CHALMERS

Chalmers Publication Library

nformation-theory-friendly models for fiber-optic channels: A primer

This document has been downloaded from Chalmers Publication Library (CPL). It is the author´s version of a work that was accepted for publication in:
IEEE Information Theory Workshop (ITW), Jerusalem, Israel [Invited]

Citation for the published paper:
Agrell, E. ; Durisi, G. ; Johannisson, P. (2015) "nformation-theory-friendly models for fiber-optic channels: A primer". IEEE Information Theory Workshop (ITW), Jerusalem, Israel [Invited]

Downloaded from: http://publications.lib.chalmers.se/publication/220564

Notice: Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source. Please note that access to the published version might require a subscription.

Chalmers Publication Library (CPL) offers the possibility of retrieving research publications produced at Chalmers University of Technology. It covers all types of publications: articles, dissertations, licentiate theses, masters theses, conference papers, reports etc. Since 2006 it is the official tool for Chalmers official publication statistics. To ensure that Chalmers research results are disseminated as widely as possible, an Open Access Policy has been adopted. The CPL service is administrated and maintained by Chalmers Library.

(article starts on next page)

Information-theory-friendly models for fiber-optic channels: A primer

Erik Agrell, Giuseppe Durisi, and Pontus Johannisson

***Abstract*—There exists a rich flora of channel models for optical fiber channels, which differ not only in the types of transmission scenario they describe but also in the type of analysis they support. In this tutorial paper, we review several channel models used in optical communications, and discuss their suitability for information-theoretic analyses. Key issues are how nonlinearity, channel memory, and multiuser interference are modeled.**

I. Introduction

For finite-bandwidth linear additive white-Gaussian-noise channels, it is well known that matched filtering followed by sampling at symbol rate is optimal. This operation, together with the use of the sampling theorem, discretizes the channel input–output (I/O) relation, thus paving the way for information-theoretic analyses (see, e.g., [1, Ch. 8]), which are typically cumbersome to carry out in continuous-time.

For nonlinear channels such as the optical fiber channel we shall focus on in this paper, however, discretizing the channel I/O relation is less straightforward. Indeed, matched filtering followed by symbol-rate sampling is not optimal, and information-theoretic analyses based on such a receiver structure provide only capacity lower bounds.

An additional challenge in the optical fiber case is that even though the physics of lightwave propagation in an optical fiber is well understood [2], the resulting continuous-time channel output is given as an *implicit* function (a differential equation) of the channel input. Discrete-time channel models for which the channel output is an *explicit* function of the channel input have been proposed in the literature, but none of them is applicable across the whole range of channel and transmission conditions. The available models can be classified depending on which physical phenomena are assumed to dominate [3, Fig. 27] and on which assumptions and approximations are made in the derivations. There is often a trade-off between the physical relevance of such models and their information-theoretic usefulness.

In this tutorial paper, we survey the main types of channel models that have been proposed and comment on their underlying assumptions and their type of I/O relation, with focus on their use in information theory.

II. Signal propagation in optical fibers

All continuous-time modeling of electromagnetic phenomena originates from Maxwell's equations. Under the slowly varying envelope approximation, i.e., under the assumption that the modulation bandwidth is small compared to the carrier frequency of light, simplified evolution equations for the optical signal can be obtained. Specifically, the impact of chromatic dispersion and the Kerr nonlinearity are described by the nonlinear Schrödinger equation

$$j\frac{\partial U}{\partial z} = \frac{\beta_2}{2}\frac{\partial^2 U}{\partial t^2} - \gamma |U|^2 U, \tag{1}$$

where $U = U(t,z)$ is the complex envelope of the optical signal as a function of time t and propagation distance z, β_2 is the group-velocity dispersion parameter, and γ is the nonlinear parameter. Note that t denotes time in a reference frame that moves with the group velocity [2, (2.3.40)].

However, there are additional physical effects that must be accounted for in order to obtain an accurate channel model. Coherent transmission systems typically use polarization multiplexing to perform modulation in a four-dimensional space spanned by the two orthogonal polarizations. This is described by $\boldsymbol{U} = \boldsymbol{U}(t,z) = (U_{\mathrm{x}}(t,z), U_{\mathrm{y}}(t,z))^{\mathrm{T}}$. Averaging over the rapidly and randomly changing polarization state, one obtains the Manakov equation. Due to losses, the power of the optical signal decreases exponentially with distance. This loss is typically compensated for by the periodic insertion of erbium-doped fiber amplifiers (EDFAs), which provide lumped gain and also add noise. Alternatively, one can use Raman amplification, which introduces distributed gain and adds noise in the propagation fiber itself. These effects are included in the generalized Manakov equation [4, Sec. 3]

$$\frac{\partial \boldsymbol{U}}{\partial z} + j\frac{\beta_2}{2}\frac{\partial^2 \boldsymbol{U}}{\partial t^2} - j\frac{8\gamma}{9}(\boldsymbol{U}^{\mathrm{H}}\boldsymbol{U})\boldsymbol{U} + \frac{\alpha - g}{2}\boldsymbol{U} = \boldsymbol{Z}, \tag{2}$$

where α is the power attenuation, $g = g(z)$ is the power gain, and $\boldsymbol{Z}(t,z)$ describes addition of complex optical amplifier noise. With this formulation, one can capture the z-dependence of both Raman and EDFA amplification, in the latter case by allowing g and $\boldsymbol{Z}$ to contain Dirac δ-functions. Other channel effects, e.g., higher order dispersion and polarization-mode dispersion, are neglected. Equation (2) can be solved numerically with the split-step Fourier method [2]. Explicit solutions exist only in special cases, such as when $\gamma = 0$ (linear channel) or $\beta_2 = 0$ (memoryless channel). In the latter case, nonlinear signal–noise interaction gives rise to *nonlinear phase noise* (NLPN) [5], [6].

III. Information-theory-friendly channel models

The channel models reviewed in Sec. II are difficult to analyze using information-theory tools, because i) they are formulated in continuous time, whereas information-theoretic analyses are easier to perform in discrete time, and ii) because the relation between the signal input $\boldsymbol{U}(t,0)$ and output $\boldsymbol{U}(t,L)$, where L is the total length of the transmission fiber(s), is given implicitly as a differential equation.

For *linear* waveform channels, a common approach to deal with their continuous-time nature is to project the input and the output signals on the singular functions of the channel operator [1], [7]. This action *diagonalizes* the continuous-time channel, i.e., it transforms it into countably many scalar noninteracting I/O relations. This is similar to the diagonalization of circulant matrices by means of the discrete Fourier transform, which is exploited in orthogonal frequency-division multiplexing systems operating over frequency-selective channels [8].

For *nonlinear* dispersive waveform channels of the type described by (1), channel diagonalization can be achieved by means of the inverse scattering transform, also known as the nonlinear Fourier transform [9]. A drawback of this approach is that it is not applicable without approximation to systems that include, e.g., gain/loss and distributed addition of noise.

A different approach to obtain discrete-time models from continuous-time models is to project input and output signals on a conveniently chosen complete orthonormal set for input and output spaces (e.g., sinc functions). Although this approach does not directly exploit the structure of the channel operator, and results in a channel I/O relation that is not in diagonal form, it is a reasonable choice in systems with weak nonlinearity. It is the dominant approach in practice and the one we shall focus on in this paper.

Once both input and output signals are discretized, an "information-theory-friendly" channel is specified by providing a sequence of finite-dimensional conditional output distributions $f_{Y^n|X^n}$, $n = 1, 2, \dots$, where $X^n = [X_1, \dots, X_n]$ contains the first n input symbols and $Y^n = [Y_1, \dots, Y_n]$ the corresponding output symbols [10]. A symbol may in this context be real or complex, scalar or vectorial. The distribution $f_{Y^n|X^n}$ may not depend on the *statistics* of X^n, which will be exemplified in Sec. IV-B and V.

For every channel defined in this way, the capacity C, i.e., the highest rate for which a sequence of codes with vanishing error probability can be found, satisfies under some technical conditions[1] [10]

$$C = \liminf_{n\to\infty} \sup \frac{1}{n} I(X^n; Y^n). \tag{3}$$

This result is usually referred to as the channel coding theorem. Here, $I(\cdot;\cdot)$ denotes the mutual information [11, Ch. 2], and the supremum is over all probability distributions for X^n that satisfy the input power constraint P

$$\frac{1}{n}\mathbb{E}\Big\{\sum_{k=1}^{n} |X_k|^2\Big\} \le P. \tag{4}$$

In the optical communications literature, the inequality in (4) is often replaced by an equality. Since the capacity C can be shown to be nondecreasing with P, the two constraints are equivalent [12].

[1] Specifically, (3) is achievable if the channel is *information stable* [10].

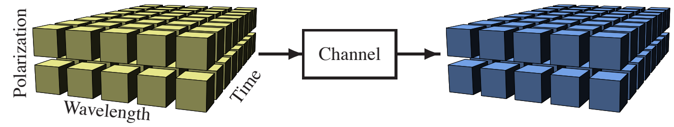

Fig. 1. The I/O space. The cuboids represent symbols $X_{k,i,\ell}$ and $Y_{k,i,\ell}$, where the three indices represent polarization, wavelength, and time.

IV. Classes of Optical Channel Models

As discussed in Sec. III, information-theoretic analyses rely to a large extent on the availability of an explicit discrete-time channel model. The input symbols lie (in the most general case) in a three-dimensional space, spanned by polarization, wavelength, and time, as shown in Fig. 1. In most models of practical interest, however, a reduced subspace is considered, for example by studying single-polarization, single-wavelength, and/or memoryless channel models.

In this section, we survey the main classes of discrete-time channel models that have been proposed for fiber-optical transmission. These classes, which are summarized in Table I, are presented each in one subsection below, except the continuous-time models, which were reviewed in Sec. II.

A. Perturbative models with deterministic nonlinearity

Since the nonlinear signal distortion in fiber-optic transmission systems is typically weak, one can linearize (2) through a perturbation analysis [30]. An equivalent approach [31] is to use a Volterra series description [32]. The basic assumption is that the approximate solution can be written as $\boldsymbol{U} \approx \boldsymbol{U}_0 + \Delta\boldsymbol{U}$, where $\boldsymbol{U}_0$ is the linear solution with $\gamma = 0$ and $\Delta\boldsymbol{U}_0$ is a small perturbation. Signal–noise interaction is neglected and noise is added at the receiver. A discrete-time expression can be obtained by matched filtering followed by sampling [33, Sec. IV], although the nonlinear signal distortion makes the filter no longer matched to the pulse. One typically assumes also that (linear) intersymbol interference can be ignored, due to the use of root-Nyquist pulses or linear equalization. Under these assumptions, the perturbation analysis gives, for a single-polarization transmission at a single wavelength, the following discrete-time I/O relation:

$$Y_k = X_k + \Delta X_k + N_k. \tag{5}$$

Here, X_k is the complex transmitted symbol, ΔX_k is the nonlinear distortion, and N_k is the optical amplifier noise.

The distortion ΔX_k is in general a function of all n transmitted symbols. The nonlinear Kerr effect is cubic and the n transmitted symbols interact in n^3 different combinations of symbol triplets, of which n^2 affect each received symbol. The distortion can be expressed as the double sum

$$\Delta X_k = \sum_{i,\ell} A_{i,\ell} X_{k+i} X_{k+\ell} X^*_{k+i+\ell}, \tag{6}$$

where the summation is over all transmitted symbols and the coefficients $A_{i,\ell}$ depend on the system parameters, e.g., the signal pulse shape. The triple products in (6) are *intrachannel four-wave mixing* (IFWM) in the general case and *intrachannel*

TABLE I

NONLINEAR MODELS FOR FIBER-OPTIC COMMUNICATION CHANNELS, CLASSIFIED ACCORDING TO THEIR FUNDAMENTAL ASSUMPTIONS (HORIZONTAL CLASSES) AND THE DIMENSIONS OF THEIR I/O SPACE (COLUMNS POLARIZATION, WAVELENGTH, AND TIME). A DESIGNATION AS "IMPLICIT" MEANS THAT THE EFFECT IS PHYSICALLY CONSIDERED BUT NOT MODELED AS A DIMENSION IN THE I/O SPACE.

Model	Polarization	Wavelength	Time	References
Differential equations				
Nonlinear Schrödinger	single	N/A	continuous time	[2, (2.3.41)]
Manakov	dual	N/A	continuous time	[4, Sec. 3]
Perturbative models with deterministic nonlinearity				
IXPM, IFWM	single	single	multiple	[13, (17)]
Cross-pol. IXPM and IFWM	dual	single	multiple	[14, (10)]
FWM	single	three	single	[15, (1)]
XPM, FWM	single	multiple	single	[16, (8)], [17, (8)]
IXPM, IFWM, XPM, FWM	single	two	multiple	[18, (34)–(35)], [19, (5)], [20]
IXPM, IFWM, XPM, FWM	single	multiple	multiple	[16, (48)], [21, (15)]
Perturbative models with random nonlinearity				
Single-channel GN	single	single	implicit	[22]
Dual-polarization GN	dual	single	implicit	[23]
Multichannel GN	single	implicit	implicit	[24]–[26]
Finite-memory GN	single	single	multiple	[27, (8)]
Phase and Gaussian noise	single	implicit	multiple	[28, (1)]
Memoryless phase-noise models				
Single-polarization NLPN	single	single	single	[29, (70)], [5, (11)]
Dual-polarization NLPN	dual	single	single	[6, (23)]

cross-phase modulation (IXPM) in the special cases $i = 0$ or $\ell = 0$. Models based on (6) were proposed in [13, Eq. (17)] and for dual polarization in [14, Eq. (10)].

The perturbation analysis can also be applied to multiple-wavelength transmission. If the channel memory is neglected, a model similar to (5)–(6) can be formulated, where the indices refer to wavelengths instead of time [15, Eq. (1)], [16, Eq. (8)], [17, Eq. (8)]. The nonlinear distortion components are, in this case, *four-wave mixing* (FWM) and *cross-phase modulation* (XPM). The generalization to interaction between symbols that are separated in both time and wavelength was considered in [16], [18]–[21]; see Table I for details.

In most fiber-optic communication systems, the bulk of the nonlinear distortion is caused by phase shifts due to XPM or IXPM. In a perturbation analysis, all phase shifts are described by their first-order linear approximation, resulting in ΔX_k being in quadrature to X_k and proportional to the magnitude of X_k. This relation can be expressed in terms of the (I)XPM phase shifts $\Delta\theta_k$ as

$$Y_k = X_k e^{j\Delta\theta_k} + N_k, \tag{7}$$

which is equivalent to (5) up to a first-order linearization. This and similar phase-shift models were studied in [17]–[20].

The channel capacity formula (3) applies to the models (5), (7), and their variations, provided that the double sum in (6) is truncated after a finite number of terms to avoid inter-packet interference. This can be done with an arbitrarily small error, since the coefficients $A_{i,\ell}$ tend to 0 as $|i|$ or $|\ell|$ approach infinity. Capacity lower bounds were obtained in [20] using a mismatched decoding approach. This class of models was employed to formulate and solve problems in multiuser information theory in [16]–[18], as summarized in Sec. V.

B. Perturbative models with random nonlinearity

The perturbation analysis in Sec. IV-A yields a significant simplification over the original model (2). However, the resulting discrete-time I/O relations are still involved. A further drastic simplification is obtained by assuming that the nonlinear distortion arises from the interaction between independent stochastic processes rather than data-dependent signals. This yields the so called *Gaussian noise* (GN) model [24]. Specifically, under some additional assumptions on these stochastic processes (e.g., [25, Eq. (3)]), one obtains an explicit solution through perturbation analysis [26]. A discrete-time model can be obtained by matched filtering and sampling. In the single-polarization, single-wavelength case [22], these steps yield an I/O relation similar to (5), where, however,

$$\Delta X_k = W_k \sqrt{\eta P_{\mathrm{i}}^3}. \tag{8}$$

Here, W_k is a complex, zero-mean, circular Gaussian random variable with variance $\mathbb{E}[|W_k|^2] = 1$, the factor η depends on system parameters including the signal pulse shape, and P_{i} quantifies the power of the interacting stochastic processes. The cubic dependency on power corresponds to the triple products in (6). The GN models can be obtained also for dual-polarization [23] and/or multi-wavelength case [24]–[26], and similar expressions apply.

Information-theoretically, the GN model (8) is fundamentally different from the perturbation model (5)–(6). In the GN model, the nonlinear distortion ΔX_k is random rather than deterministic. Furthermore, the channel becomes memoryless. Since, according to the GN model, ΔX_k is Gaussian and independent of X_k, mutual information is maximized using Gaussian inputs, and the capacity of this channel is given by the familiar $\log(1+\mathrm{snr})$ formula, where snr is the signal-to-noise ratio.

A crucial question, often neglected in the literature, is how to define P_{i} mathematically, and how to relate it to the input power constraint P in (4). Physically, these two quantities are similar: both quantify the average input power. However, care must be exercised in using a channel model that depends on the transmitted sequence only through its power as in (8) with $P_{\mathrm{i}} = P$. Indeed, it is unclear how to cast this model within the framework presented in Section III, where the channel should not depend on statistical properties of the input sequence, but on the input sequence itself.

From an information-theory perspective, a more satisfactory way to interpret the GN model is to view the capacity predicted by this model as a lower bound on the capacity of the channel (5) when one uses Gaussian inputs and distortion is treated as noise. Indeed, the well-known Gaussian saddle point theorem (see, e.g., [11, p. 298]) implies that the capacity of (5)–(6) can be lower-bounded as follows: $C \geq I(X_{\mathrm{G}}; Y) \geq I(X_{\mathrm{G}}; Y_{\mathrm{G}})$, where X_{G} is a complex Gaussian random variable that satisfies the average power constraint, Y is the channel output when the input is X_{G}, and Y_G is the channel output when the input is X_{G} and when the distortion ΔX_k in (5) is replaced by a Gaussian random variable having the same variance as ΔX_k. The results in [24], [33] and other papers that report the channel capacity of the GN model (8) should thus be interpreted as capacity lower bounds.

In an attempt to make the GN model amenable to information-theoretic analysis, the authors in [27] proposed to replace P_{i} in (8) by a time-varying power P_k defined as

$$P_k = \frac{1}{2M+1} \sum_{i=k-M}^{k+M} |X_i|^2, \tag{9}$$

which obviously depends on the input sequence, not on its statistics as in (8). Lower bounds on the capacity of this channel were derived in [27]. Another modification of the GN model was proposed in [28], where it was observed that part of the distortion ΔX_k manifests itself as slowly varying phase noise. A block-memoryless model was proposed in which $\Delta X_k = X_k e^{j\theta} + N_k^{\mathrm{NL}}$, where θ is the phase noise, which is constant throughout the block, and N_k^{NL} is an independent Gaussian variable, whose variance implicitly depends on the input sequence. The capacity of this channel was lower-bounded in [28] using Gaussian inputs.

C. Memoryless phase-noise models

When dispersion can be neglected, i.e., β_2 can be set to zero, the main remaining channel effect is the Kerr nonlinearity. Without optical amplifier noise, an explicit analytic solution to (2) can be found [6, Eq. (2)]. The introduction of optical amplifier noise leads to signal–noise interaction, NLPN, which changes the shape of the probability density function (pdf) of the received signal. This modified pdf has been found analytically and is in the single-polarization case [34, Sec. 5.1]

$$\begin{aligned} & f_{R,\Theta|R_0,\Theta_0}(r,\theta|r_0,\theta_0) \\ & \quad = \frac{f_{R|R_0}(r|r_0)}{2\pi} + \frac{1}{\pi}\sum_{m=1}^{\infty} \Re\left\{C_m(r,r_0)e^{jm(\theta-\theta_0)}\right\} \end{aligned} \tag{10}$$

where R_0, Θ_0 and R, Θ are the polar coordinates of the channel input $X = R_0 e^{j\Theta_0}$ and output $Y = Re^{j\Theta}$, resp., $f_{R|R_0}$ is a Rice distribution giving the conditional marginal distribution of R, and $C_m(r, r_0)$ are Fourier coefficients that depend on the system configuration, including fiber parameters and amplification scheme. This model is applicable for single-channel systems at low to moderate symbol frequencies, even in the strongly nonlinear regime, but is unfortunately not accurate in wavelength-division multiplexing systems. A dual-polarization extension was provided in [6]. The channel capacity of the memoryless channel (10) was lower-bounded in [5] and shown to grow unbounded with increasing P.

V. Multiuser information theory

When more than one user is present in the system, the analysis of the largest achievable rates at which communication is possible is made complicated by the nonlinear interaction between the signals associated to different users. Information-theoretically, the notion of capacity is replaced by that of *capacity region*, which is the set of rate-tuples (one per user) for which reliable communication can occur. Calculating the capacity region is an untractable problem in the presence of nonlinear distortion. Even for simple linear channel models, the capacity region of channels subject to unwanted multiuser interference is unknown [35, Ch. 6].

In long-haul fiber-optical networks, multiple users typically transmit on the same fiber using separate wavelengths. Bounds on the capacity region for such wavelength-division multiplexing systems were obtained in [12], [16]. In [17], [18], it was shown that improved rates can be achieved in a two-user system if both users apply *interference focusing* to reduce the interference caused by one user on the other user's transmission. These studies, which are all based on selected perturbation models with deterministic nonlinearity (Sec. IV-A), are carried out under somewhat idealized conditions and may have limited practical impact. In real optical networks, users may be geographically distributed (no joint processing possible) and may transmit at different symbol rates; even when they transmit at the same symbol rate, their symbol clocks are typically not synchronous.

Instead of studying the capacity region, a common approach in the literature about optical channel capacity [3], [15], [24], [33] is to consider the point-to-point capacity of one user in the network, while the transmissions by the other users are considered to be interference and modeled as stochastic processes. The rationale, apart from making the problem more tractable, is that many practical transmission scenarios are symmetric in the sense that signals on different wavelengths are modulated similarly and experience similar interference statistics, so that a characterization of a single user would give a good picture of the overall system performance. However, similarly to the GN model, care must be exercised if the resulting channel model depends on the statistical properties of the input, via a relation between input and interference. The capacity results obtained by such analysis should be seen as lower (inner) bounds on the capacity region of the underlying channel, rather than ultimate performance limits.

VI. Conclusions

The discrete-time channel models developed so far for fiber-optic channels are obtained under specific assumptions and approximations on the underlying continuous time I/O relation. This limits their validity to a specific range of channel and transmission conditions, a fact that must be taken into account when performing information-theoretic analyses. Furthermore, most discrete-time channel model are obtained by applying matched filtering and sampling. Since this procedure is not optimal for the case of nonlinear channels, analyses based on these discrete-time models can only provide capacity lower bounds, and are unable to characterize the ultimate performance limits of communication over fibers. "Information-theory friendly" models able to provide a satisfactory characterization of this ultimate limit need to be developed.

Although simplicity and tractability are crucial properties of every channel model, an oversimplified channel model may yield erroneous conclusions when analyzed with information-theory tools. One such example is the GN model (8), whose capacity is just a lower bound (obtained by using Gaussian input and treating distortion as noise) on the capacity of the more refined discrete-time channel (5)–(6). An approach where the channel is characterized as conditional distribution as specified in Sec. III and then information-theory inequalities are used to characterize the rates achievable with suboptimal (but easier to implement) strategies seems to be preferable.

References

[1] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY: Wiley, 1968.

[2] G. P. Agrawal, *Nonlinear Fiber Optics*, 3rd ed. Academic Press, 2001.

[3] R.-J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity limits of optical fiber networks," *J. Lightw. Technol.*, vol. 28, no. 4, pp. 662–701, Feb. 2010.

[4] C. R. Menyuk, "Application of multiple-length-scale methods to the study of optical fiber transmission," *Journal of Engineering Mathematics*, vol. 36, pp. 113–136, 1999.

[5] 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, no. 20, pp. 203 901-1–4, Nov. 2003.

[6] L. Beygi, E. Agrell, M. Karlsson, and P. Johannisson, "Signal statistics in fiber-optical channels with polarization multiplexing and self-phase modulation," *J. Lightw. Technol.*, vol. 29, no. 16, pp. 2379–2386, Aug. 2011.

[7] A. D. Wyner, "The capacity of the band-limited Gaussian channel," *Bell Syst. Tech. J.*, vol. 45, no. 3, pp. 359–395, Mar. 1966.

[8] D. N. C. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge University Press, 2005.

[9] M. I. Yousefi and F. R. Kschischang, "Information transmission using the nonlinear Fourier transform, part I: Mathematical tools," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 4312–4328, Jul. 2014.

[10] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1147–1157, July 1994.

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

[12] E. Agrell, "Conditions for a monotonic channel capacity," *IEEE Trans. Commun.*, 2015.

[13] H. Song and M. Brandt-Pearce, "A discrete-time polynomial model of single channel long-haul fiber-optic communication systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kyoto, Japan, June 2011.

[14] 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, Sept. 2011.

[15] E. Agrell and M. Karlsson, "WDM channel capacity and its dependence on multichannel adaptation models," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, Anaheim, CA, Mar. 2013, p. OTu3B.4.

[16] M. H. Taghavi, G. C. Papen, and P. H. Siegel, "On the multiuser capacity of WDM in a nonlinear optical fiber: Coherent communication," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5008–5022, Nov. 2006.

[17] H. Ghozlan and G. Kramer, "Interference focusing for mitigating cross-phase modulation in a simplified optical fiber model," in *Proc. IEEE Int. Symp. Inf. Theory*, Austin, TX, June 2010, pp. 2033–2037.

[18] ——, "Interference focusing for simplified optical fiber models with dispersion," in *Proc. IEEE Int. Symp. Inf. Theory*, Saint Petersburg, Russia, July–Aug. 2011, pp. 376–379.

[19] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "Properties of nonlinear noise in long, dispersion-uncompensated fiber links," *Opt. Exp.*, vol. 21, no. 22, pp. 25 685–25 699, Oct. 2013.

[20] M. Secondini, E. Forestieri, and G. Prati, "Achievable information rate in nonlinear WDM fiber-optic systems with arbitrary modulation formats and dispersion maps," *J. Lightw. Technol.*, vol. 31, no. 23, pp. 3839–3852, Dec. 2013.

[21] H. Song and M. Brandt-Pearce, "A 2-D discrete-time model of physical impairments in wavelength-division multiplexing systems," *J. Lightw. Technol.*, vol. 30, no. 5, pp. 713–726, Mar. 2012.

[22] A. Bononi, P. Serena, N. Rossi, E. Grellier, and F. Vacondio, "Modeling nonlinearity in coherent transmissions with dominant intrachannel-four-wave-mixing," *Opt. Exp.*, vol. 20, no. 7, pp. 7777–7791, Mar. 2012.

[23] L. Beygi, E. Agrell, P. Johannisson, M. Karlsson, and H. Wymeersch, "A discrete-time model for uncompensated single-channel fiber-optical links," *IEEE Trans. Commun.*, vol. 60, no. 11, pp. 3440–3450, Nov. 2012.

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

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

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

[27] E. Agrell, A. Alvarado, G. Durisi, and M. Karlsson, "Capacity of a nonlinear optical channel with finite memory," *J. Lightw. Technol.*, vol. 32, no. 16, pp. 2862–2876, Aug. 2014.

[28] 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.

[29] A. Mecozzi, "Limits to long-haul coherent transmission set by the Kerr nonlinearity and noise of the in-line amplifiers," *J. Lightw. Technol.*, vol. 12, no. 11, pp. 1993–2000, Nov. 1994.

[30] A. Mecozzi, C. B. 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.

[31] A. Vannucci, P. Serena, and A. Bononi, "The RP method: A new tool for the iterative solution of the nonlinear Schrödinger equation," *J. Lightw. Technol.*, vol. 20, no. 7, pp. 1102–1112, Jul. 2002.

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

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

[34] K.-P. Ho, *Phase-Modulated Optical Communication Systems*. New York, NY: Springer, 2005.

[35] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, UK: Cambridge University Press, 2011.