# The Nonlinear Fiber-optic Channel: Modeling and Achievable Information Rate

**E. Forestieri and M. Secondini**
TeCIP Institute, Scuola Superiore Sant'Anna, Italy

**Abstract**— The problem of evaluating the channel capacity of a nonlinear fiber-optic system is reviewed. After properly defining the nonlinear fiber-optic channel, we consider a few alternative models and discuss their suitability for system design and performance evaluation. It is shown that using an accurate model can allow devising modulation and detection schemes providing higher robustness to nonlinear effects than those obtainable by simpler models.

## 1. INTRODUCTION

The evaluation of channel capacity in nonlinear fiber-optic systems is still an open issue. The main difficulty is due to the unavailability of an exact and mathematically tractable channel model. In fact, most of the results obtained so far either refer to highly simplified models or provide loose bounds to the capacity [1–15]. A useful and practical approach to lower bound the capacity is that of evaluating the achievable information rate for an arbitrarily fixed input distribution (modulation format) and a suboptimal decoder optimized for an approximate version of the channel (mismatched decoding). In this context, the capacity problem can be seen as a joint optimization of the achievable information rate over both the input distribution and the channel model [10, 15, 18]. It is thus clear that finding a good channel model for the optical fiber is the first step to design more efficient systems and derive tight bounds to the capacity.

According to the popular Gaussian noise (GN) model [19, 20], nonlinearity is simply modeled as additive white Gaussian noise: it cannot be compensated for, but it can be easily accounted for in the evaluation of channel capacity, leading to the conclusion that capacity vanishes at high powers. This result, however, is not supported by theory and is contradicted by some counter examples. Though it is now clear that the GN model has some limitations [15, 21, 22], it is still unclear whether these limitations are only of theoretical interest or do have a practical value.

The choice of the right model depends on the problem at hand: while numerical methods, such as the split-step Fourier method (SSFM) [23], are suitable for compensating deterministic nonlinearity through the digital backpropagation (DBP) technique [24], analytical approximations — based, for instance, on perturbation methods or Volterra series expansion — are more suitable for modeling stochastic effects such as signal noise-interaction and inter-channel nonlinearity [13, 25, 29]. Finally, an exact analytical model, accounting only for dispersion and nonlinearity, can be obtained by employing the inverse scattering (nonlinear Fourier) transform (IST) [30]. It is still unclear, however, whether this approach is competitive with the split-step Fourier method in terms of computational complexity and whether it can be used for modeling stochastic nonlinear effects.

By employing the aforementioned models, it is possible to show that the optimum modulation and detection strategies differ from those designed according to the GN model and typically used in fiber-optic systems. Therefore, in the second part of the paper, we consider different modulation and detection schemes aimed at providing higher robustness to nonlinear effects and investigate their actual advantage over GN-model-based strategies in terms of achievable information rate. Depending on the considered scenario, a non-negligible improvement is achieved. However, at high powers, the gap between lower and upper bounds to channel capacity remains large and does not allow to establish whether capacity is bounded or not. Possible directions for reducing the gap are finally discussed.

## 2. THE OPTICAL FIBER CHANNEL

In this section, we define the optical fiber channel and shortly discuss and compare different channel models available in the literature. We consider a generic multi-span optical fiber link, employing single-mode optical fibers and optical amplifiers exactly recovering the attenuation. Neglecting polarization effects (whose impact is shortly discussed at the end of this section) and accounting for attenuation/amplification, group velocity dispersion, Kerr nonlinearity, and amplified spontaneous emission (ASE) noise, the nonlinear evolution equation that governs the propagation of the normalized complex envelope of the optical signal $u(z,t)$ through the optical fiber channel is

$$\frac{\partial u}{\partial z} = j\frac{\beta_2}{2}\frac{\partial^2 u}{\partial t^2} - ja(z)\gamma|u|^2 u + n(z,t) \tag{1}$$

where $\beta_2$ is the group-velocity-dispersion parameter, $\gamma$ is the nonlinear coefficient, $a(z)$ accounts for variations of the average signal power along the link due to attenuation or amplification[1], and $n(z,t)$ is a forcing term that represents noise injected by optical amplifiers. The optical fiber channel can be considered as a waveform channel that, given the input waveform $u(0,t)$, produces the output waveform $u(z,t)$. In order to model and characterize the channel, it is customary to resort to a discrete-time channel model. This can be done by making some assumptions on the modulation and demodulation stages at the input and output of the channel, respectively. This assumptions usually corresponds to some constraints (e.g., bandwidth constraints) on the possible waveforms and have, therefore, an impact on the capacity of the resulting discrete-time channel.

We start by describing a single-user system (often referred to as single-channel system), in which only one user has access to the optical fiber. At the modulation stage, we assume that the input waveform $u(0,t)$ is strictly band-limited with low-pass bandwidth $B$, which is equivalent to say that the channel input can be represented by the sequence of symbols $\mathbf{x} = \{x_1, \ldots, x_N\}$, which are mapped at rate $1/T = 2B$ onto the continuous waveform according to

$$u(0,t) = \sum_{k=1}^{N} x_k g(t - kT) \tag{2}$$

where $g(t) = \mathrm{sinc}(2Bt)$. Because of the nonlinearity of (1), the bandwidth of the output waveform can be wider than $B$. Therefore, as opposed to the case of a linear channel, a demodulator with bandwidth $B$ is not enough to provide, in general, a sufficient statistic. Nevertheless, in order to obtain a discrete-time representation of the output and a practical detection scheme, we assume that also the demodulator is band-limited, with a low-pass bandwidth $\eta B$, where $\eta \geq 1$ is a parameter that accounts for a possible spectral broadening of the waveforms [31]. In principle, $\eta$ can be set arbitrarily large to obtain a sufficient statistic. However, this is not practical and we will rather consider $\eta$ as a technological constraint on the analog demodulator bandwidth that limits the achievable information rate on the channel. Thus, the received waveform $u(L,t)$ is filtered by a low-pass filter with transfer function $H(f) = \mathcal{F}\{h(t)\} = \mathrm{rect}\big(f/(2\eta B)\big)$ and sampled at rate $2\eta B$ to obtain the output samples $\mathbf{y} = \{y_1, \ldots, y_{\eta N}\}$. The equivalent low-pass model of the system is represented in Fig. 1(a), where $G(f) = \mathcal{F}\{g(t)\}$.

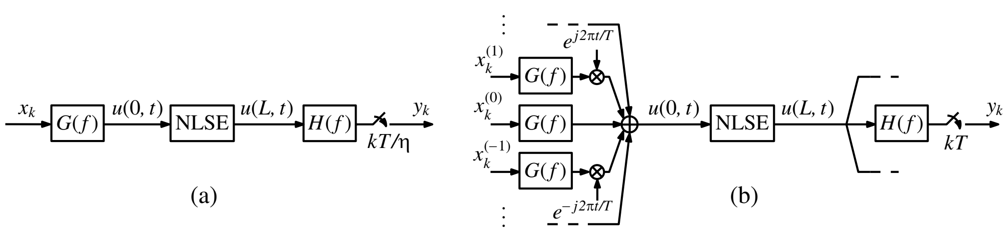



Figure 1: Equivalent low-pass system model: (a) single-user, (b) multi-user.

For the multi-user scenario, we consider WDM as the most practical (and widely employed) multiplexing technique. In this case, the input waveform is obtained by multiplexing different waveforms generated by different users according to (see Fig. 1(b))

$$u(0,t) = \sum_{\ell=-M}^{M}\sum_{k=1}^{N} x_k^{(\ell)} g(t - kT) e^{j2\pi\ell t/T} \tag{3}$$

where $\mathbf{x}^{(\ell)} = \{x_1^{(\ell)}, \ldots, x_N^{(\ell)}\}$ is the sequence of symbols transmitted by user $\ell$ and $2M+1$ is the number of users. We look at this system from the perspective of the central user ($\ell = 0$), assuming

[1]For a link with $K$ spans of same length $D$, it would be $a(z) = \sum_{k=0}^{K-1} e^{-\alpha(z-kD)}\big[u(z-kD) - u(z-kD-D)\big]$, where $\alpha$ is the attenuation parameter and $u(x)$ the unit-step function.

that the other users are out of his control and are a source of disturbance. In the sequel, by letting $\mathbf{x}^{(0)} = \mathbf{x}$, this scenario is formally described as in the single-user case, but replacing (2) with (3). Moreover, we set $\eta = 1$ since we assume that each user can demodulate only its "own" frequency band.

Equations (1)–(3) define the described channels only in an implicit way, as they do not provide an explicit expression of the output $\mathbf{y}$ given the input $\mathbf{x}$. Such an explicit knowledge is fundamental when dealing with communication and information theory problems, such as system design or evaluation of performance and information theoretical limits. We distinguish between two different types of channel models: a deterministic model, which solves the propagation Equation (1) when all the involved quantities are deterministic; and a stochastic model, which provides the conditional distribution $p(\mathbf{y}|\mathbf{x})$ of the output given the input. Usually, the former is just a part of the latter. For instance, statistics of the stochastic model can be represented by collecting many random realizations obtained from the deterministic model; or a specific deterministic solution (e.g., noise free) can be part of the stochastic model when considering a perturbation approach. The deterministic model can be important even by itself, for instance to compensate for nonlinear effects through digital backpropagation.

Unfortunately, even in the deterministic case, Equation (1) cannot be exactly solved. However, it can be approximated with arbitrary accuracy by numerical methods. The SSFM is the most used one and allows to trade off between accuracy and complexity by selecting the step size. When neglecting the noise term, and assuming a constant power along the link (e.g., loss exactly compensated by distributed amplification), Equation (1) is exactly integrable by the IST method, which is the analogous of the Fourier transform for nonlinear systems. The optical waveform $u(z, t)$ is represented in the spectral domain by means of a set of scattering data, whose propagation is governed by a simple linear equation. While the propagation step is trivial, the operations of direct and inverse scattering (going from time domain to spectral domain and back) are quite involved compared to direct and inverse Fourier transform. Their numerical implementation poses many issues in terms of accuracy and complexity and is a subject of current research [32, 33]. Perturbation methods can also be used to solve (1) [13, 26–29]. Formally, an approximate (zeroth-order) solution is found by setting $\gamma = 0$ in (1), and then expressing the exact solution as a power series of $\gamma$, whose coefficients are evaluated by substituting the power series back into (1).

In the deterministic case, the most relevant issue is the required computational complexity to achieve a prescribed accuracy. As mentioned before, the SSFM is probably the most effective numerical approach in this sense. Perturbation methods and Volterra series provide explicit finite-order solutions, but have a limited accuracy (at a given order) and a higher computational complexity than SSFM when used to described the whole link. Also the IST, so far, does not offer significant advantages over the SSFM in terms of computational complexity, even if more efficient algorithms are being developed. In terms of accuracy and complexity, the most promising approach seems to be that of combining perturbation methods with the SSFM approach to obtain an enhanced SSFM [34]: the idea is that of keeping the SSFM approach of dividing the fiber into many (dispersive and nonlinear) steps, but using perturbation methods to increase the accuracy of each step. In fact, through this approach, an experimental implementation of DBP with a reasonably low complexity (comparable to a conventional frequency-domain dispersion equalizer) has been demonstrated [35], showing that DBP is a feasible processing strategy.

## 3. ACHIEVABLE INFORMATION RATE

The evaluation of channel capacity [36], both for the single-user or multi-user scenario described in the previous section, is still an open issue, mainly because of the unavailability of an exact stochastic model $p(\mathbf{y}|\mathbf{x})$. In fact, even the information rate $I(\mathbf{X}; \mathbf{Y})$ for a given input distribution $p(\mathbf{x})$ is hard to evaluate without explicit knowledge of $p(\mathbf{y}|\mathbf{x})$. In this case, it is useful to resort to the concept of mismatched decoding and introduce the following auxiliary-channel lower bound [16–18]

$$\hat{I}(\mathbf{X}; \mathbf{Y}) \triangleq \lim_{N \to \infty} \frac{1}{N} E\left\{ \log \frac{q(\mathbf{y}|\mathbf{x})}{q_P(\mathbf{y})} \right\} \leq I(\mathbf{X}; \mathbf{Y}) \tag{4}$$

In (4), $q(\mathbf{y}|\mathbf{x})$ and $q_P(\mathbf{y})$ are, respectively, the conditional distribution for an arbitrary auxiliary channel and the output distribution obtained by connecting input (with distribution $p(\mathbf{x})$) to the same auxiliary channel. Expectation $E\{\cdot\}$, on the other hand, is taken with respect to the distribution $p(\mathbf{x})p(\mathbf{y}|\mathbf{x})$ of the real channel. The importance of the quantity defined in (4) is in its properties, which hold for any real and auxiliary channel: it is a lower bound to $I(\mathbf{X};\mathbf{Y})$ on the real channel; it is achievable by the maximum *a posteriori* probability (MAP) detector designed for the selected auxiliary channel; and it can be simply evaluated through simulations [17, 18] (the expectation with respect to $p(\mathbf{x})p(\mathbf{y}|\mathbf{x})$ can be computed by averaging over many realizations, without explicit knowledge of $p(\mathbf{y}|\mathbf{x})$). In the following, we will refer to (4) as the achievable information rate (AIR) for a given input distribution and mismatched decoder. According to this definition, it becomes clear what is the role of the approximated channel models introduced in the previous section: they provide the auxiliary-channel distribution $q(\mathbf{y}|\mathbf{x})$, which is sufficient to evaluate the corresponding AIR and to design a detector to achieve it. The actual information rate (and the optimum detector) is therefore obtained through a maximization of the AIR over all possible channel models. Finally, channel capacity is obtained as a maximization of the information rate over all possible input distributions $p(\mathbf{x})$ [36]. In the following, we will focus on the first maximization step (over channel models), assuming a fixed input distribution $p(\mathbf{x})$. This will provide us with different AIRs, corresponding to different lower bounds to the information rate (and, hence, to channel capacity). Moreover, in the multi-user scenario, we will assume that all users transmit with same input distribution and power (corresponding to behavioral model $c$ in [37]).

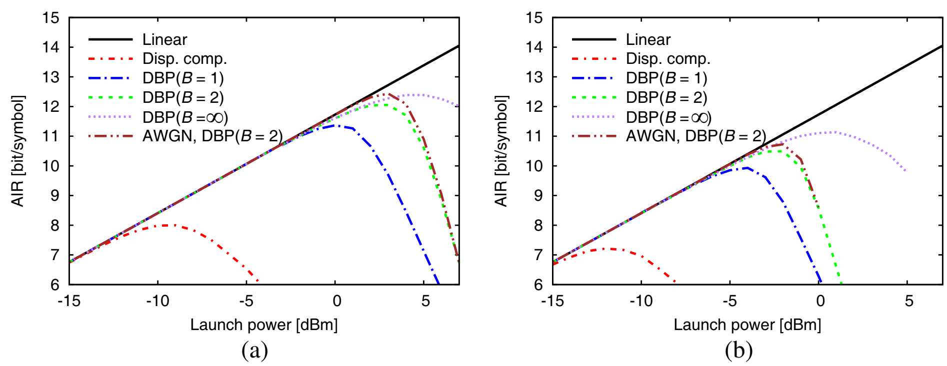


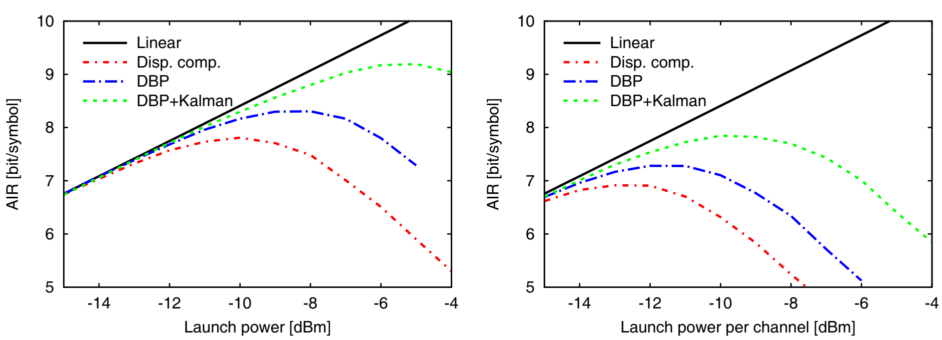

Figure 2: AIR for a single-channel system with different demodulation and detections: (a) standard fiber ($D = 17\,\mathrm{ps/nm/km}$); (b) low-dispersion fiber ($D = 2\,\mathrm{ps/nm/km}$).

For a single-user scenario, we consider an equivalent low-pass bandwidth $B = 25$ GHz, corresponding to an optical bandwidth of 50 GHz and a symbol rate of $1/T = 50$ GBd, and assume that the elements of $\mathbf{x}$ are i.i.d. Gaussian symbols (the capacity-achieving distribution in the absence of nonlinear effects). The optical fiber link is a 1000 km dispersion-unmanaged link with ideal distributed amplification (the distributed gain exactly equals fiber loss, such that $a(z) = 1$, and the spontaneous emission factor is set to 1). The transmission fiber is a single-mode fiber with attenuation coefficient $\alpha = 0.2\,\mathrm{dB/km}$ and nonlinear coefficient $\gamma = 1.27\,\mathrm{W}^{-1}\mathrm{km}^{-1}$. Polarization effects are neglected, such that the propagation of the complex envelope is governed by (1). Two different values of the dispersion coefficients are considered: $D = 17\,\mathrm{ps/nm/km}$ (a standard fiber) and $D = 2\,\mathrm{ps/nm/km}$ (a low-dispersion fiber). Figs. 2(a) and (b) show the corresponding AIRs, obtained by using different receivers optimized for different approximated channel models. The first receiver is designed by completely neglecting nonlinear effects, i.e., assuming that the received signal is affected only by dispersion and AWGN. The output samples $\mathbf{y}$ are obtained from the received waveform $u(L, t)$ by means of the demodulation scheme shown in Fig. 1, with $\eta = 1$ (which, in the absence of nonlinear effects, provides a sufficient statistic). After linear dispersion compensation, they are finally sent to a MAP detector optimized for the AWGN channel. The second receiver has the same demodulation bandwidth ($\eta = 1$) of the first one, but employs DBP[2] in place of linear dispersion compensation to compensate for deterministic nonlinear effects. After DBP (which employs

[2]The implementation of DBP requires setting some parameters which affect its numerical accuracy and complexity, namely, the number of steps along the propagation direction, the sampling rate, the number of samples in each processed block, and the number of overlapping samples — the last two parameters being required to employ the overlap-and-save algorithm to process long waveforms. Given the demodulation bandwidth, we consider the best possible implementation of DBP in terms of accuracy, without caring for complexity. Thus, all the aforementioned parameters are increased until no further accuracy improvements are observed. In particular, for what concerns sampling rate, though the signal after the demodulation filter is strictly band limited to a bandwidth $\eta B$ and completely represented by samples $\mathbf{y}$ taken at rate $\eta/T$, the nonlinear processing of DBP may require a higher sampling rate to minimize numerical errors. Thus, digital resampling is employed to achieve the required sampling rate.

oversampling), samples are filtered again by an ideal rectangular filter of bandwidth $B$, resampled at symbol rate $1/T$, and sent to the MAP detector for the AWGN channel. The third receiver is similar to the second one, but with a wider demodulation bandwidth ($\eta = 2$) to account for spectral broadening during propagation. Finally, the fourth receiver assumes no bandwidth limitation in the demodulator (in practice, the demodulation bandwidth is increased until no further improvements of the AIR are observed). For comparison, we report also the AIR on the linear ($\gamma = 0$) channel (which is obtained with any of the previous receivers and equals the Shannon capacity of the AWGN channel) and the AIR for the third receiver ($\eta = 2$), when signal-noise interaction is artificially removed by injecting all the ASE noise at the receiver (after DBP). It can be observed that, depending on the approximated model considered to optimize the detector, different AIRs are obtained. The most relevant improvement is obtained by including DBP at the receiver (more than 3 bit/symbol for both dispersion values). Also doubling the bandwidth of the demodulator provides a relevant improvement (more than 1 bit/symbol). However, an additional increase of the bandwidth can provide only a marginal improvement, since the effect of signal-noise interaction becomes dominant. In order to address the problem of signal-noise interaction, a more accurate channel model should be considered, as done in [31] or [38], for instance. However, this would be irrelevant when neglecting spectral broadening ($\eta = 1$) and only slightly beneficial (as shown in the figures) when accounting for spectral broadening up to $\eta = 2$. Similar results are obtained for both dispersion values, though all the curves are shifted toward lower powers and lower AIRs when reducing dispersion. The problem of signal-noise interaction can be well addressed at zero dispersion [4]. In this case, nonlinearity is instantaneous and affects only the phase of the optical signal, while the amplitude, being unaltered, remains a reliable information carrier at any power. As a result, it can be demonstrated that channel capacity grows unbounded at zero dispersion [4], provided that an unlimited demodulator bandwidth is considered to account for a severe spectral broadening. It would be interesting (though, perhaps, not of practical value) to understand if a similar result can be extended also to the case of non-zero dispersion.

As a multi-user scenario, we consider exactly the same modulation, bandwidth, and link configurations considered in the single-user scenario, assuming that three users ($M = 1$) are present and transmit with same power and distribution. Fig. 3 shows the AIRs obtained (for the central user) by considering three different receivers optimized according to different approximated channel models. A standard ($D = 17\,\mathrm{ps/nm/km}$) and low-dispersion ($D = 2\,\mathrm{ps/nm/km}$) fiber are considered in Fig. 3(a) and (b), respectively. As in te first receiver is designed by completely neglecting nonlinear effects (i.e., assuming that the received signal is affected only by dispersion and AWGN) and the second receiver by replacing linear dispersion compensation with DBP. On the other hand, in this scenario, we do not consider wider demodulation bandwidths. Finally, the third receiver is optimized according to the FRLP model [13, 15] and is, therefore, capable of partially mitigating the effect of inter-channel nonlinearity. In particular, a Kalman equalization strategy is employed in [39], though similar results can be obtained through a simpler detection strategy when employing multi-carrier modulation [40]. In this case, compared to the single-user scenario, the channel seen by the central user is impaired also by inter-channel nonlinearity. As a result, the AIRs obtained by the first receiver (optimized for a linear channel) are slightly lower than in the single-user scenario. This is due to the fact that deterministic intra-channel nonlinearity and inter-channel nonlinearity




(a) (b)

Figure 3: AIR for a WDM system with different demodulation and detections: (a) standard fiber ($D = 17\,\mathrm{ps/nm/km}$); (b) low-dispersion fiber ($D = 2\,\mathrm{ps/nm/km}$).

have a similar impact. On the other hand, the AIRs obtained by the second receiver (using DBP) are much lower as, in this case, inter-channel nonlinearity are dominant compared to spectral broadening and signal-noise interaction effects. In fact, only $\sim 0.5$ bit/symbol of AIR is gained by using DBP on the standard fiber and slightly less ($\sim 0.4$ bit/symbol) on the low-dispersion fiber. On the other hand, a larger gain (an additional $\sim 1$ bit/symbol on the standard fiber and $\sim 0.6$ bit/symbol on the low-dispersion fiber) is obtained by designing the receiver according to a more accurate channel model that accounts also for inter-channel nonlinearity. It is worth noting that these gains are obtained by keeping the demodulator bandwidth limited to $B$, which is the practical constraint imposed for this scenario. In fact, higher gains could be obtained by performing DBP on a wider bandwidth (including the frequency bands assigned to the other users and, therefore, compensating also for inter-channel nonlinearity). In this case, signals from many users can be seen as originated and demodulated by a single super-user (constituting a super-channel, according to a common definition adopted in the literature), and the obtained results still hold, provided that $B$ is the overall bandwidth of the super-user, while the remaining users are out of his control.

## 4. DISCUSSION AND CONCLUSION

The main message of this work is that an accurate modeling of the nonlinear optical fiber channel is fundamental to evaluate its capacity and design efficient communication systems that can approach such an ultimate theoretical limit. Even though simple models, such as the popular GN model, have proved accurate to predict the performance of conventional systems over the nonlinear optical fiber channel, they do not capture all the characteristics of the channel and are inadequate to study novel modulation and detection strategies that can deal with nonlinear effects and outperform conventional systems. In fact, by modeling nonlinear effects as AWGN, the GN model correctly predicts their impact on conventional systems, but also predicts that they cannot be compensated for or mitigated. On the other hand, we have shown that, by optimizing the receiver according to more accurate models, it is possible to outperform conventional systems and achieve higher information rates. A simple and well known example is that of DBP: even if deterministic intra-channel nonlinearity can be modeled as AWGN for what concerns its impact on conventional systems, it is easy to find out that it can be compensated for by employing a more accurate model (e.g., the SSFM for implementing DBP). In the same way, a proper model can help mitigating other nonlinear effects which, at first sight, might look like AWGN. This is the case, for instance, of inter-channel nonlinearity or signal-noise interaction.

The analysis and results presented in this work are based on some simplifying assumptions. First of all, single-polarization signals are considered and polarization effects are neglected. This is not realistic in practical systems and some care should be taken to extend the analysis to dual polarization systems (through the Manakov equation) and include polarization effects such as polarization mode dispersion. Nevertheless, the main message remains true and, in fact, some examples of DBP and mitigation of inter-channel nonlinearity and signal-noise interaction for dual polarization systems can be found also in the literature. Another important assumption is that of ideal distributed amplification. This is "less unrealistic", as this scenario can be approached at will by employing Raman amplification and/or reducing amplifier spacing. Nevertheless, attenuation plays an important role in determining the dynamics of inter-channel nonlinearity and can significantly reduce the effectiveness of mitigation strategies [15, 29]. Finally, we have considered only three users (channels) in our multi-user scenario to speed up the simulations. This is, however, only a minor issue since the presence of other (more distant) channels has only a small additional impact on the information rate, especially when considering mitigation strategies, which become more effective on more distant channels [39].

The capacity problem remains open. In this work, we have addressed the problem from a receiver perspective, trying to maximize the achievable information rate for a given input modulation by acting on the detection strategy. Though there is still margin for improvement in this sense, we believe that more significant gains (and tighter lower bounds to capacity) can be achieved by removing the assumption of Gaussian i.i.d. input symbols and optimizing the input distribution.

## ACKNOWLEDGMENT

This work was supported in part by the Italian Ministry for Education University and Research (MIUR) under the FIRB project COTONE

## REFERENCES


1. Mitra P. P. and J. B. Stark, "Nonlinear limits to the information capacity of optical fiber communications," *Nature*, Vol. 411, No. 6841, 1027–1030, 28 Jun. 2001.
2. Narimanov, E. E. and P. Mitra, "The channel capacity of a fiber optics communication system: Perturbation theory," *J. Lightwave Technol.*, Vol. 20, No. 3, 530–537, Mar. 2002.
3. Ho, K.-P. and J. M. Kahn, "Channel capacity of WDM systems using constant-intensity modulation formats," *Proc. Opt. Fiber Commun. (OFC)*, 731–733, paper ThGG85, 2002.
4. Turitsyn, K. S., 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, No. 20, paper 203901, Nov. 2003.
5. Djordjevic, I., B. Vasic, M. Ivkovic, and I. Gabitov, "Achievable information rates for high-speed long-haul optical transmission," *J. Lightwave Technol.*, Vol. 23, No. 11, 3755–3763, Nov. 2005.
6. Taghavi, M. H., G. C. Papen, and P. H. Siegel, "On the multiuser capacity of WDM in a nonlinear optical fiber: Coherent communication," *IEEE Trans. Inform. Theory*, Vol. 52, No. 11, 5008–5022, Nov. 2006.
7. Tang, J., "A comparison study of the Shannon channel capacity of various nonlinear optical fibers," *J. Lightwave Technol.*, Vol. 24, No. 5, 2070–2075, May 2006.
8. Essiambre, R.-J., G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity limits of optical fiber networks," *J. Lightwave Technol.*, Vol. 28, No. 4, 662–701, 15 Feb. 2010.
9. Ellis, A. D., Z. Jian, and D. Cotter, "Approaching the non-linear Shannon limit," *J. Lightwave Technol.*, Vol. 28, No. 4, 423–433, Feb. 2010.
10. Djordjevic, I. B., "Ultimate information capacity of fiber-optic networks," *Proc. SPIE*, 2010, Vol. 7621, No. 4, 7619–7621.
11. Bosco, G., 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, B438–B449, 16 Dec. 2011.
12. Mecozzi, A. and R.-J. Essiambre, "Nonlinear Shannon limit in pseudolinear coherent systems," *J. Lightwave Technol.*, Vol. 30, No. 12, 2011–2024, 15 Jun. 2012.
13. Secondini, M. and E. Forestieri, "Analytical fiber-optic channel model in the presence of cross-phase modulation," *IEEE Photon. Technol. Lett.*, Vol. 24, No. 22, 2016–2019, 15 Nov. 2012.
14. Serena, P., A. Bononi, and G. Colavolpe, "On the nonlinear capacity with memory of PS-QPSK and PDM-QPSK in WDM non-dispersion managed links," *Proc. Europ. Conf. Optical Commun. (ECOC)*, paper We.2.C.1, 2012.
15. Secondini, M., E. Forestieri, and G. Prati, "Achievable information rate in nonlinear WDM fiber-optic systems with arbitrary modulation formats and dispersion maps," *J. Lightwave Technol.*, Vol. 31, No. 23, 3839–3852, 2013.
16. Merhav, N., G. Kaplan, A. Lapidoth, and S. S. Shitz, "On information rates for mismatched decoders," *IEEE Trans. Inform. Theory*, Vol. 40, No. 6, 1953–1967, Nov. 1994.
17. Pfister, H., J. B. Soriaga, and P. Siegel, "On the achievable information rates of finite state ISI channels," *IEEE GLOBECOM'01*, Vol. 5, 2992–2996, 2001.
18. Arnold, D. M., H.-A. Loeliger, P. O. Vontobel, A. Kavvcic, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Trans. Inform. Theory*, Vol. 52, No. 8, 3498–3508, Aug. 2006.
19. Poggiolini, P., "The GN model of non-linear propagation in uncompensated coherent optical systems," *J. Lightwave Technol.*, Vol. 30, No. 24, 3857–3879, Dec. 2012.
20. Poggiolini, P., G. Bosco, A. Carena, V. Curri, Y. Jiang, and F. Forghieri, "The GN-model of fiber non-linear propagation and itsapplications," *J. Lightwave Technol.*, Vol. 32, No. 4, 694–721, Feb. 2014.
21. Dar, R., M. Feder, A. Mecozzi, and M. Shtaif, "Properties of nonlinear noise in long, dispersion-uncompensated fiber links," *Optics Express*, Vol. 21, No. 22, 25 685–25 699, Nov. 2013.
22. Agrell, E., A. Alvarado, G. Durisi, and M. Karlsson, "Capacity of a nonlinear optical channel with finite memory," *J. Lightwave Technol.*, Vol. 32, No. 16, 2862–2876, Aug. 2014.
23. Taha, T. R. and M. J. Ablowitz, "Analytical and numerical aspects of certain nonlinear evolution equation, II, numerical, nonlinear Schroedinger equation," *J. Computat. Phys.*, Vol. 5, 203–230, 1984.

24. Ip, E. and J. Kahn, "Compensation of dispersion and nonlinear impairments using digital backpropagation," *J. Lightwave Technol.*, Vol. 26, No. 20, 3416–3425, 2008.
25. Peddanarappagari, K. V. and M. Brandt-Pearce, "Volterra series transfer function of single-mode fibers," *J. Lightwave Technol.*, Vol. 15, 2232–2241, 1997.
26. Vannucci, A., P. Serena, and A. Bononi, "The RP method: a new tool for the iterative solution of the nonlinear Schröinger equation," *J. Lightwave Technol.*, Vol. 20, No. 7, 1102–1112, Jul. 2002.
27. Song, H. and M. Brandt-Pearce, "A discrete-time polynomial model of single channel long-haul fiber-optic communication systems," *Proc. IEEE Int. Conf. Commun. (ICC)*, 1–6, 2011.
28. Secondini, M., 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, 3358–3369, 15 Aug. 2009.
29. Dar, R., M. Feder, A. Mecozzi, and M. Shtaif, "Time varying ISI model for nonlinear interference noise," *Proc. Opt. Fiber Commun. (OFC)*, Mar. 2014.
30. Zakharov, V. E. and A. B. Shabat, "Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media," *Sov. Phys.-JETP*, Vol. 34, 62–69, 1972.
31. Marsella, D., M. Secondini, and E. Forestieri, "Maximum likelihood sequence detection for mitigating nonlinear effects," *J. Lightwave Technol.*, Vol. 32, No. 5, 908–916, Mar. 2014.
32. Wahls, S. and H. V. Poor, "Introducing the fast nonlinear Fourier transform," *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 5780–5784, May 2013.
33. Wahls, S. and H. V. Poor, "Inverse nonlinear Fourier transforms via interpolation: The Ablowitz-Ladik case," *Proc. Int. Symp. Math Theory Networks Systems (MTNS)*, 1848–1855, Jul. 2014.
34. Secondini, M., D. Marsella, and E. Forestieri, "Enhanced split-step Fourier method for digital backpropagation," *Proc. Europ. Conf. Optical Commun. (ECOC)*, paper We.3.3.5., 2014
35. Secondini, M., S. Rommel, F. Fresi, E. Forestieri, G. Meloni, and L. Potì, "Coherent 100G nonlinear compensation with single-step digital backpropagation," *Proc. Opt. Networks Design Model. (ONDM)*, 2015.
36. Cover, T. M. and J. A. Thomas, *Elements of Information Theory*, Wiley, New York, 1991.
37. Agrell, E. and M. Karlsson, "WDM channel capacity and its dependence on multichannel adaptation models," *Proc. Opt. Fiber Commun. (OFC)*, paper OTu3B.4, 2013.
38. Irukulapati, N., H. Wymeersch, P. Johannisson, and E. Agrell, "Stochastic digital backpropagation," *IEEE Trans. Commun.*, Vol. 62, No. 11, 3956–3968, Nov. 2014.
39. Secondini, M. and E. Forestieri, "On XPM mitigation in WDM fiber-optic systems," *IEEE Photon. Technol. Lett.*, Vol. 26, No. 22, 2252–2255, 2014.
40. Marsella, D., M. Secondini, E. Agrell, and E. Forestieri, "A simple strategy for mitigating XPM in nonlinear WDM optical systems," *Proc. Opt. Fiber Commun. Conf. (OFC)*, paper Th4D–3, 2015.