. Zhao, L. Yan, M. Karlsson, and E. Agrell, 'Traffic-Grooming- and Multipath- Routing-Enabled Impairment-Aware Elastic Optical Networks,' J. Opt. Commun. Netw., Vol. 8, No. 2, Feb. 2016.

[167] D. J. Ives, P. Bayvel, S. J. Savory, "Adapting transmitter power and modulation format to improve optical network performance utilizing the Gaussian noise model of nonlinear impairments," J. Lightw. Technol. vol. 32, no. 21, p. 3485 (2014).

[168] D. J. Ives, P. Bayvel, and S. J. Savory, 'Routing, modulation, spectrum and launch power assignment to maximize the traffic throughput of a nonlinear optical mesh network,' Photon. Netw. Commun., vol. 29, no. 3, pp. 244–256 (2015).

[169] K. Christodoulopoulos, I. Tomkos, and E. A. Varvarigos, "Elastic Bandwidth Allocation in Flexible OFDM-Based Optical Networks," J. Lightw. Technol., vol. 29, pp. 1354-1366, May 2011.

[170] L. Yan, E. Agrell, M. N. Dharmaweera, and H. Wymeersch, "Joint Assignment of Power, Routing, and Spectrum in Static Flexible-Grid Networks," J. Lightw. Technol., vol. 35, no. 10, pp. 1766–1774, May 2017.

[171] Y. Xu, "Resource Allocation in Elastic Optical Networks with Physical-Layer Impairments", PhD thesis, University of Virginia, Aug. 2017.

[172] A. Bononi, P. Serena, A. Morea, and G. Picchi, "Regeneration Savings in Flexible Optical Networks with a New Load-aware Reach Maximization," invited paper, Opt. Switch. Netw. (2016), pp. 123-134.

[173] T. Panayiotou, S. P. Chatzis, and G. Ellinas, 'Performance Analysis of a Data-Driven Quality-of-Transmission Decision Approach on a Dynamic Multicast- Capable Metro Optical Network,' J. Opt. Commun. and Netw., vol. 9, no. 1, pp. 98-108, Jan. 2017.

[174] R. Hardin, F. D. Tappert, 'Applications of the split-step Fourier method to the numerical solution of nonlinear and variable coefficient wave equations,' *SIAM Rev. Chronicle*, vol. 15**,** p. 423, 1973.

[175] A. Hasegawa, F. Tappert, 'Transmission of stationary nonlinear optical pulses in dispersive dielectric fibers. I. Anomalous dispersion,' *Appl. Phys. Lett.*, vol. 23, pp. 142–144, 1973.

[176] T. R. Taha, M. J. Ablowitz, 'Analytical and numerical aspects of certain nonlinear evolution equation, ii, numerical, nonlinear schroedinger equation,' *J. Computat. Phys.*, vol. 5, pp. 203–230, 1984.

[177] O. V. Sinkin, R. Holzlöhner, J. Zweck, C. R. Menyuk, 'Optimization of the split-step Fourier method in modeling optical-fiber communications systems,' *J. Lightwave Technol.*, vol. 21, no. 1, pp. 61–68, Jan. 2003.

[178] Q. Zhang, M. I. Hayee, 'Symmetrized Split-Step Fourier Scheme to Control Global Simulation Accuracy in Fiber-Optic Communication Systems,' *J. Lightwave Technol.*, vol. 26, no. 2, pp. 302-316, Jan.15, 2008.

[179] A. V. Oppenheim, R. W. Schafer, *Discrete-time signal processing*, Prentice Hall, 1999.

[180] J. Shao, X. Liang, S. Kumar, 'Comparison of split-step Fourier schemes for simulating fiber optic communication systems,' *IEEE Photonics J.*, vol. 6, no. 4, 2014.

[181] S. Verdú, 'Fifty years of Shannon theory,' *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2057--2078, 1998.

[182] 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. Opt.*, vol. 18, no. 6, p. 063002, 2016. [Online]. Available: \url{http://stacks.iop.org/2040-8986/18/i=6/a=063002}

[183] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: Wiley, 1968.

[184] F. D. Neeser and J. L. Massey, 'Proper complex random processes with applications to information theory,' *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1293–1302, Jul. 1993.

[185] G. Kramer, M. Yousefi, and F. Kschischang, 'Upper bound on the capacity of a cascade of nonlinear and noisy channels,' in *Proc. IEEE Inf. Theory Workshop*, Apr. 2015, pp. 1–4.

[186] E. Agrell and M. Karlsson, 'Influence of behavioral models on multiuser channel capacity,' *J. Lightw. Technol.*, vol. 33, no. 17, pp. 3507–3515, Sep. 2015.

[187] D. M. Arnold, H.-A. Loeliger, P. O. Vontobel, A. Kavvcic, and W. Zeng, 'Simulation-based computation of information rates for channels

with memory,' *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3498–3508, Aug. 2006.

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

[189] R. Dar, M. Shtaif, and M. Feder, 'New bounds on the capacity of the nonlinear fiber-optic channel,' *Opt. Lett.*, vol. 39, no. 2, pp. 398–401, Jan. 2014.

[190] 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. Exhib.*, 2015, Paper Th4D.3.

[191] E. Agrell, 'Conditions for a monotonic channel capacity,' *IEEE Trans. Commun.*, vol. 63, no. 3, pp. 738–748, Mar. 2015.

[192] M. Secondini and E. Forestieri, 'Scope and limitations of the nonlinear Shannon limit,' *J. Lightwave Technol.*, vol. 35, no. 4, pp. 893–902, 2017.

[193] M. I. Yousefi, "The Asymptotic Capacity of the Optical Fiber", available online https://arxiv.org/abs/1610.06458.

[194] J. C. Cartledge, F. P. Guiomar, F. R. Kschischang, G. Liga, M. P. Yankov, "Digital signal processing for fiber nonlinearities," *Opt. Exp.*, vol. 25, no. 3, pp. 1916–1936, Feb. 2017.

[195] R.-J. Essiambre and P. J. Winzer, 'Fibre nonlinearities in electronically pre-distorted transmission,' in *Proc. Eur. Conf. Opt. Commun.*, 2005, Paper Tu.3.2.2.

[196] X. Li, X. Chen, G. Goldfarb, E. Mateo, I. Kim, F. Yaman, G. Li, 'Electronic post-compensation of WDM transmission impairments using coherent detection and digital signal processing,' *Opt. Express*, vol. 16, pp. 880–888, 2008.

[197] L. Li, Z. Tao, L. Dou, W. Yan, S. Oda, T. Tanimura, T. Hoshida, and J. C. Rasmussen, 'Implementation efficient nonlinear equalizer based on correlated digital backpropagation,' presented at the Optical Fiber Communication Conf., Los Angeles, CA, USA, 2011, Paper OWW3.

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

[199] A. Napoli, Z. Maalej, V. A. J. M. Sleiffer, M. Kuschnerov, D. Rafique, E. Timmers, B. Spinnler, T. Rahman, L. D. Coelho, and N. Hanik, 'Reduced complexity digital back-propagation methods for optical communication systems,' *J. Lightw. Technol.*, vol. 32, no. 7, pp. 1351–1362, 2014.

[200] Z. Tao, L. Dou, Y. Zhao, W. Yan, T. Oyama, S. Oda, T. Tanimura, T. Hoshida, and J. C. Rasmussen, 'Nonlinear Compensation beyond Conventional Perturbation Based Algorithms,' in *Proc. ACP/IPOC 13*, 2013, vol. ATh4E.1, pp. 4–6.

[201] Q. Zhuge, M. Reimer, A. Borowiec, M. O'Sullivan, and D. V. Plant, 'Aggressive quantization on perturbation coefficients for nonlinear pre-distortion,' in *Proc. OFC 2014*, 2014, vol. Th4D.7.

[202] Z. Tao, L. Dou, W. Yan, L. Li, T. Hoshida, and J. C. Rasmussen, 'Multiplier-Free Intrachannel Nonlinearity Compensating Algorithm Operating at Symbol Rate,' *J. Lightw. Technol.*, vol. 29, no. 17, pp. 2570–2576, 2011.

[203] Y. Gao, J. C. Cartledge, A. S. Karar, S. S.-H. Yam, M. O'Sullivan, C. Laperle, A. Borowiec, and K. Roberts, 'Reducing the complexity of perturbation based nonlinearity pre-compensation using symmetric EDC and pulse shaping,' *Opt. Exp.*, vol. 22, no. 2, p. 1209, Jan. 2014.

[204] F. P. Guiomar and A. Pinto, 'Simplified Volterra Series Nonlinear Equalizer for Polarization-Multiplexed Coherent Optical Systems,' *J. Lightw. Technol.*, vol. 31, no. 23, pp. 3879–3891, 2013.

[205] J. Gonçalves, C. S. Martins, F. P. Guiomar, T. Cunha, J. Pedro, A. N. Pinto, and P. Lavrador, 'Nonlinear compensation with DBP aided by a memory polynomial,' *Opt. Exp.*, vol. 24, no. 26, pp. 30309–30316, 2016.

[206] L. Zhu and G. Li, 'Folded digital backward propagation for dispersion-managed fiber-optic transmission,' *Opt. Exp.*, vol. 19, no. 7, p. 5953, 2011.

[207] X. Liang and S. Kumar, 'Multi-stage perturbation theory for compensating intra-channel nonlinear impairments in fiber-optic links,' *Opt. Exp.*, vol. 22, no. 24, pp. 29733–29745, 2014.

[208] A. Ghazisaeidi and R.-J. Essiambre, "Calculation of coefficients of perturbative nonlinear pre-compensation for Nyquist pulses," in *Proc. European Conference on Optical Communications (ECOC)*, Cannes (FR), Sept. 2014.

[209] A. Ghazisaeidi, I. Fernandez de Jauregui Ruiz, L. Schmalen, P. Tran, C. Simonneau, E. Awwad, B. Uscumlic, P. Brindel, and G. Charlet, "Submarine transmission systems using digital nonlinear compensation and adaptive rate forward error correction," *J. Lightw. Technol.*, vol. 34, no. 8, pp. 1886–1895, Apr. 2016.

[210] D. Rafique, 'Fiber Nonlinearity Compensation: Commercial Applications and Complexity Analysis,' *J. Lightw. Technol.*, vol. 34, no. 2, pp. 544--553, 2016.

[211] G. Liga, T. Xu, A. Alvarado, R. I. Killey, and P. Bayvel, 'On the performance of multichannel digital backpropagation in high-capacity long-haul optical transmission,' *Opt. Express*, vol. 22, pp. 30053–30062, 2014.

[212] 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, no. 3, pp. 493–502, 2016.

[213] M. Secondini, D. Marsella, and E. Forestieri, 'Enhanced split-step fourier method for digital backpropagation,' in *Proc. European Conf. on Opt. Commun.* 2014, paper We.3.3.5.

[214] K. Goroshko, H. Louchet, and A. Richter, 'Overcoming performance limitations of digital back propagation due to polarization mode dispersion,' in *Proc. of International Conference on Transparent Optical Networks (ICTON)*, Trento, Italy, 2016, p. Mo.B1.4.

[215] C. B. Czegledi, G. Liga, D. Lavery, M. Karlsson, E. Agrell, S. J. Savory, and P. Bayvel, 'Polarization-mode dispersion aware digital backpropagation,' in *Proc. of European Conference on Optical Communication (ECOC)*, Dusseldorf, Germany, 2016, pp. 1091–1093.

[216] 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. Express*, vol. 25, pp. 1903–1915, 2017.

[217] N. V. Irukulapati, H. Wymeersch, P. Johannisson, and E. Agrell, 'Stochastic digital backpropagation,' *IEEE Transactions on Communications*, vol. 62, no. 11, pp. 3956–3968, Nov. 2014.

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

[219] 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, no. 9, pp. 2256–2266, May 2016.

[220] A. Napoli, M. Kuschnerov, C.-Y. Lin, B. Spinnler, M. Bohn, D. Rafique, V. A. Sleiffer, and B. Schmauss, 'Adaptive digital backpropagation for optical communication systems,' in *Proc. Opt. Fiber Commun. Conf.*, 2014, Paper M3C. 4.

[221] E. P. da Silva, R. Asif, K. J. Larsen, and D. Zibar, 'Nonlinear compensation with modified adaptive digital backpropagation in flexigrid networks,' in *Proc. Conference on Lasers and Electro-Optics* 2015, San Jose, United States, 2015, Paper SM2M.

[222] C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura, 'Method for solving the Korteweg-de Vries equation,' *Phys. Rev. Lett.*, vol. 19, no. 19, pp. 1095–1097, Nov. 1967.

[223] V. E. Zakharov and A. B. Shabat, 'Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of wave in nonlinear media,' *Soviet J. Exp. Theory Phys. (JETP)*, vol. 34, no. 1, pp. 62–69, Jan. 1972.

[224] S. V. Manakov, 'On the theory of two-dimensional stationary self-focusing of electromagnetic waves,' *Soviet J. Exp. Theory Phys. (JETP)*, vol. 38, pp. 248–253, Feb. 1974.

[225] M. J. Ablowitz and H. Segur, *Solitons and the Inverse Scattering Transform*. Philadelphia, PA, USA: SIAM, 1981.

[226] M. I. Yousefi and F. R. Kschischang, 'Information transmission using the nonlinear Fourier transform, Part I–III,' *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 4312–4369, Jul. 2014.

[227] S. K. Turitsyn, J. E. Prilepsky, S. Thai Le, S. Wahls, L. L. Frumin, M. Kamalian, and S. A. Derevyanko, 'Nonlinear Fourier transform for optical data processing and transmission: advances and perspectives,' *Optica*, vol. 4, no. 3, pp. 307-322, Mar. 2017.

[228] A. Hasegawa and Y. Kodama, *Solitons in Optical Communications*. Oxford, U.K.: Clarendon Press, 1995.

[229] A. Hasegawa and T. Nyu, 'Eigenvalue communication,' *IEEE J. Lightw. Technol.*, vol. 11, no. 3, pp. 395–399, Mar. 1993.

[230] J. E. Prilepsky, S. A. Derevyanko, K. J. Blow, I. Gabitov, and S. K. Turitsyn, 'Nonlinear inverse synthesis and eigenvalue division multiplexing in optical fiber channels,' *Phys. Rev. Lett.*, vol. 113, p. 013901, Jul. 2014.

[231] S. A. Derevyanko, J. E. Prilepsky, and S. K. Turitsyn, 'Capacity estimates for optical transmission based on the nonlinear Fourier transform,' *Nature Commun.*, vol. 7, Jul. 2016, Art. no. 12710.

[232] E. G. Turitsyna and S. K. Turitsyn, 'Digital signal processing based on inverse scattering transform,' *Opt. Lett.*, vol. 38, no. 20, pp. 4186–4188, Oct. 15th 2013.

[233] H. Bülow, 'Experimental demonstration of optical signal detection using nonlinear Fourier transform,' *J. Lightwave Technol.*, vol. 33, no. 7, pp. 1433–1439, Apr. 2015.

[234] H. Terauchi and A. Maruta, 'Eigenvalue modulated optical transmission system based on digital coherent technology,' in *OptoElectronics and Communications Conference (OECC)*, Kyoto, Japan, 2013, paper WR2-5.

[235] S. Civelli, E. Forestieri and M. Secondini, "Decision-feedback detection strategy for nonlinear frequency-division multiplexing", *Opt. Express*, vol. 26, no. 9, pp. 12057-12071, Apr. 2018.

[236] H. Bülow,, V. Aref, and W. Idler, 'Transmission of waveform determined by 7 eigenvalues with PSK-modulated spectral amplitude,' in *European Conference on Optical Communications (ECOC)*, Germany, 2016, paper Tu.3.E.2.

[237] S. Civelli, E. Forestieri and M. Secondini, 'Why Noise and Dispersion May Seriously Hamper Nonlinear Frequency-Division Multiplexing,' *IEEE Photon. Technol. Lett.*, vol. 29, no. 16, pp. 1332-1335, Aug.15, 2017.

[238] V. Aref, S. Thai Le, and H. Bülow,, 'Demonstration of fully nonlinear spectrum modulated system in the highly nonlinear optical transmission regime,' in *Proc. European Conference on Optical Communication (ECOC) – postdeadline paper*, Germany, 2016.

[239] S. Wahls and H. V. Poor, "Fast numerical nonlinear Fourier transforms," *Trans. Inf. Theory*, vol. 61, no. 12, pp. 6957–6974, 2015.

[240] M. Kamalian, J. E. Prilepsky, S. T. Le, and S. K. Turitsyn, 'Periodic nonlinear Fourier transform for fiber-optic communications, Part I–II,' *Opt. Express*, vol. 24, no. 16, pp. 18353–18381, Aug. 2016.

[241] I. T. Lima, T. D. S. DeMenezes, V. S. Grigoryan, M. O'sullivan, and C. R. Menyuk, 'Nonlinear Compensation in Optical Communications Systems With Normal Dispersion Fibers Using the Nonlinear Fourier Transform,' *J. Lightw. Technol.*, vol. 35, no. 23, pp. 5056–5068, Dec. 2017.

[242] A. Maruta and Y. Matsuda, 'Polarization division multiplexed optical eigenvalue modulation,' in *International Conference on Photonics in Switching (PS)*, Florence, Italy, 2015, pp. 265–267.

[243] S. Gaiarin, A. M. Perego, E. P. da Silva, F. Da Ros, and D. Zibar, 'Experimental Demonstration of Dual Polarization Nonlinear Frequency Division Multiplexed Optical Transmission System,' in *Proc. European Conference on Optical Communications*, Sweden, 2017.

[244] 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.

[245] 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.

[246] T. Fehenberger, A. Alvarado, G. Bocherer, 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.

[247] 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.

[248] 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.

[249] 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.

[250] S. Blanes, F. Casas, J. Oteo, and J. Ros, 'The Magnus expansion and some of its applications,' *Physics Reports*, vol. 470, no. 5–6, pp. 151–238, Jan. 2009.

[251] M. Reimer, D. Yevick, and D. Dumas, 'The accuracy of the Magnus expansion for polarization-mode dispersion and polarization- dependent loss,' *J. Lightw. Technol.*, vol. 26, no. 19, pp. 3337–3344, Oct. 2008.

[252] M. Secondini and E. Forestieri, 'A Comparative Analysis of Different Perturbation Models for the Nonlinear Fiber Channel,' in *Asia Communications and Photonics Conference 2015*, OSA Technical Digest (online), paper AM3D.1.

[253] A. G. Green, P. B. Littlewood, P. P. Mitra, and L. G. L. Wegener, 'Schroedinger equation with a spatially and temporally random potential: Effects of cross-phase modulation in optical communication,' *Phys. Rev. E*, vol. 66, no. 4, 2002.

[254] M. Sorokina, S. Sygletos and S. Turitsyn, 'Ripple distribution for nonlinear fiber-optic channels,' *Opt. Exp.*, vol. 25, no. 3, pp. 2228–2238, Feb. 2017.

[255] M. Secondini, E. Agrell, E. Forestieri, D. Marsella, 'Fiber Nonlinearity Mitigation in WDM Systems: Strategies and Achievable Rates,' in *Proc. European Conference and Exhibition on Optical Communications*, Sweden, 2017.

[256] 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.

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

[258] Matlab support website:
`https://it.mathworks.com/help/matlab/functionlist.html`

[259] P. Poggiolini, Y. Jiang, A. Carena, F. Forghieri, 'Analytical Modeling of the Impact of Fiber Non-Linear Propagation on Coherent Systems and Networks,' book chapter in *Enabling Technologies for High Spectral-Efficiency Coherent Optical Communication Networks*, Editors Xiang Zhou and Chongjin Xie, John Wiley & Sons, Inc., Apr. 13th 2016.

[260] P. Poggiolini, Y. Jiang 'Recent Advances in the Modeling of the Impact of Nonlinear Fiber Propagation Effects on Uncompensated Coherent Transmission Systems,' Tutorial Review, *J. of Lightw. Technol.*, vol. 35, no. 3, pp. 458-480, Feb. 2017.

[261] P. Poggiolini, G. Bosco, A. Carena, F. Guiomar, M. Ranjbar Zefreh, F. Forghieri, S. Piciaccia, 'Non-Linearity Modeling at Ultra-High Symbol Rates,' in *Proc. of OFC 2018*, paper W1G.3, San Diego (CA), Mar. 2018.

[262] P. Poggiolini, G. Bosco, A. Carena, D. Pilori, A. Nespola, M. Ranjbar Zefreh, M. Bertino, F. Forghieri, 'Non-Linearity Modeling for Gaussian-Constellation Systems at Ultra-High Symbol Rates,' in *Proc. of ECOC 2018*, paper Tu4G.3, Rome (IT), Sept. 2018.

[263] D. Semrau, R. I. Killey, P. Bayvel, 'A Closed-Form Approximation of the Gaussian Noise Model in the Presence of Inter-Channel Stimulated Raman Scattering,' www.arXiv.org, paper: arXiv:1808.07940 [eess.SP], Aug. 2018.

[264] P. Poggiolini, 'A generalized GN-model closed-form formula,' www.arXiv.org, paper arXiv:1810.06545 [eess.SP], Sept. 2018.

[265] D. Semrau, G. Saavedra, D. Lavery, R. I. Killey, P. Bayvel 'A Closed-Form Expression to Evaluate Nonlinear Interference in Raman-Amplified Links' *J. Lightw. Technol.* , vol. 35, no. 19, pp. 4316–4328, Oct. 2017.

[266] M. Cantono, J.-L. Auge, V. Curri, 'Modelling the Impact of SRS on NLI Generation in Commercial Equipment: an Experimental Investigation,' in *Proc. of OFC 2018*, paper M1D.2, San Diego (CA), Mar. 2018.

[267] M. Cantono, D. Pilori, A. Ferrari, C. Catanese, J. Thouras, J.L. Auge, and V. Curri, 'On the Interplay of Nonlinear Interference Generation with Stimulated Raman Scattering for QoT Estimation,' to appear on *J. Lightw. Technol.*.

[268] P. Poggiolini, A. Nespola, Y. Jiang, G. Bosco, A. Carena, L. Bertignono, S. M. Bilal, S. Abrate, F. Forghieri, 'Analytical and Experimental Results on System Maximum Reach Increase Through Symbol Rate Optimization,' *J. Lightw. Technol.* , vol. 34, no. 8, pp. 1872–1885, Apr. 2016.

[269] V. Kamalov, L. Jovanovski, V. Vusirikala, S. Zhang, F. Yaman, K. Nakamura, T. Inoue, E. Mateo, Y. Inada 'Evolution from 8QAM live traffic to PS 64-QAM with Neural-Network Based Nonlinearity Compensation on 11000 km Open Subsea Cable', in in *Proc. of OFC 2018*, post-deadline paper Th4D.5, San Deigo (CA), Feb. 2018.

[270] A. D. Ellis, M. A. Zaki Al Khateeb, M. E. McCarthy 'Impact of optical phase conjugation on the nonlinear Shannon limit,' *J. Lightw. Technol.* , vol. 35, no. 4, pp. 792–798, Feb. 2017.