



# CHALMERS

## Chalmers Publication Library

### **Capacity of a nonlinear optical channel with finite memory**

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:

**Journal of Lightwave Technology (ISSN: 0733-8724)**

Citation for the published paper:

Agrell, E. ; Alvarado, A. ; Durisi, G. (2014) "Capacity of a nonlinear optical channel with finite memory". *Journal of Lightwave Technology*, vol. 32(16), pp. 2862-2876.

Downloaded from: <http://publications.lib.chalmers.se/publication/210422>

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)

# Capacity of a Nonlinear Optical Channel with Finite Memory

Erik Agrell, Alex Alvarado, Giuseppe Durisi, and Magnus Karlsson

(Invited Paper)

**Abstract**—The channel capacity of a nonlinear, dispersive fiber-optic link is revisited. To this end, the popular Gaussian noise (GN) model is extended with a parameter to account for the finite memory of realistic fiber channels. This finite-memory model is harder to analyze mathematically but, in contrast to previous models, it is valid also for nonstationary or heavy-tailed input signals. For uncoded transmission and standard modulation formats, the new model gives the same results as the regular GN model when the memory of the channel is about 10 symbols or more. These results confirm previous results that the GN model is accurate for uncoded transmission. However, when coding is considered, the results obtained using the finite-memory model are very different from those obtained by previous models, even when the channel memory is large. In particular, the peaky behavior of the channel capacity, which has been reported for numerous nonlinear channel models, appears to be an artifact of applying models derived for independent input in a coded (i.e., dependent) scenario.

**Index Terms**—Channel capacity, channel model, fiber-optic communications, Gaussian noise model, nonlinear distortion.

## I. INTRODUCTION

The introduction of coherent optical receivers has brought significant advantages in fiber optical communications, e.g., enabling efficient polarization demultiplexing, higher-order modulation formats, increased sensitivity, and electrical mitigation of transmission impairments [1], [2]. Even if the linear transmission impairments (such as chromatic and polarization-mode dispersion) can be dealt with electronically, the Kerr nonlinearity in the fiber remains a significant obstacle. Since the nonlinearity causes signal distortions at high signaling powers, arbitrarily high signal-to-noise ratios are inaccessible, which limits transmission over long distances and high spectral efficiencies. This is sometimes referred to as the “nonlinear Shannon limit” [3], [4].

Research supported by the Swedish Research Council (VR) under grant no. 2012-5280, the Swedish Foundation for Strategic Research (SSF) under grant no. RE07-0026, and the European Community’s Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 271986. This paper was presented in part at the 2013 European Conference on Optical Communication.

E. Agrell and G. Durisi are with the Department of Signals and Systems, Chalmers University of Technology, SE-41296 Gothenburg, Sweden (email: {agrell,durisi}@chalmers.se).

M. Karlsson is with the Department of Microtechnology and Nanoscience, Chalmers University of Technology, SE-41296 Gothenburg, Sweden (email: magnus.karlsson@chalmers.se).

A. Alvarado was with the Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, UK. He is now with the Optical Networks Group, Department of Electronic and Electrical Engineering, University College London, London WC1E 7JE, UK (email: alex.alvarado@ieee.org).

Copyright © 2014 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

For systems with large accumulated dispersion and weak nonlinearity, the joint effect of chromatic dispersion and the Kerr effect is similar to that of additive Gaussian noise. This was pointed out already by Splett [5] and Tang [6]. The emergence of this Gaussian noise is prevalent in links that have no inline dispersion compensation, such as today’s coherent links, where the dispersion compensation takes place electronically in the receiver signal processing. This Gaussian noise approximation has been recently rediscovered and applied to today’s coherent links in a series of papers by Poggiolini *et al.* (see [7], [8] and references therein) and other groups [9]–[11]. The resulting so-called Gaussian noise model, or *GN model* for short, is valid for multi-channel (wavelength- and polarization-division multiplexed) signals. It has also been shown to work for single-channel and single-polarization transmission if the dispersive decorrelation is large enough [9], [12].

The GN model belongs to a more general family of additive Gaussian noise channels, which we study in this paper. In such channels, for a given transmitted complex symbol  $X_k$ , the (complex) single-channel output at each discrete-time  $k \in \mathbb{Z}$  is modeled as

$$Y_k = X_k + Z_k, \quad (1)$$

where  $\{Z_k\}$  is a circularly symmetric, complex, white, Gaussian random sequence, independent of  $X_k$ .

A crucial assumption in the derivation of the GN model, as well as most other fiber-optical channel models, is that of independent, identically distributed (i.i.d.) inputs: the transmitted symbols are independent of each other, are drawn from the same constellation, and have the same average transmit power. Under these assumptions, the model has been experimentally verified to be very accurate [13], [14] for the most common modulation formats, such as quadrature amplitude modulation (QAM) or phase-shift keying.

In this paper, the assumption of i.i.d. inputs is relaxed. This is done by introducing a modified GN model, which we call the *finite-memory GN model*. This new model includes the memory of the channel as a parameter and differs from previous channel models in that it is valid also when the channel input statistics are time-varying, or when “heavy-tailed” constellations are used.

The performance predicted by the regular GN model (both in terms of uncoded error probability and channel capacity) is compared with the ones predicted by the finite-memory GN model. The *uncoded* performance is characterized in terms of symbol error rate (SER) and bit error rate (BER), assuming i.i.d. data. Exact analytical expressions are obtained for 16-ary QAM (16-QAM), which show that the GN model is accurate

for uncoded transmission and standard modulation formats, confirming previous results.

The main contributions of the paper are in terms of *coded* performance. Shannon, the father of information theory, proved that for a given channel, it is possible to achieve an arbitrarily small error probability, if the transmission rate in bits per symbol is small enough. A rate for which virtually error-free transmission is possible is called an *achievable rate* and the supremum over all achievable rates for a given channel, represented as a statistical relation between its input  $X$  and output  $Y$ , is defined as the *channel capacity* [15], [16, p. 195]. A capacity-approaching transmission scheme operates in general by grouping the data to be transmitted into blocks, encoding each block into a sequence of coded symbols, modulating and transmitting this sequence over the channel, and decoding the block in the receiver. This coding process introduces, by definition, dependencies among the transmitted symbols, which is the reason why channel models derived for i.i.d. inputs may be questionable for the purpose of capacity analysis.

More fundamentally, the regular GN model is not well-suited to capacity analysis, because in this model each output sample depends on the statistics of the previously transmitted input symbols (through their average power) rather than on their actual value. This yields artifacts in capacity analysis. One such artifact is the peaky behavior of the capacity of the GN model as a function of the transmit power. Indeed, through a capacity lower bound it is shown in this paper that this peaky behavior does not occur for the finite-memory GN model, even when the memory is taken to be arbitrary large.

The analysis of channel capacity for fiber-optical transmission dates back to 1993 [5], when Splett *et al.* quantified the impact of nonlinear four-wave mixing on the channel capacity. By applying Shannon's formula for the additive white Gaussian noise (AWGN) channel capacity to a channel with power-dependent noise, Splett *et al.* found that there exists an "optimal" *finite* signal-to-noise ratio that maximizes capacity. Beyond this value, capacity starts decreasing. It was however not motivated in [5] why the noise was assumed Gaussian. Using a different model for four-wave mixing, Stark [17] showed that capacity saturates, but does not decrease, at high power. In the same paper, the capacity loss due to the quantum nature of light was quantified. In 2001, Mitra and Stark [18] considered the capacity in links where cross-phase modulation dominates, proved that the capacity is lower-bounded by the capacity of a linear, Gaussian channel with the same input-output covariance matrix, and evaluated this bound via Shannon's AWGN formula. The obtained bound vanishes at high signal power. The authors claimed, without providing evidence, that the true capacity would have the same qualitative nonmonotonic behavior.

Since 2001, the interest in optical channel capacity has virtually exploded. The zero-dispersion channel was considered by Turitsyn *et al.* [19]. The joint effect of nonlinearity and dispersion was modeled by Djordjevic *et al.* [20] as a finite-state machine, which allowed the capacity to be estimated using the Bahl-Cocke-Jelinek-Raviv (BCJR) algorithm. Taghavi *et al.* [21] considered a fiber-optical multiuser system as a multiple-

access channel and characterized its capacity region. In a very detailed tutorial paper, Essiambre *et al.* [22] applied a channel model based on extensive lookup tables and obtained capacity lower bounds for a variety of scenarios. Secondini *et al.* [23] modeled the nonlinear interference (NLI) in wavelength-division multiplexing (WDM) systems as a slowly varying intersymbol interference channel, based on the continuous-time analysis in [24], and obtained lower bounds using the theory of mismatched decoding. Dar *et al.* used a blockwise constant phase noise to represent the NLI [25], [26] and recently generalized the model into an intersymbol interference channel [27].

Detailed literature reviews are provided in [28] for the early results, and in [22] for more recent results. Other capacity estimates, or lower bounds thereon, were reported for various nonlinear transmission scenarios in, e.g., [4], [29]–[35]. Most of these estimates or bounds decrease to zero as the power increases, in contrast to the results presented in this work.

This paper is organized as follows. In Sec. II, the GN model is reviewed and the finite-memory GN model is introduced. In Sec. III, the uncoded error performance of the new finite-memory model is studied. The channel capacity results, which are the main results in this paper, are presented in Sec. IV and conclusions are drawn in Sec. V. The mathematical proofs are relegated to appendices.

*Notation:* Throughout this paper, vectors are denoted by boldface letters  $\mathbf{x}$  and sets are denoted by calligraphic letters  $\mathcal{X}$ . Random variables are denoted by uppercase letters  $X$  and their (deterministic) outcomes by the same letter in lowercase  $x$ . Probability density functions (PDFs) and conditional PDFs are denoted by  $f_Y(y)$  and  $f_{Y|X}(y|x)$ , respectively, and probability mass functions (PMF) are denoted by  $P_X(x)$ . Expectations are denoted by  $\mathbb{E}[\cdot]$  and random sequences by  $\{X_k\}$ .

## II. CHANNEL MODELING: FINITE AND INFINITE MEMORY

In this section, we will begin with a high-level description of the nonlinear interference in optical dual-polarization WDM systems, highlighting the role of the channel memory, and thereafter in Sec. II-B–II-D describe in detail the channel models considered in this paper.

### A. Nonlinear Interference in Optical Channels

A coherent optical communication link converts a discrete, complex-valued electric data signal  $X_k$  to a modulated, continuous optical signal, which is transmitted through an optical fiber, received coherently, and then converted back to a discrete output sequence  $Y_k$ . The coherent link is particularly simple theoretically, in that the transmitter and receiver directly map the electric data to the optical *field*, which is a linear operation (in contrast with, e.g., direct-detection receivers), and can ideally be performed without distortions. The channel is then well described by the propagation of the (continuous) optical field in the fiber link. It should be emphasized that this assumes the coherent receiver to be ideal, with perfect synchronization and negligible phase noise. Experiments have shown [2] that commercial coherent receivers can indeed perform well enough

for the fiber propagation effects to be the main limitations. Two main *linear* propagation effects in the fiber need to be addressed: *dispersion* and *attenuation*. The attenuation effects can be overcome by periodic optical amplification, at the expense of additive Gaussian noise from the inline amplifiers. The dispersion effects are usually equalized electronically by a filter in the coherent receiver. Such a linear optical link can be well-described by an AWGN channel, the capacity of which is unbounded with the signal power.

However, the fiber Kerr-nonlinearity introduces signal distortions, and greatly complicates the transmission modeling. The nonlinear signal propagation in the fiber is described by a nonlinear partial differential equation, the *nonlinear Schrödinger equation* (NLSE), which includes dispersion, attenuation, and nonlinearity. At high power levels, the three effects can no longer be conveniently separated. However, in contemporary coherent links (distance at least 500 km and symbol rate at least 28 Gbaud), the nonlinearity is significantly weaker than the other two effects, and a perturbation approach can be successfully applied to the NLSE [5], [8]–[10]. This leads to the GN model, which will be described in Sec. II-C.

### B. Finite Memory

Even today's highly dispersive optical links have a finite memory. For example, a signal with dispersive length  $L_D = 1/(\Delta\omega^2|\beta_2|)$ , where  $\beta_2$  is the group velocity dispersion and  $\Delta\omega$  the optical bandwidth, broadens (temporally) a factor  $L/L_D$  over a fiber of length  $L$ . With typical dispersion lengths of 5–50 km, this broadening factor can correspond to hundreds to thousands of adjacent symbols, a large but *finite* number. The same will hold for interaction among WDM channels; if one interprets  $\Delta\omega$  as the channel separation,  $L/L_D$  will give an approximation on the number of symbols that two WDM channels separate due to walk-off (and hence interact with nonlinearly during transmission). The channel memory will thus be even larger in the WDM case, and increase with channel separation, but the nonlinear interaction will decrease due to the shorter  $L_D$ . Thus, the principle of a finite channel memory holds also for WDM signals. To keep notation as simple as possible, we will consider a single, scalar, wavelength channel in this paper. Extensions to dual polarizations and WDM are possible, but will involve obscuring complications such as four-dimensional constellation space [36] in the former case and behavioral models [37] in the latter. We can thus say that in an optical link a certain signal may sense the interference from  $N \approx L/L_D$  neighboring symbols, which is the physical reason for introducing a finite-memory model.

If we let the number  $N$  of interfering symbols go to infinity, an even simpler type of model is obtained. The interference is now averaged over infinitely many transmitted symbols. Assuming that an i.i.d. sequence is transmitted, this time average converges to a statistical average, which greatly simplifies the analysis. Many models suggested for dispersive optical channels so far belong to this category [4], [5], [9], [10], [12], [22], [38], [39], of which the GN model described in Sec. II-C is the most common. The finite-memory models by Secondini *et al.* [23] and Dar *et al.* [26], [27] are notable exceptions.

### C. The Regular GN Model

For coherent long-haul fiber-optical links without dispersion compensation, Splett *et al.* [5], Poggiolini *et al.* [7], and Beygi *et al.* [9] have all derived models where the NLI appears as Gaussian noise, whose statistics depend on the transmitted signal power via a cubic relationship. The models assume that the transmitted symbols  $X_k$  in time slot  $k \in \mathbb{Z}$  are i.i.d. In this model, the additive noise in (1) is given by

$$Z_k = \tilde{Z}_k \sqrt{P_{\text{ASE}} + \eta P^3}, \quad (2)$$

where  $\{\tilde{Z}_k\}$  are i.i.d. zero-mean unit-variance circularly symmetric complex Gaussian random variables,  $P_{\text{ASE}}$  and  $\eta$  are real, nonnegative constants, and  $P = \mathbb{E}[|X|^2]$  is the *average transmit power*. Therefore, the noise  $Z_k$  is distributed as  $Z_k \sim \mathcal{CN}(0, P_{\text{ASE}} + \eta P^3)$ , where  $\mathcal{CN}(0, \sigma^2)$  denotes a circularly symmetric complex Gaussian random variable with mean 0 and variance  $\sigma^2$ . The parameter  $P$ , which is a property of the transmitter, governs the behavior of the channel model. This can be intuitively understood as a long-term average of the signal power. Mathematically,

$$P = \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{i=k-N}^{k+N} |X_i|^2, \quad (3)$$

with convergence almost surely for any given  $k$ , still assuming i.i.d. symbols  $X_k$ . For this reason, we will refer to models that depend on infinitely many past and/or future symbols, via  $P$  in (3) or in some other way, as *infinite-memory* models.

The cubic relation in (2) between the transmit power and the additive noise variance  $P_{\text{ASE}} + \eta P^3$  is a consequence of the Kerr nonlinearity, and holds for both lumped and distributed amplification schemes. The constant  $P_{\text{ASE}}$  represents the total amplified spontaneous emission (ASE) noise of the optical amplifiers for the channel under study, while  $\eta$  quantifies the NLI. Several related expressions for this coefficient have been proposed. For example, for distributed amplification and WDM signaling over the length  $L$ ,

$$\eta = \frac{4\gamma^2 L}{\pi|\beta_2|B^2} \log_e(2\pi e|\beta_2|LB^2), \quad (4)$$

$$\eta = \frac{16\gamma^2 L}{27\pi|\beta_2|R_s^2} \log_e\left(\frac{2}{3}\pi^2|\beta_2|LB^2\right), \quad (5)$$

were proposed in [5] and [40], resp., where  $\gamma$  is the fiber nonlinear coefficient,  $B$  is the total WDM bandwidth, and  $R_s$  is the symbol rate. Obviously, the expressions in (4) and (5) are qualitatively similar. For dual polarization and single channel transmission over  $M$  lumped amplifier spans, the expression

$$\eta = \frac{3\gamma^2}{\alpha^2} M^{1+\epsilon} \tanh\left(\frac{\alpha}{4|\beta_2|R_s^2}\right) \quad (6)$$

was proposed in [12], and a qualitatively similar formula can be obtained from the results in [8]. Here,  $\alpha$  is the attenuation coefficient of the fiber and the coefficient  $\epsilon$  is between 0 and 1 (see [8], [12]) depending on how well the nonlinear interference decorrelates between each amplifier span. Although empirical formulas for  $\epsilon$  have been proposed [39], for simplicity we assume independent decorrelated waveforms

inducing the nonlinear interference at each amplifier, i.e.,  $\epsilon = 0$ . In this paper, we use the single-polarization version of (6), in which the coefficient 3 should be replaced by 2 [9], which yields

$$\eta = \frac{2\gamma^2 M}{\alpha^2} \tanh\left(\frac{\alpha}{4|\beta_2|R_s^2}\right). \quad (7)$$

The benefits of the GN model is that it is very accurate for uncoded transmission with traditional modulation formats,<sup>1</sup> as demonstrated in experiments and simulations [13], [14], [42], and that it is very simple to analyze. It is, however, not intended for nonstationary input sequences, i.e., sequences whose statistics vary with time, because of the i.i.d. assumption on the symbols  $X_k$ . In order to capture the behavior of a wider class of transmission schemes, the GN model can be modified to depend on a time-varying transmit power, which is the topic of the next section.

#### D. The Finite-Memory GN Model

As mentioned in Sec. I and II-C, a finite-memory model is essential in order to model the channel output corresponding to time-varying input distributions. Therefore, we refine the GN model in Sec. II-C to make it explicitly dependent on the channel memory  $N$ , in such a way that the model “converges” to the regular GN model as  $N \rightarrow \infty$ . Many such models can be formulated. In this paper, we aim for simplicity rather than accuracy.

The proposed model assumes that the input–output relation is still given by (1), but the average transmit power  $P$  in (2) is replaced by an *empirical power*, i.e., by the arithmetic average of the squared magnitude of the symbol  $X_k$  and of the  $2N$  symbols around it. Mathematically, (2) is replaced by

$$Z_k = \tilde{Z}_k \sqrt{P_{\text{ASE}} + \eta \left( \frac{1}{2N+1} \sum_{i=k-N}^{k+N} |X_i|^2 \right)^3} \quad (8)$$

for any  $k \in \mathbb{Z}$ , where  $N$  is the (one-sided) *channel memory*. We refer to (1) and (8) as the *finite-memory GN model*. Since (second-order) group velocity dispersion causes symmetric broadening with respect to the transit time of the signal, intersymbol interference from dispersion will act both backwards and forwards in terms of the symbol index. This is why both past and future inputs contribute to the noise power in (8). A somewhat related model for the additive noise in the context of data transmission in electronic circuits has been recently proposed in [43], where the memory is single-sided and the noise scales linearly with the signal power, not cubically as in (8).

Having introduced the finite-memory GN model, we now discuss some particular cases. First, the memoryless AWGN channel model can be obtained from both the GN and finite-memory GN models by setting  $\eta = 0$ . In this case, the noise variance is  $\mathbb{E}[|Z_k|^2] = P_{\text{ASE}}$  for all  $k$ . Second, let us consider the scenario where the transmitted symbols is the random process  $\{X_i\}$ . Then the empirical power  $(1/(2N +$

<sup>1</sup>The model is not valid for exotic modulation formats such as satellite constellations [41].

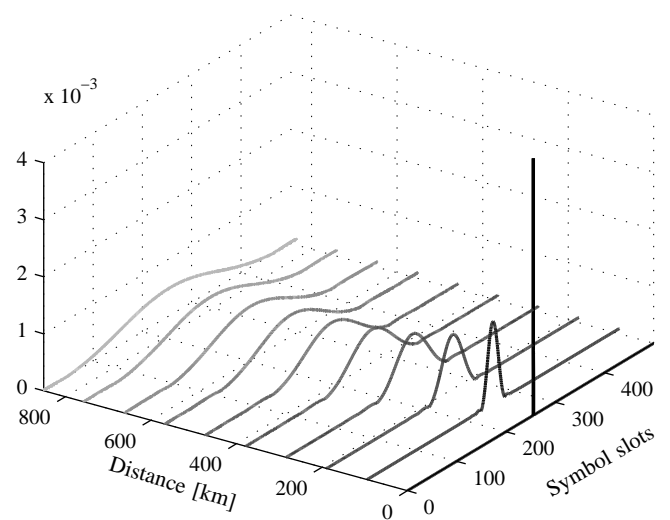


Fig. 1. Amplitude for a linearly propagating 15.6 ps raised-cosine pulse (compatible with 32 GBaud) over 850 km fiber with  $\beta_2 = -21.7 \text{ ps}^2/\text{km}$ . The lossy NLSE over 10 amplifier spans was simulated, with ASE noise switched off for clarity, and the peak power used was 0.1 mW.

1))  $\sum_{i=k-N}^{k+N} |X_i|^2$  at any discrete time  $k$  is a random variable that depends on the magnitude of the  $k$ th symbol and the  $2N$  symbols around it. In the limit  $N \rightarrow \infty$ , this empirical power converges to the “statistical” power  $P$  in (3), for any i.i.d. process with power  $P$ , as mentioned in Sec. II-C. This observation shows that the proposed finite-memory model in (8) “converges” to the GN model in (2), provided that the channel memory  $N$  is sufficiently large and that the process consists of i.i.d. symbols with zero mean and variance  $P$ .

The purpose of the finite-memory model is to be able to predict the output of the channel when the transmitted symbols are not i.i.d. This is the case for example when the transmitted symbols are a nonstationary process (as will be exemplified in Sec. II-E) and also for coded sequences (which we discuss in Sec. IV). An advantage of the finite-memory model, from a theoretic viewpoint, is that the input–output relation of the channel is modeled as a fixed conditional probability of the output given the input and its history, which is the common notion of a channel model in communication and information theory ever since the work of Shannon [15], [44, p. 74]. This is in contrast to the regular GN model and other channel models, whose conditional distribution change depending on which transmitter the channel is connected to. Specifically, the GN model is represented by a family of such conditional distributions, one for each value of the transmitter parameter  $P$ .

A drawback with the proposed finite-memory model is that it is more complex than the GN model. Also, our model is not accurate for small values of  $N$ , since the GN assumption relies on the central limit theorem [7], [9], [10]. Furthermore, we assumed that all the  $2N$  symbols around the symbol  $X_k$  affect the noise variance equally. In practice, this is not the case. We nevertheless use the proposed model in this paper because it is relatively easy to analyze (see Sec. III and IV)

TABLE I  
SYSTEM PARAMETERS USED IN THE PAPER.

Symbol	Value	Meaning
$\alpha$	0.2 dB/km	Fiber attenuation
$\beta_2$	$-21.7 \text{ ps}^2/\text{km}$	Group velocity dispersion
$\gamma$	$1.27 \text{ (W km)}^{-1}$	Fiber nonlinear coefficient
$M$	10	Number of amplifier spans
$L$	850 km	System length
$R_s$	32 Gbaud	Symbol rate
$P_{\text{ASE}}$	$4.1 \cdot 10^{-6} \text{ W}$	Total ASE noise
$\eta$	$7244 \text{ W}^{-2}$	NLI coefficient

and because even this simple finite-memory model captures the quantitative effects caused by non-i.i.d. symbols, which is essential for the capacity analysis in Sec. IV.

### E. Numerical Comparison

Before analyzing the finite-memory GN model, we first quantify the chromatic dispersion of an optical fiber. To this end, we simulated the transmission of a single symbol pulse over a over a single-channel, single-polarization fiber link without dispersion compensation. Ten amplifiers spans over a total distance of 850 km are simulated using the lossy NLSE model. We used a raised-cosine pulse with peak power 0.1 mW and a duration of 15.6 ps at half the maximum amplitude, which corresponds to half the symbol slot in a 32 Gbaud transmission system. The result is illustrated in Fig. 1. At this low power, the nonlinear effects are almost negligible. For clarity of illustration, the ASE noise was neglected by setting  $P_{\text{ASE}} = 0$ . The remaining system parameters are given in Table I and will be used throughout the paper. As we can see, the pulse broadens as it propagates along the fiber, having a width corresponding to hundreds of data symbols after 850 km of transmission. This is in good agreement with the relation for symbol memory used in [45, p. 2037], which gives  $2N \approx 2\pi|\beta_2|LR_s^2 = 119$ .

Next, to validate the behavior of the finite-memory model with nonstationary input symbol sequences, we simulated the transmission of independent quadrature phase-shift keying (QPSK) data symbols with a time-varying magnitude, over the same 700 km fiber link, at  $R_s = 32$  Gbaud. The transmitted sequence consists of 200 symbols with 6 mW of signal power, 200 symbols at 0 mW power, 200 symbols at 6 mW, and so on. The statistical average power is then 3 mW. The chosen pulse shape is a raised-cosine return-to-zero pulse. In Fig. 2, we show the amplitude of the transmitted symbols  $|X_k|$  (red) and received symbols  $|Y_k|$  (blue) with three different models: the NLSE, the finite-memory GN model with  $N = 60$ , and the regular GN model. In the middle and lower plots of Fig. 2, we used the NLI coefficient  $\eta = 7244 \text{ W}^{-2}$ , which was calculated from (7). The single-polarization ASE noise power was calculated as [46, Eqs. (7.4.1), (7.2.15)]  $P_{\text{ASE}} = h\nu MFG R_s/2 = 4.1 \mu\text{W}$ , where  $h\nu = 1.28 \cdot 10^{-19} \text{ J}$  is the photon energy,  $F = 6 \text{ dB}$  is the amplifier noise figure, and  $G = \alpha L/M = 17 \text{ dB}$  is the amplifier gain.

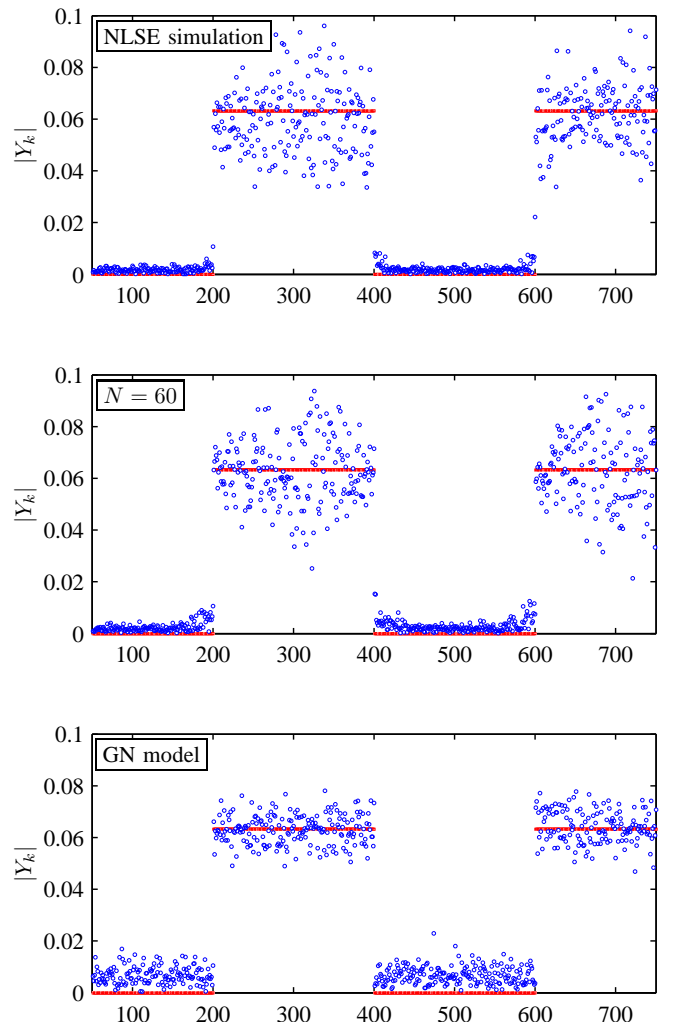


Fig. 2. Amplitude of the transmitted QPSK symbols  $|X_k|$  (red squares) and received symbols  $|Y_k|$  (blue circles) transmitted in a 850 km fiber link. The received symbols are obtained using (top) the NLSE, (middle) the finite-memory GN model (8) with  $N = 60$ , and (bottom) the regular GN model (2).

As can be seen, the agreement between the NLSE simulations and the finite memory model is quite reasonable, but the GN model cannot capture the nonstationary dynamics. The results in Fig. 2 also show that the noise variance in the NLSE simulation is low around the symbols with low signal power and high around the symbols with high signal power. This behavior is captured by the finite-memory GN model but not by the regular GN model, for which the variance of the noise is the same for any time instant. This illustrates that the GN model (2) should be avoided with nonstationary transmit signals as the ones used in Fig. 2, at least if the time constant of the transmitted signal (as quantified by, e.g., the periodicity or block length) is shorter than the time constant of the physical channel. This is not surprising, as the model was derived under an i.i.d. assumption. In Sec. V, we will return to this observation when analyzing coded transmission.

### III. UNCODED ERROR PROBABILITY

We assume that the transmitted symbols  $\{X_k\}$  are independently drawn from a discrete constellation  $\mathcal{S} = \{s_1, \dots, s_{2^m}\}$ . The symbols are assumed to be selected with the same probability, and thus, the average transmit (statistical) power is given by

$$P = \mathbb{E}[|X|^2] = \frac{1}{2^m} \sum_{s \in \mathcal{S}} |s|^2. \quad (9)$$

For each time instant  $k$ , we denote the sequence of the  $2N$  symbols transmitted around  $X_k$  by

$$\mathbf{X}_k^{\text{mem}} \triangleq [X_{k-N}, \dots, X_{k-1}, X_{k+1}, \dots, X_{k+N}], \quad (10)$$

where the notation emphasizes that  $\mathbf{X}_k^{\text{mem}}$  is a random vector describing the channel memory at time instant  $k$ .

For a given sequence of  $2N$  symbols  $\mathbf{X}_k^{\text{mem}} = \mathbf{x}_k^{\text{mem}}$  and a given transmitted symbol  $X_k = x_k$ , the noise in (8) can be expressed as

$$Z_k = \tilde{Z}_k \sqrt{\rho(|x_k|^2 + \|\mathbf{x}_k^{\text{mem}}\|^2)}, \quad (11)$$

where

$$\rho(a) \triangleq P_{\text{ASE}} + \eta \left( \frac{a}{2N+1} \right)^3. \quad (12)$$

and  $\|\mathbf{x}\|$  denotes the Euclidean norm of  $\mathbf{x}$ . The channel law for the finite-memory model is then

$$\begin{aligned} f_{Y_k|X_k, \mathbf{X}_k^{\text{mem}}}(y|x_k, \mathbf{x}_k^{\text{mem}}) \\ \triangleq \frac{1}{\pi \rho(|x_k|^2 + \|\mathbf{x}_k^{\text{mem}}\|^2)} \exp\left(-\frac{|y-x_k|^2}{\rho(|x_k|^2 + \|\mathbf{x}_k^{\text{mem}}\|^2)}\right). \end{aligned} \quad (13)$$

#### A. Error Probability Analysis

We consider the equally spaced 16-QAM constellation shown in Fig. 3. In this case,  $\mathcal{S} = \{a + b\sqrt{-1} : a, b \in \{\pm\Delta, \pm3\Delta\}\}$ , the minimum Euclidean distance (MED) of the constellation is  $2\Delta$ , and the statistical average power (9) is  $P = 10\Delta^2$ . The binary labeling is the binary reflected Gray code (BRGC) [47], where the first two bits determine the in-phase (real) component of the symbols and the last two bits determine the quadrature (imaginary) components of the symbols. This is shown with colors in Fig. 3.

The maximum-likelihood (ML) symbol-by-symbol detection rule for a given sequence  $\mathbf{x}_k^{\text{mem}}$  chooses the symbol  $s \in \mathcal{S}$  that maximizes  $f_{Y_k|X_k, \mathbf{X}_k^{\text{mem}}}(y|s, \mathbf{x}_k^{\text{mem}})$  in (13). The decision made by this detector can be expressed as

$$\hat{X}_k^{\text{ML}} \triangleq \underset{s \in \mathcal{S}}{\text{argmin}} \left\{ \log \rho(|s|^2 + \|\mathbf{x}_k^{\text{mem}}\|^2) + \frac{|y-s|^2}{\rho(|s|^2 + \|\mathbf{x}_k^{\text{mem}}\|^2)} \right\}, \quad (14)$$

which shows that, due to the dependency of  $\log \rho(|s|^2 + \|\mathbf{x}_k^{\text{mem}}\|^2)$  on  $s$ , this detector is not an MED detector. For simplicity, however, we disregard this term and study the MED detector, which chooses the symbol  $s$  being

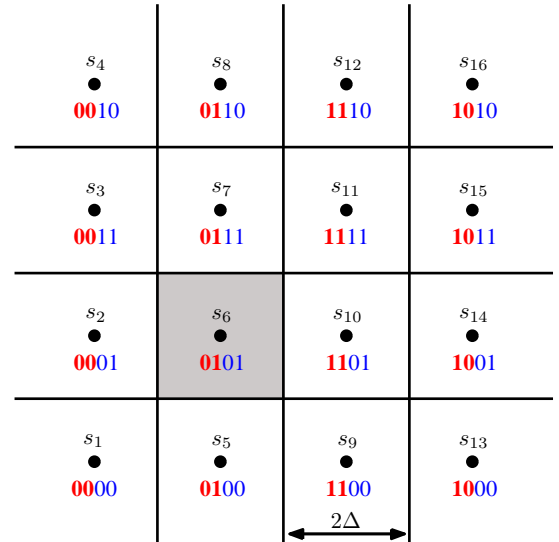


Fig. 3. The 16-QAM constellation  $\mathcal{S}$  and its binary labeling. The binary labeling of the constellation is based on the Cartesian product of the BRGC for 4-ary pulse amplitude modulation in phase (bold, red) and quadrature (blue). The Voronoi regions of the symbols and the MED of the constellation are also shown. The Voronoi region  $\mathcal{V}_6$  is highlighted in gray.

closest, in Euclidean distance, to the channel output  $Y_k$ . Thus

$$\begin{aligned} \hat{X}_k &= \underset{s \in \mathcal{S}}{\text{argmin}} |Y_k - s|^2 \\ &= s_i, \quad \text{if } Y_k \in \mathcal{V}_i, \end{aligned} \quad (15)$$

where  $\mathcal{V}_i$  denotes the decision region, or *Voronoi region*, of  $s_i$ .

*Remark 1:* As we will see later, for increasing memory  $N$ , the MED detector in (15) in fact converges to the detector in (14). Intuitively, this holds because the relative error in the approximation  $\|\mathbf{x}_k^{\text{mem}}\|^2 + |s|^2 \approx \|\mathbf{x}_k^{\text{mem}}\|^2$  vanishes when  $N$  is large.

*Remark 2:* In this paper, we focus on the MED detector in (15) because of its simplicity and widespread practical usage. The ML symbol-by-symbol detector in (14) is considered as a theoretical benchmark, but it is still suboptimal. We do not in this paper consider the optimal detector for channels with memory, the ML sequence detector, which can be implemented using the Viterbi algorithm but is very complex for channels with long memory [48].

The following two theorems give closed-form expressions for the BER and SER for the constellation in Fig. 3 when used over the finite-memory GN model.

*Theorem 1:* Gray-coded 16-QAM transmission over the finite-memory GN model with memory  $N$  yields after MED detection the BER

$$\begin{aligned} \text{BER} &= \frac{1}{2^{4N+3}} \sum_{l=0}^{4N} \binom{4N}{l} (2\lambda_{l,1,1} + 3\lambda_{l,1,5} + \lambda_{l,1,9} \\ &\quad + \lambda_{l,3,1} + 2\lambda_{l,3,5} + \lambda_{l,3,9} - \lambda_{l,5,5} - \lambda_{l,5,9}), \end{aligned} \quad (16)$$



where

$$\lambda_{l,r,t} \triangleq Q\left(\sqrt{\frac{r^2 P/5}{\rho((2N+4l+t)P/5)}}\right) \quad (17)$$

and  $Q(x) = (1/\sqrt{2\pi}) \int_x^\infty \exp(-t^2/2) dt$  is the Gaussian Q function.

*Proof:* See Appendix A.  $\square$

**Theorem 2:** 16-QAM transmission over the finite-memory GN model with memory  $N$  yields after MED detection the SER

$$\begin{aligned} \text{SER} = \frac{1}{2^{4N+2}} \sum_{l=0}^{4N} \binom{4N}{l} (4\lambda_{l,1,1} - 4\lambda_{l,1,1}^2 \\ + 6\lambda_{l,1,5} - 4\lambda_{l,1,5}^2 + 2\lambda_{l,1,9} - \lambda_{l,1,9}^2), \end{aligned} \quad (18)$$

with  $\lambda_{l,r,t}$  defined in (17).

*Proof:* See Appendix B.  $\square$

The BER and SER in the limit  $N \rightarrow \infty$  can be inferred from Theorems 1 and 2 as shown in the next corollary.

**Corollary 1:** The BER and SER for the finite-memory GN model in the limit  $N \rightarrow \infty$  are

$$\begin{aligned} \text{BER} = \frac{3}{4} Q\left(\sqrt{\frac{P/5}{P_{\text{ASE}} + \eta P^3}}\right) + \frac{1}{2} Q\left(\sqrt{\frac{9P/5}{P_{\text{ASE}} + \eta P^3}}\right) \\ - \frac{1}{4} Q\left(\sqrt{\frac{5P}{P_{\text{ASE}} + \eta P^3}}\right), \end{aligned} \quad (19)$$

$$\text{SER} = 3Q\left(\sqrt{\frac{P/5}{P_{\text{ASE}} + \eta P^3}}\right) - \frac{9}{4} Q^2\left(\sqrt{\frac{P/5}{P_{\text{ASE}} + \eta P^3}}\right). \quad (20)$$

*Proof:* See Appendix C.  $\square$

The other extreme case to consider is the memoryless AWGN channel. The BER and SER expressions in this case are given in the following corollary.

**Corollary 2:** The BER and SER for the memoryless AWGN channel are given by

$$\begin{aligned} \text{BER} = \frac{3}{4} Q\left(\sqrt{\frac{P}{5P_{\text{ASE}}}}\right) + \frac{1}{2} Q\left(\sqrt{\frac{9P}{5P_{\text{ASE}}}}\right) \\ - \frac{1}{4} Q\left(\sqrt{\frac{5P}{P_{\text{ASE}}}}\right), \end{aligned} \quad (21)$$

$$\text{SER} = 3Q\left(\sqrt{\frac{P}{5P_{\text{ASE}}}}\right) - \frac{9}{4} Q^2\left(\sqrt{\frac{P}{5P_{\text{ASE}}}}\right). \quad (22)$$

*Proof:* Set  $\eta = 0$  in (19) and (20).  $\square$

The results in Corollaries 1 and 2 correspond to well-known expressions for the BER and SER for the AWGN channel. In particular, (21) can be found in [49, eq. (10)], [50, eq. (10.36a)] and (22) in [50, eq. (10.32)]. Also, the results in Corollary 2 together with (2) show that the BER and SER for the finite-memory GN model when  $N \rightarrow \infty$  converge to the BER and SER for the regular GN model.

### B. Numerical Results

We consider the same scenario as in Sec. II-E, with parameters according to Table I. The BER and SER for the

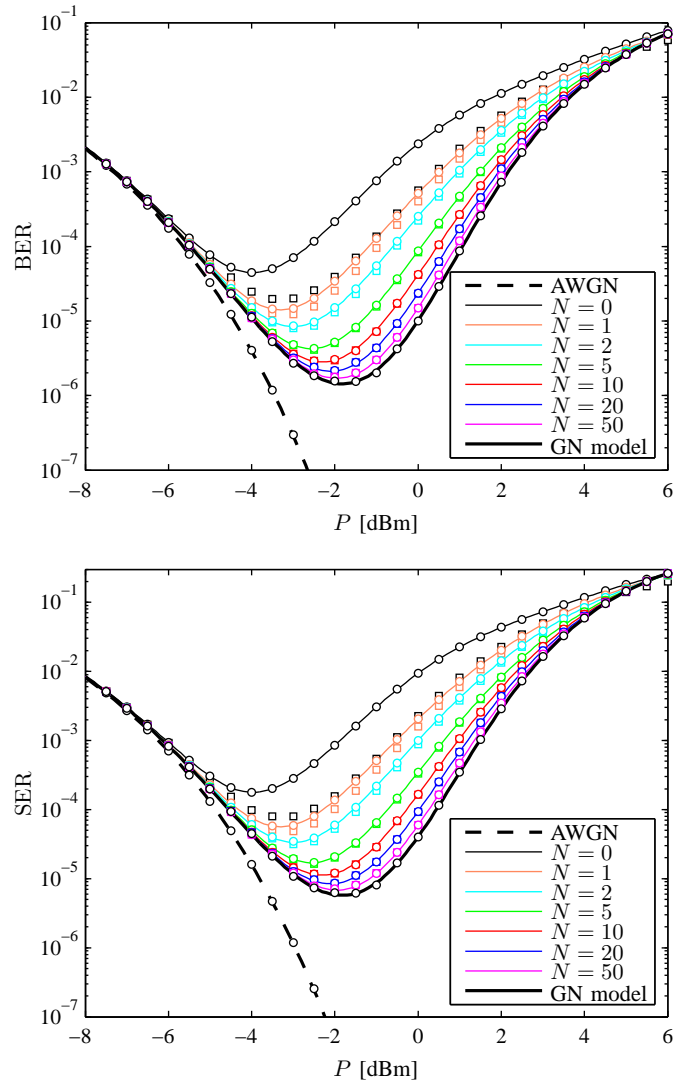


Fig. 4. Analytical BER (top) and SER (bottom) of 16-QAM transmission with the finite-memory GN model, for different values of  $N$  (solid lines). Markers show simulation results with the ML detector in (14) (squares) and the MED detector in (15) (circles). The results for the memoryless AWGN channel and the regular GN model are included for comparison.

16-QAM constellation in Fig. 3 given by Theorems 1 and 2 are shown in Fig. 4 for different values of  $N$ . Fig. 4 also shows the results for the asymptotic case  $N \rightarrow \infty$  and for the memoryless AWGN channel given by Corollaries 1 and Corollary 2, respectively. Furthermore, results obtained via computer simulations of (1)–(2) are included using the ML detector in (14), marked with squares, and the MED detector in (15), marked with circles. As expected, the MED detector yields a perfect match with the analytical expressions, whereas the ML detector deviates for small channel memories ( $N \leq 2$ ).

The results in Fig. 4 show that in the low-input-power regime, the memory in the channel plays no role for the BER and SER, and all the curves follow closely the BER and the SER of a memoryless AWGN channel. However, as  $P$  increases, the nonlinearity kicks in, causing the BER and SER to have a minimum whose value depends on  $N$ ,



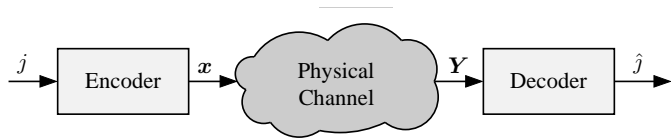


Fig. 5. Encoder and decoder pair. The encoder maps a message  $j$  to a codeword  $\mathbf{x} = [x_1, \dots, x_n]$ . The decoder uses the noisy observation  $\mathbf{Y} = [Y_1, \dots, Y_n]$  to provide an estimate  $\hat{j}$  of the message  $j$ .

and then to increase as  $P$  increases. Physically, this can be explained as follows: In the low-power regime, the BER is limited by the ASE noise, which is independent of the memory depth. In the high-power regime, the Kerr-induced noise dominates, resulting in increasing BER with power. Similar behavior has been reported in most experiments and simulations on nonlinearly-limited links, e.g., [9], [14], [51], [52], [46, Ch. 9]. The reason why the performance improves slightly with the memory depth  $N$  is the nonlinear scaling of the Kerr-induced noise. For  $N = 1$ , sequences of two or more high-amplitude symbols will receive high noise power and dominate the average BER. For higher  $N$ , longer (and less probable) sequences of high-amplitude symbols are required to receive the same, high, noise power. Thus on average the performance improves with  $N$ , up to a limit given by the GN model.

The results in Fig. 4 also show how the finite-memory model in the high signal power regime approaches the GN model, as predicted analytically in Sec. III-A. For  $N = 50$ , the two models yield very similar BER and SER curves.

#### IV. CHANNEL CAPACITY

In this section, some fundamentals of information theory are first reviewed. Then a lower bound on the capacity of the finite-memory GN model is derived and evaluated numerically.

##### A. Preliminaries

Fig. 5 shows a generic coded communication system where a message  $j$  is mapped to a codeword  $\mathbf{x} = [x_1, \dots, x_n]$ . This codeword is then used to modulate a continuous-time waveform, which is then transmitted through the physical channel. At the receiver's side, the continuous-time waveform is processed (filtered, equalized, synchronized, matched filtered, sampled, etc.) resulting in a discrete-time observation  $\mathbf{Y} = [Y_1, \dots, Y_n]$ , which is a noisy version of the transmitted codeword  $\mathbf{x}$ . The decoder uses  $\mathbf{Y}$  to estimate the transmitted message  $j$ .

When designing a coded communication system, the first step is to choose the set of codewords that will be transmitted through the channel. This set, called the *codebook*, is stored in both the encoder and decoder. Once the codebook has been chosen, the mapping rule between messages and codewords should be chosen, which fully determines the encoding procedure. At the receiver side, the decoder block will use the mapping rule used at the transmitter (as well as the channel characteristics) to give an estimate  $\hat{j}$  of the message  $j$ . The triplet codebook, encoder, and decoder forms a so-called

*coding scheme*. Practical coding schemes are designed so as to minimize the probability that  $\hat{j}$  differs from  $j$ , while at the same time keeping the complexity of both encoder and decoder low.

Channel capacity is the largest transmission rate at which reliable communications can occur. More formally, let  $(n, M, \epsilon)$  be a coding scheme consisting of:

- An encoder that maps a message  $j \in \{1, \dots, M\}$  into a block of  $n$  transmitted symbols  $\mathbf{x} = [x_1, \dots, x_n]$  satisfying a per-codeword power constraint

$$\frac{1}{n} \sum_{l=1}^n |x_l|^2 = P. \quad (23)$$

- A decoder that maps the corresponding block of received symbols  $\mathbf{Y} = [Y_1, \dots, Y_n]$  into a message  $\hat{j} \in \{1, \dots, M\}$  so that the average error probability, i.e., the probability that  $\hat{j}$  differs from  $j$ , does not exceed  $\epsilon$ .

Observe that  $P$  here is defined differently from the previous sections. It still represents the average transmit power, but while this quantity in Sec. II–III was interpreted in a statistical sense as the mean of an i.i.d. random variable, it is in this section the exact power of *every* codeword.

The maximum coding rate  $R^*(n, \epsilon)$  (measured in bit/symbol) for a given block length  $n$  and error probability  $\epsilon$  is defined as the largest ratio  $(\log_2 M)/n$  for which an  $(n, M, \epsilon)$  coding scheme exists. The channel capacity  $C$  is the largest coding rate for which a coding scheme with vanishing error probability exists, in the limit of large block length,

$$C \triangleq \lim_{\epsilon \rightarrow 0} \lim_{n \rightarrow \infty} R^*(n, \epsilon). \quad (24)$$

##### B. Memoryless Channels

By Shannon's *channel coding theorem*, the channel capacity (in bit/symbol) of a discrete-time memoryless channel with codewords subject to the power constraint (23) can be calculated as [15], [16, Ch. 7]

$$C = \sup I(X; Y), \quad (25)$$

where  $I(X; Y)$  is the *mutual information* (MI)

$$I(X; Y) = \iint f_{X,Y}(x, y) \log_2 \frac{f_{X,Y}(x, y)}{f_X(x)f_Y(y)} dx dy \quad (26)$$

and the maximization in (25) is over all probability distributions  $f_X$  that satisfy  $\mathbb{E}[|X|^2] = P$ , for a given channel  $f_{Y|X}$ .

Roughly speaking, a transmission scheme that operates at an arbitrary rate strictly less than  $C$  can be designed by creating a codebook of  $M = 2^{nR}$  codewords of length  $n$ , whose elements are i.i.d. random samples from the distribution  $f_X$  that maximizes the mutual information in (25). During transmission, the encoder maps each message  $j$  into a unique codeword  $\mathbf{x}$ , and the decoder identifies the codeword that is most similar, in some sense, to the received vector  $\mathbf{Y}$ . An arbitrarily small error probability  $\epsilon$  can be achieved by choosing  $n$  large enough. This *random coding* paradigm was proposed already by Shannon [15]. In practice, however, randomly constructed codebooks are usually avoided for complexity reasons.

Since the additive noise in (2) is statistically independent of  $X_k$ , the channel is memoryless, and thus, the channel capacity of the GN model (2) can be calculated exactly as [5], [35]

$$C = \log_2 \left( 1 + \frac{P}{P_{\text{ASE}} + \eta P^3} \right) \quad (27)$$

using Shannon's well-known capacity expression [15, Sec. 24], [16, Ch. 9]. The capacity in (27) can be achieved by choosing the elements of all codewords  $\mathbf{x}$  to be drawn independently from a Gaussian distribution  $\mathcal{CN}(0, P)$ .

Considered as a function of the transmitted signal power  $P$ , the capacity in (27) has the peculiar behavior of reaching a peak and eventually decreasing to zero at high enough power, since the denominator of (27) increases faster than the numerator. This phenomenon, sometimes called the "nonlinear Shannon limit" in the optical communications community, conveys the message that reliable communication over nonlinear optical channels becomes impossible at high powers. In the following sections, we shall question this pessimistic conclusion.

### C. Channels with Memory

The capacity of channels with memory is, under certain assumptions on information stability [53, Sec. I],

$$C = \limsup_{n \rightarrow \infty} \frac{1}{n} I(\mathbf{X}_1^n; \mathbf{Y}_1^n), \quad (28)$$

where  $\mathbf{X}_i^j = (X_i, X_{i+1}, \dots, X_j)$ ,  $I(\mathbf{X}_i^j; \mathbf{Y}_i^j)$  is defined as a multidimensional integral analogous to (26), and the maximization is over all joint distributions of  $X_1, \dots, X_n$  satisfying  $\mathbb{E}[\|\mathbf{X}_1^n\|^2] = nP$ . In this context, it is worth emphasizing that the maximization in (28) includes sequences  $X_1, \dots, X_n$  that are not i.i.d. Hence, in order to calculate the channel capacity of a transmission link, it is essential that the employed channel model allows non-i.i.d. inputs.

An exact expression for the channel capacity of the finite-memory GN model (8) is not available. Shannon's formula, which leads to (27), does not apply here, because the sequences  $\{X_k\}$  and  $\{Z_k\}$ , where  $Z_k$  was defined in (8), are dependent. A capacity estimation via (28) is numerically infeasible, since it involves integration and maximization over high-dimensional spaces. We therefore turn our attention to bounds on the capacity for the finite-memory model. Every joint distribution of  $X_1, \dots, X_k$  satisfying  $\mathbb{E}[\|\mathbf{X}_1^n\|^2] = nP$  gives us a lower bound on capacity. Thus,

$$C \geq \lim_{n \rightarrow \infty} \frac{1}{n} I(\mathbf{X}_1^n; \mathbf{Y}_1^n), \quad (29)$$

for any random process  $\{X_k\}$  such that the limit exists.

### D. Lower Bound

In this section, a lower bound on (28) is derived by applying (29) to the following random input process. In every block of  $2N + 1$  consecutive symbols, we let the first  $N$  symbols and the last  $N$  symbols have a constant amplitude, whereas the amplitude of the symbol in the middle of the block follows an arbitrary distribution. The phase of each symbol in the block is assumed uniform. With this random input process,

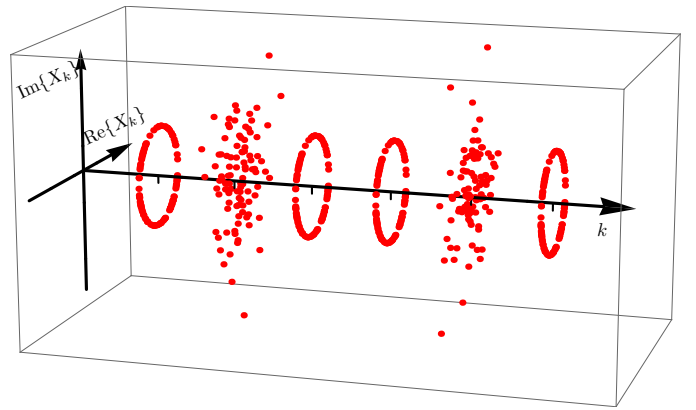


Fig. 6. Six samples of the random input process  $\{X_k\}$  used to generate the lower bound in Theorem 3. The channel memory is here  $N = 1$ , meaning that  $2N + 1 = 3$  input symbols  $X_k$  influence each output symbol. The distributions are illustrated as scatter plots of 1000 realizations for each sample.

illustrated in Fig. 6, the memory in (8) depends only on a single variable-amplitude symbol. This enables us to derive an analytical expression for the resulting capacity lower bound in (29).

*Theorem 3:* For every  $r_1 \geq 0$  and every probability distribution  $f_R$  over  $\mathbb{R}^+$  such that

$$\frac{2Nr_1^2 + \mathbb{E}[R^2]}{2N + 1} = P, \quad (30)$$

where  $R \sim f_R$ , the channel capacity of (8) is lower-bounded as

$$C \geq -\frac{\mathbb{E}[\log_2 f_{\mathbf{U}}(\mathbf{U})]}{2N + 1} - \int_0^\infty f_R(r) \log_2(e\rho(2Nr_1^2 + r^2)) dr. \quad (31)$$

Here,  $\mathbf{U} \triangleq [U_{-N}, U_{-N+1}, \dots, U_N]$  is a random vector whose probability density function  $f_{\mathbf{U}}$  is

$$f_{\mathbf{U}}(\mathbf{u}) = \int_0^\infty f_R(r) \frac{\exp\left(-\frac{\sum_{k=-N}^N u_k + 2Nr_1^2 + r^2}{\rho(2Nr_1^2 + r^2)}\right)}{(\rho(2Nr_1^2 + r^2))^{2N+1}} \cdot I_0\left(\frac{2r\sqrt{u_0}}{\rho(2Nr_1^2 + r^2)}\right) \cdot \prod_{\substack{k=-N \\ k \neq 0}}^N I_0\left(\frac{2r_1\sqrt{u_k}}{\rho(2Nr_1^2 + r^2)}\right) dr, \quad (32)$$

where the function  $\rho(\cdot)$  is defined in (12), and  $I_0(u)$  is the modified Bessel function of the first kind.

*Proof:* See Appendix D.  $\square$

The bound will be numerically computed in the next section.

### E. Numerical Results

Theorem 3 yields a lower bound on capacity for every constant  $r_1$  and every probability distribution  $f_R$  satisfying (30). Instead of optimizing the bound over all distributions  $f_R$ , which is of limited interest, since the theorem itself provides only a lower bound on capacity, we study a heuristically

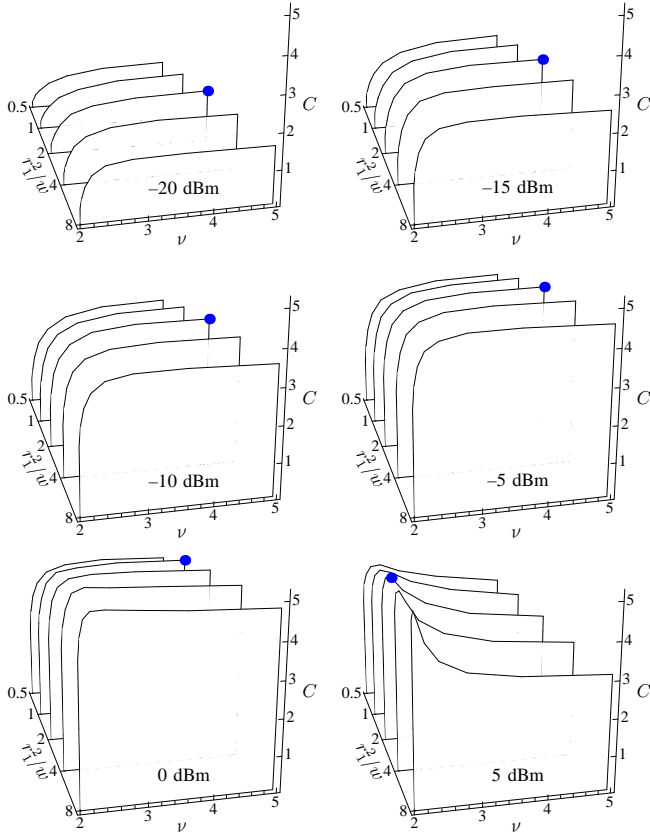


Fig. 7. Lower bounds on capacity from Theorem 3 as a function of  $\nu$ , for various parameters  $P$  and  $r_1^2/w$ . Dots indicate the highest MI for each  $P$ . The memory is  $N = 1$ .

chosen family of distributions and optimize its parameters along with the constant amplitude  $r_1$ .

An attractive distribution in this context is to let the variable-amplitude symbols follow a complex, circularly symmetric *bivariate t-distribution* [54, p. 86], [55, p. 1],

$$f_X(x) = \frac{1}{2\pi w} \left(1 + \frac{|x|^2}{\nu w}\right)^{-(1+\nu/2)}, \quad (33)$$

where  $X$  (with magnitude  $R = |X|$ ) denotes one such variable-amplitude symbol,  $\nu$  is a shape parameter, and  $w$  scales the variance, which equals [55, p. 11]  $\mathbb{E}[|X|^2] = \mathbb{E}[R^2] = 2\nu w/(\nu - 2)$  if  $\nu > 2$  and is otherwise undefined. The shape of this distribution is similar to a Gaussian, but the heaviness of the tail can be controlled via the shape parameter  $\nu$ : the closer  $\nu$  is to 2, the heavier tail. This is, as we shall see later, what makes it an interesting choice for nonlinear optical channels.

Again, we consider the same scenario as in Sec. II-E, with the system parameters given in Table I. The distribution of  $R = |X|$  is given by  $f_R(r) = 2\pi r f_X(r)$ , with  $f_X$  given by (33). The power constraint (30), which reduces to

$$P = \frac{1}{2N+1} \left(2N r_1^2 + \frac{2\nu w}{\nu-2}\right),$$

leaves two degrees of freedom to optimize for each  $P$ , which

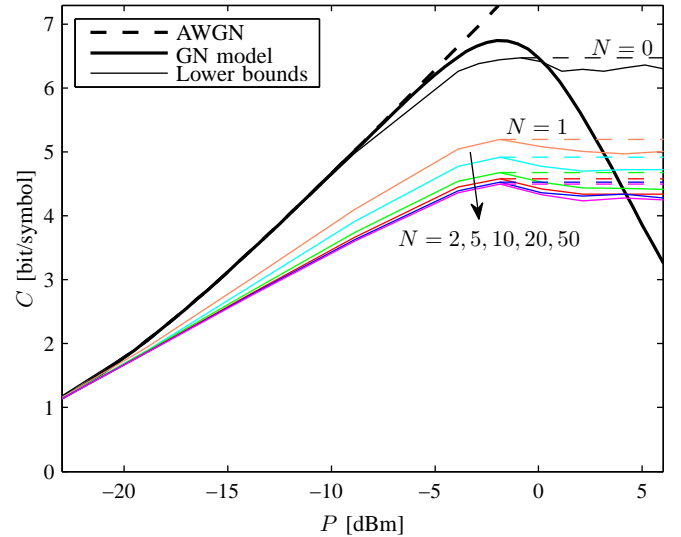


Fig. 8. Lower bounds from Theorem 3 on the capacity of the finite-memory model for different values of  $N$ . The exact capacities of the AWGN channel and the GN model in (27) are included for comparison. Observe that the capacity of the finite-memory model does not converge to the capacity of the GN model as the memory  $N$  increases. Dashed lines indicate improved lower bounds via the law of monotonic channel capacity.

we can take to be the shape parameter  $\nu$  and the ratio  $r_1^2/w$ .

The lower bound on the capacity of the finite-memory model given by Theorem 3 is shown in Fig. 7 as a function of  $P$ ,  $\nu$ , and  $r_1^2/w$ , for the special case  $N = 1$ . The expectation in (31) was estimated by Monte Carlo integration. It can be seen that at low to medium power, the optimum shape parameter  $\nu$  is high, which means that the distribution is similar to a Gaussian. As the transmit power  $P$  increases above approximately 0 dBm, the optimum shape parameter  $\nu$  gets closer and closer to 2. This means that the tail gets heavier, so that at high power, it consumes almost all power, while the probability of transmitting a high amplitude  $R$  is still small. In this sense, a t-distribution with a shape parameter near 2 is similar to a satellite constellation [41].

Selecting the optimum parameters  $\nu$  and  $r_1^2/w$  for every power  $P$ , the capacity bound is plotted in Fig. 8 as a function of transmit power  $P$ , for selected values of the channel memory  $N$ . The figure also shows the AWGN channel capacity and the exact capacity of the GN model given by (27). In the low-power (linear) regime, the capacity bound is close to the AWGN capacity if  $N = 0$ , because the t-distribution is, at high values of  $\nu$ , approximately equal to the capacity-achieving Gaussian distribution. As  $N$  increases, the capacity bound tends, still in the linear regime, to the mutual information of constant-amplitude transmission [56], [57].

Interestingly, we can see that as  $N$  increases, the curves approach an asymptotic bound (the curves for  $N = 10, 20,$  and  $50$  almost overlap). It follows that reliable communication in the high signal power regime is indeed possible for every finite  $N$ . Qualitatively similar results can be obtained using satellite constellations. These results should be compared with the regular GN model, whose capacity (27) decreases to zero at high average transmit power [35]. It may seem contradictory

that the GN model, which can be characterized as a limiting case of the finite-memory model (cf. (8) and (2)–(3)), nevertheless exhibits a fundamentally different channel capacity. This can be intuitively understood as follows. For every block of  $2N + 1$  symbols, we transmit  $2N$  constant-amplitude symbols with low power and only one symbol with variable (potentially very large) power. Although the amplitude of this variable-power symbol is chosen so that the average power constraint is satisfied according to (30) (which requires averaging across many blocks of length  $2N + 1$ ), the convergence to average power illustrated in (3) does not occur *within* a block, even when  $N$  is taken very large.

While the results in this section apply exclusively to the finite-memory model given by (1) and (8), qualitatively similar results can be expected for any other finite-memory model. This is a result of the *law of monotonic channel capacity*, which, cast in the framework of this paper, states that the channel capacity never decreases with power for any finite-memory channel [58]. This law does not give a capacity lower bound *per se*, but it provides an instrument by which a lower bound at a certain power  $P$  can be propagated to any power greater than  $P$ . Hence, observing that the lower bounds in Fig. 8 all exhibit a low peak before they converge to their asymptotic values at high  $P$ , we conclude that marginally tighter lower bounds can be obtained by flattening out the part of the curves to the right of these peaks (dashed lines).

## V. DISCUSSION AND CONCLUSIONS

We extended the popular GN model for nonlinear fiber channels with a parameter to account for the channel memory. The extended channel model, which is given by (8), is able to model the time-varying output of an optical fiber whose input is a nonstationary process. If the input varies on a time scale comparable to or longer than the memory of the channel, then this model gives more realistic results than the regular GN model, as we showed in Fig. 2.

The validity of the GN model remains undisputed in the case of i.i.d. input symbols, such as in an uncoded scenario with a fixed, not too heavy-tailed modulation format<sup>2</sup> and a fixed transmit power. These are the conditions under which the GN model was derived and validated. The uncoded bit and symbol error rates computed in Sec. III confirm that the finite-memory model behaves similarly to the GN model as the channel memory  $N$  increases.

The scene changes completely if we instead study capacity, as in Fig. 8. In this case, the finite-memory GN model does not, even at high  $N$ , behave as the regular GN model. This is because the channel capacity by definition involves a maximization over all possible transmission schemes, including nonstationary input, heavy-tailed modulation formats, etc. In the high-power (nonlinear) regime, it turns out to be beneficial to transmit using a heavy-tailed input sequence, whose output the GN model cannot reliably predict. Hence, the GN model and other infinite-memory models (in the sense defined in Sec. II-C) should be used with caution in capacity

<sup>2</sup>Examples of “heavy-tailed” modulation formats are t-distributions (Sec. IV-E) and satellite constellations [41].

analysis. It is still possible (and often easy) to calculate the capacity of such channel models, but this capacity should not be interpreted as the capacity of some underlying physical phenomenon with a finite memory. As a rule of thumb, if the model depends on the average transmit power, we recommend to avoid it in capacity analysis.

A challenging area for future work would be to derive more realistic finite-memory models than (8), i.e., discrete-time channel models that give the channel output as a function of a finite number of input symbols, ideally including not only a time-varying sequence of symbols but also symbols in other wavelengths, polarizations, modes, and/or cores, and to analyze these models from an information-theoretic perspective. This may lead to innovative new transmission techniques, which may potentially increase the capacity significantly over known results in the nonlinear regime. The so-called nonlinear Shannon limit, which has only been derived for infinite-memory channel models, does not prevent the existence of such techniques.

## APPENDIX A PROOF OF THEOREM 1

We first compute the PMF  $P_{\|\mathbf{X}_k^{\text{mem}}\|^2}$ . As  $\|\mathbf{X}_k^{\text{mem}}\|^2$  is a sum of  $2N$  i.i.d. random variables, its PMF is the  $2N$ -fold self-convolution of the PMF of one such random variable. This convolution can be readily computed using probability generating functions [59, Sec. 5.1]. Let

$$\begin{aligned} \hat{P}_{|X_k|^2}(z) &= \frac{1}{4}(z^{2\Delta^2} + 2z^{10\Delta^2} + z^{18\Delta^2}) \\ &= \frac{1}{4}(z^{\Delta^2} + z^{9\Delta^2})^2 \end{aligned} \quad (34)$$

denote the probability generating function of  $|X_k|^2$ . The probability generating function of  $\|\mathbf{X}_k^{\text{mem}}\|^2$  is

$$\hat{P}_{\|\mathbf{X}_k^{\text{mem}}\|^2}(z) = (\hat{P}_{|X_k|^2}(z))^{2N} \quad (35)$$

$$= \frac{1}{4^{2N}}(z^{\Delta^2} + z^{9\Delta^2})^{4N} \quad (36)$$

$$= \sum_{l=0}^{4N} \frac{1}{2^{4N}} \binom{4N}{l} z^{(4N+8l)\Delta^2}. \quad (37)$$

We see from (37) that the PMF of  $\|\mathbf{X}_k^{\text{mem}}\|^2$  is

$$\Pr\{\|\mathbf{X}_k^{\text{mem}}\|^2 = \delta_l\} = \frac{1}{2^{4N}} \binom{4N}{l}, \quad l = 0, 1, \dots, 4N, \quad (38)$$

where

$$\delta_l \triangleq (4N + 8l)\Delta^2. \quad (39)$$

Let  $\text{BER}_{l,q,i}$  denote the conditional error probability of bit  $q$ , given  $\|\mathbf{X}_k^{\text{mem}}\|^2 = \delta_l$  and  $X_k = s_i$ . The overall BER is obtained by averaging  $\text{BER}_{q,l,i}$  over all  $q = 1, \dots, 4$ ,  $l = 1, \dots, 4N$ , and  $i = 1, \dots, 16$ . However, due to the symmetric constellation and labeling in Fig. 3, it is sufficient to average

over  $q = 1, 2$ ,  $i \in \{1, 2, 5, 6\}$ , and all  $l$ , i.e.,

$$\text{BER} = \sum_{l=0}^{4N} \Pr\{\|\mathbf{X}_k^{\text{mem}}\|^2 = \delta_l\} \cdot \frac{1}{8} \sum_{q=1}^2 \sum_{i \in \{1, 2, 5, 6\}} \text{BER}_{l,q,i}. \quad (40)$$

By (13), the received symbol  $Y_k$  given  $\|\mathbf{X}_k^{\text{mem}}\|^2 = \delta_l$  and  $X_k = s_i$  is a complex, circularly symmetric Gaussian variable with mean  $s_i$  and variance  $\rho(\delta_l + |s_i|^2)/2$  per real dimension. Therefore,  $\text{BER}_{q,l,i}$  can be expressed in terms of the Gaussian Q-function

$$\lambda_{l,r,t} \triangleq Q\left(\sqrt{\frac{2r^2\Delta^2}{\rho(\delta_l + 2t\Delta^2)}}\right). \quad (41)$$

These expressions are obtained by inspecting Fig. 3 as

$$\begin{aligned} \text{BER}_{l,1,1} &= \lambda_{l,3,9}, & \text{BER}_{l,2,1} &= \lambda_{l,1,9} - \lambda_{l,5,9}, \\ \text{BER}_{l,1,2} &= \lambda_{l,3,5}, & \text{BER}_{l,2,2} &= \lambda_{l,1,5} - \lambda_{l,5,5}, \\ \text{BER}_{l,1,5} &= \lambda_{l,1,5}, & \text{BER}_{l,2,5} &= \lambda_{l,1,5} + \lambda_{l,3,5}, \\ \text{BER}_{l,1,6} &= \lambda_{l,1,1}, & \text{BER}_{l,2,6} &= \lambda_{l,1,1} + \lambda_{l,3,1}. \end{aligned} \quad (42)$$

Finally, (17) is obtained by substituting (39) and  $\Delta^2 = P/10$  into (41), and (16) is obtained by substituting (38) and (42) into (40).

#### APPENDIX B PROOF OF THEOREM 2

In analogy with (40), the SER is

$$\text{SER} = \sum_{l=0}^{4N} \Pr\{\|\mathbf{X}_k^{\text{mem}}\|^2 = \delta_l\} \cdot \frac{1}{4} \sum_{i \in \{1, 2, 5, 6\}} \text{SER}_{l,i}, \quad (43)$$

where  $\text{SER}_{l,i}$  is the conditional symbol error probability given  $\|\mathbf{X}_k^{\text{mem}}\|^2 = \delta_l$  and  $X_k = s_i$ . Note that there is no dependency on the bit  $q$ , as the SER does not depend on the labeling. The conditional SER can again be identified from Fig. 3 as

$$\begin{aligned} 1 - \text{SER}_{l,1} &= (1 - \lambda_{l,1,9})^2, \\ 1 - \text{SER}_{l,2} &= (1 - \lambda_{l,1,5})(1 - 2\lambda_{l,1,5}), \\ 1 - \text{SER}_{l,5} &= (1 - \lambda_{l,1,5})(1 - 2\lambda_{l,1,5}), \\ 1 - \text{SER}_{l,6} &= (1 - 2\lambda_{l,1,1})^2. \end{aligned} \quad (44)$$

The proof of (18) is completed by solving (44) for  $\text{SER}_{l,i}$ , substituting the result together with (38) into (43), and simplifying.

#### APPENDIX C PROOF OF COROLLARY 1

The SER in (18) can be expressed as

$$\text{SER} = u_{1,1} - u_{1,2} + \frac{3}{2}u_{5,1} - u_{5,2} + \frac{1}{2}u_{9,1} - \frac{1}{4}u_{9,2}, \quad (45)$$

where

$$u_{t,K} \triangleq \frac{1}{2^{4N}} \sum_{l=0}^{4N} \binom{4N}{l} \lambda_{l,1,t}^K. \quad (46)$$

We can interpret the sum in (46) in probabilistic terms as  $u_{t,K} = \mathbb{E}[\lambda_{L_{4N},1,t}^K]$ , where  $L_{4N}$  is a binomial random variable

with parameters  $(4N, 1/2)$ , i.e.,  $S_{4N}$  is the sum of  $4N$  i.i.d. random variables that take values 0 and 1 with the same probability. We use the notation  $L_{4N}$  to emphasize the dependency on  $N$ . To prove (20), we need to calculate  $\lim_{N \rightarrow \infty} u_{t,K}$ .

First, we write  $\lambda_{L_{4N},1,t}$  using (17) and (12) as

$$\lambda_{L_{4N},1,t} = Q\left(\sqrt{\frac{P/5}{P_{\text{ASE}} + \eta \left(\frac{(2N+4L_{4N}+t)P/5}{2N+1}\right)^3}}\right). \quad (47)$$

By the strong law of large numbers [59, Sec. 7.4.(3)], the ratio  $L_{4N}/(2N+1)$  converges almost surely to 1 as  $N \rightarrow \infty$ . Hence,

$$\lim_{N \rightarrow \infty} \lambda_{L_{4N},1,t} = Q\left(\sqrt{\frac{P/5}{P_{\text{ASE}} + \eta P^3}}\right) \quad (48)$$

for any constant  $t$ . Finally,

$$\begin{aligned} \lim_{N \rightarrow \infty} u_{t,K} &= \lim_{N \rightarrow \infty} \mathbb{E}[\lambda_{L_{4N},1,t}^K] \\ &= \mathbb{E}\left[\lim_{N \rightarrow \infty} \lambda_{L_{4N},1,t}^K\right] \end{aligned} \quad (49)$$

$$= \left[Q\left(\sqrt{\frac{P/5}{P_{\text{ASE}} + \eta P^3}}\right)\right]^K, \quad (50)$$

where the exchange of limit and expectation in (49) is possible because  $\lambda_{L_{4N},1,5}^K$  is bounded [59, Sec. 5.6.(12).(b)] and (50) follows from (48). The proof of (20) is completed by substituting (50) into (45) and simplifying.

The proof of the BER expression in (19) follows analogous steps and is omitted for space reasons.

#### APPENDIX D PROOF OF THEOREM 3

Consider a sequence of independent symbols  $X_k = R_k e^{j\Phi_k}$ ,  $k \in \mathbb{Z}$ , where for each  $k$ , the magnitude  $R_k$  is independent of the phase  $\Phi_k$ , which is uniform in  $[0, 2\pi)$ . The magnitude  $R_k$  is distributed according to  $f_R$  if  $k = 0 \pmod{(2N+1)}$  and is otherwise equal to the constant  $r_1$ . Furthermore,  $f_R$  and  $r_1$  are chosen so that (30) holds, which guarantees that the average power constraint is satisfied. We will next show that the right-hand side of (31) is the mutual information (in bits per channel use) obtainable with this input distribution. Hence, it is a lower bound on capacity.

We define blocks of length  $2N+1$  of transmitted and received symbols as

$$\begin{aligned} \mathbf{Y}_l &\triangleq \mathbf{Y}_{l(2N+1)-N}^{l(2N+1)+N}, \\ \mathbf{X}_l &\triangleq \mathbf{X}_{l(2N+1)-N}^{l(2N+1)+N} \end{aligned}$$

for  $l \in \mathbb{Z}$ . Let us focus for a moment on the received block  $\mathbf{Y}_0$ . Let  $Y_k$  be the  $k$ th element ( $k = -N, \dots, N$ ) of  $\mathbf{Y}_0$ . It follows from (8) that the additive noise contribution to  $Y_k$  depends on the input vector  $\|\mathbf{X}_{k-N}^{k+N}\|$ , which may span more than one input block. By construction, however, all elements of  $\mathbf{X}_{k-N}^{k+N}$  with the exception of  $X_0$  have constant magnitude equal to  $r_1$ . Hence,

$$\|\mathbf{X}_{k-N}^{k+N}\|^2 = |X_0|^2 + 2Nr_1^2. \quad (51)$$

This implies that

$$f_{Y_k|\mathbf{X}_{k-N}^{k+N}}(y_k|\mathbf{x}_{k-N}^{k+N}) = \frac{1}{\pi\rho(2Nr_1^2 + |x_0|^2)} \exp\left(-\frac{|y_k - x_k|^2}{\rho(2Nr_1^2 + |x_0|^2)}\right). \quad (52)$$

We see from (52) that each output sample  $Y_k$  in  $\mathbf{Y}_0$  actually depends on the input symbols only through  $X_k$  and  $X_0$ . We then conclude that  $\mathbf{Y}_0$  depends on the whole input sequence only through  $\mathbf{X}_0$ . But this, together with the assumption of independent input symbols, implies that the output blocks  $\{\mathbf{Y}_l\}$  are independent. Hence, from (29),

$$C \geq \frac{1}{2N+1} I(\mathbf{X}_l; \mathbf{Y}_l) \quad (53)$$

for an arbitrary  $l \in \mathbb{Z}$ , say,  $l = 0$ .

Next, we calculate  $I(\mathbf{X}_0; \mathbf{Y}_0)$ . The mutual information can be decomposed into differential entropies as

$$I(\mathbf{X}_0; \mathbf{Y}_0) = h(\mathbf{Y}_0) - h(\mathbf{Y}_0|\mathbf{X}_0), \quad (54)$$

where

$$h(\mathbf{Y}_0) = -\mathbb{E}[\log_2 f_{\mathbf{Y}_0}(\mathbf{Y}_0)], \quad (55)$$

$$h(\mathbf{Y}_0|\mathbf{X}_0) = -\mathbb{E}[\log_2 f_{\mathbf{Y}_0|\mathbf{X}_0}(\mathbf{Y}_0|\mathbf{X}_0)]. \quad (56)$$

We start by evaluating (56). Because of (52), the conditional distribution of  $\mathbf{Y}_0$  given  $\mathbf{X}_0$  is the multivariate Gaussian density

$$f_{\mathbf{Y}_0|\mathbf{X}_0}(\mathbf{y}_0|\mathbf{x}_0) = \frac{1}{(\pi\rho(2Nr_1^2 + |x_0|^2))^{2N+1}} \exp\left(-\frac{\|\mathbf{y}_0 - \mathbf{x}_0\|^2}{\rho(2Nr_1^2 + |x_0|^2)}\right). \quad (57)$$

Using [16, Theorem 8.4.1], we conclude that

$$h(\mathbf{Y}_0|\mathbf{X}_0) = (2N+1)\mathbb{E}[\log_2 \pi\rho(2Nr_1^2 + |X_0|^2)], \quad (58)$$

where the expectation is with respect to the random variable  $|X_0|$ , which is distributed according to  $f_R$ .

To evaluate (55), we start by noting that all elements of  $\mathbf{Y}_0$  have uniform phase because the transmitted symbols and the additive noise samples have uniform phase by assumption. We use this property to simplify (55). Specifically, let  $U_k = |Y_k|^2$  and

$$\mathbf{U} \triangleq [U_{-N}, U_{-N+1}, \dots, U_N]. \quad (59)$$

By [60, eq. (320)]

$$h(\mathbf{Y}_0) = (2N+1) \log_2 \pi + h(\mathbf{U}). \quad (60)$$

To evaluate  $h(\mathbf{U}) = -\mathbb{E}[\log_2(f_{\mathbf{U}}(\mathbf{U}))]$ , we first derive the conditional distribution  $f_{\mathbf{U}||X_0|}$  of  $\mathbf{U}$  given  $|X_0|$ . Note that  $U_k$  has the same distribution as

$$\left| |X_k| + \sqrt{\rho(2Nr_1^2 + |X_0|^2)} \tilde{Z}_k \right|^2$$

(see (1) and (8)). Hence, given  $|X_0| = r$ , the random variables  $\{2U_k/\rho(2Nr_1^2 + r^2)\}$  follow a noncentral chi-square distribution with two degrees of freedom and noncentrality parameters

$\{2|X_k|^2/\rho(2Nr_1^2 + r^2)\}$ , where  $|X_k| = r_1$  if  $k \neq 0$  and  $|X_k| = r$  otherwise. Furthermore, these random variables are conditionally independent given  $|X_0|$ . Using the change of variable theorem for transformation of random variables, we finally obtain after algebraic manipulations

$$f_{\mathbf{U}||X_0|}(\mathbf{u}|r) = \frac{\exp\left(-\frac{\sum_{k=-N}^N u_k + 2Nr_1^2 + r^2}{\rho(2Nr_1^2 + r^2)}\right)}{(\rho(2Nr_1^2 + r^2))^{2N+1}} \cdot I_0\left(\frac{2r\sqrt{u_0}}{\rho(2Nr_1^2 + r^2)}\right) \cdot \prod_{\substack{k=-N \\ k \neq 0}}^N I_0\left(\frac{2r_1\sqrt{u_k}}{\rho(2Nr_1^2 + r^2)}\right). \quad (61)$$

The probability distribution  $f_{\mathbf{U}}$ , which is given in (32), is obtained from (61) by taking the expectation with respect to  $f_R$ , the probability distribution of  $|X_0|$ . Finally, we obtain the capacity lower bound (31) by substituting (32) into (55) and (58) into (56), by computing the difference between the two resulting differential entropies according to (54), and by dividing by  $2N+1$ .

## REFERENCES

- [1] H. Sun, K.-T. Wu, and K. Roberts, "Real-time measurements of a 40 Gb/s coherent system," *Opt. Express*, vol. 16, no. 2, pp. 873–879, 2008.
- [2] K. Roberts, M. O'Sullivan, K.-T. Wu, H. Sun, A. Awadalla, D. J. Krause, and C. Laperle, "Performance of dual-polarization QPSK for optical transport systems," *J. Lightw. Technol.*, vol. 27, no. 16, pp. 3546–3559, Aug. 2009.
- [3] A. D. Ellis, J. Zhao, and D. Cotter, "Approaching the non-linear Shannon limit," *J. Lightw. Technol.*, vol. 28, no. 4, pp. 423–433, Feb. 2010.
- [4] A. Mecozzi and R.-J. Essiambre, "Nonlinear Shannon limit in pseudolinear coherent systems," *J. Lightw. Technol.*, vol. 30, no. 12, pp. 2011–2024, June 2012.
- [5] A. Splett, C. Kurtzke, and K. Petermann, "Ultimate transmission capacity of amplified optical fiber communication systems taking into account fiber nonlinearities," in *Proc. Eur. Conf. Opt. Commun. (ECOC)*, Montreux, Switzerland, Sept. 1993.
- [6] J. Tang, "The channel capacity of a multispan DWDM system employing dispersive nonlinear optical fibers and an ideal coherent optical receiver," *J. Lightw. Technol.*, vol. 20, no. 7, pp. 1095–1101, July 2002.
- [7] 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.
- [8] P. Poggiolini, G. Bosco, A. Carena, V. Curri, Y. Jiang, and F. Forghieri, "The GN-model of fiber non-linear propagation and its applications," *J. Lightw. Technol.*, vol. 32, no. 4, pp. 694–721, Feb. 2014.
- [9] 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.
- [10] 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.
- [11] E. Grellier and A. Bononi, "Quality parameter for coherent transmissions with Gaussian-distributed nonlinear noise," *Opt. Express*, vol. 19, no. 13, pp. 12781–12788, June 2011.
- [12] L. Beygi, N. V. 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. Express*, vol. 21, no. 22, pp. 26376–26386, Nov. 2013.
- [13] F. V. 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," *Opt. Express*, vol. 20, no. 2, pp. 1022–1032, Jan. 2012.



- [14] A. Carena, V. Curri, G. Bosco, P. Poggiolini, and F. Forghieri, "Modeling of the impact of nonlinear propagation effects in uncompensated optical coherent transmission links," *J. Lightw. Technol.*, vol. 30, no. 10, pp. 1524–1539, May 2012.
- [15] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 623–656, July, Oct. 1948.
- [16] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ: Wiley, 2006.
- [17] J. B. Stark, "Fundamental limits of information capacity for optical communications channels," in *Proc. Eur. Conf. Opt. Commun. (ECOC)*, Nice, France, Sep. 1999.
- [18] P. P. Mitra and J. B. Stark, "Nonlinear limits to the information capacity of optical fibre communications," *Nature*, vol. 411, pp. 1027–1030, June 2001.
- [19] 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.
- [20] I. B. Djordjevic and B. Vasic, "Achievable information rates for high-speed long-haul optical transmission," *J. Lightw. Technol.*, vol. 11, no. 23, pp. 3755–3763, Nov. 2005.
- [21] 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.
- [22] 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.
- [23] 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.
- [24] M. Secondini and E. Forestieri, "Analytical fiber-optic channel model in the presence of cross-phase modulation," *IEEE Photon. Technol. Lett.*, vol. 24, no. 22, pp. 2016–2019, Nov. 2012.
- [25] R. Dar, M. Shttaif, and M. Feder, "Information rates in the optical nonlinear phase noise channel," in *Proc. Allerton Conf. Commun., Control, and Comput.*, Monticello, IL, Oct. 2013.
- [26] —, "New bounds on the capacity of the nonlinear fiber-optic channel," *Optics Letters*, vol. 39, no. 2, pp. 398–401, Jan. 2014.
- [27] R. Dar, M. Feder, A. Mecozzi, and M. Shttaif, "Time varying ISI model for nonlinear interference noise," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, San Francisco, CA, Mar. 2014.
- [28] J. M. Kahn and K.-P. Ho, "Spectral efficiency limits and modulation/detection techniques for DWDM systems," *IEEE J. Sel. Top. Quantum Electron.*, vol. 10, no. 2, pp. 259–272, Mar./Apr. 2004.
- [29] E. Narimanov and P. Mitra, "The channel capacity of a fiber optics communication system: Perturbation theory," *J. Lightw. Technol.*, vol. 20, no. 3, pp. 530–537, Mar. 2002.
- [30] L. G. L. Wegener, M. L. Povinelli, A. G. Green, P. P. Mitra, J. B. Stark, and P. B. Littlewood, "The effect of propagation nonlinearities on the information capacity of WDM optical fiber systems: Cross-phase modulation and four-wave mixing," *Physica D: Nonlinear Phenomena*, vol. 189, no. 1–2, pp. 81–99, Feb. 2004.
- [31] R.-J. Essiambre, G. J. Foschini, G. Kramer, and P. J. Winzer, "Capacity limits of information transport in fiber-optic networks," *Phys. Rev. Lett.*, vol. 101, no. 16, pp. 163 901–1–4, Oct. 2008.
- [32] T. Freckmann, R.-J. Essiambre, P. J. Winzer, G. J. Foschini, and G. Kramer, "Fiber capacity limits with optimized ring constellations," *IEEE Photon. Technol. Lett.*, vol. 21, no. 20, pp. 1496–1498, Oct. 2009.
- [33] I. B. Djordjevic, H. G. Batshon, L. Xu, and T. Wang, "Coded polarization-multiplexed iterative polar modulation (PM-IPM) for beyond 400 Gb/s serial optical transmission," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, San Diego, CA, Mar. 2010.
- [34] R. I. Killey and C. Behrens, "Shannon's theory in nonlinear systems," *J. Mod. Opt.*, vol. 58, no. 1, pp. 1–10, Jan. 2011.
- [35] G. Bosco, P. Poggiolini, A. Carena, V. Curri, and F. Forghieri, "Analytical results on channel capacity in uncompensated optical links with coherent detection," *Opt. Express*, vol. 19, no. 26, pp. B440–B449, Dec. 2011.
- [36] E. Agrell and M. Karlsson, "Power-efficient modulation formats in coherent transmission systems," *J. Lightw. Technol.*, vol. 27, no. 22, pp. 5115–5126, Nov. 2009.
- [37] —, "WDM channel capacity and its dependence on multichannel adaptation models," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, Anaheim, CA, Mar. 2013.
- [38] B. Goebel, R.-J. Essiambre, G. Kramer, P. J. Winzer, and N. Hanik, "Calculation of mutual information for partially coherent Gaussian channels with applications to fiber optics," *IEEE Trans. Inf. Theory*, vol. 57, no. 9, pp. 5720–5736, Sep. 2011.
- [39] P. Poggiolini, "The GN model of non-linear propagation in uncompensated coherent optical systems," *J. Lightw. Technol.*, vol. 24, no. 30, pp. 3875–3879, Dec. 2012.
- [40] G. Bosco, P. Poggiolini, A. Carena, V. Curri, and F. Forghieri, "Analytical results on channel capacity in uncompensated optical links with coherent detection: Erratum," *Opt. Express*, vol. 20, no. 17, pp. 19 610–19 611, Aug. 2012.
- [41] E. Agrell and M. Karlsson, "Satellite constellations: Towards the nonlinear channel capacity," in *Proc. IEEE Photon. Conf. (IPC)*, Burlingame, CA, Sept. 2012.
- [42] A. Carena, G. Bosco, V. Curri, P. Poggiolini, M. T. Taiba, and F. Forghieri, "Statistical characterization of PM-QPSK signals after propagation in uncompensated fiber links," in *Proc. Eur. Conf. Opt. Commun. (ECOC)*, London, U.K., Sept. 2010.
- [43] T. Koch, A. Lapidath, and P. Sotiriadis, "Channels that heat up," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3594–3612, Aug. 2009.
- [44] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY: Wiley, 1968.
- [45] E. Ip and J. M. Kahn, "Digital equalization of chromatic dispersion and polarization mode dispersion," *J. Lightw. Technol.*, vol. 25, no. 8, pp. 2033–2043, Aug. 2007.
- [46] G. P. Agrawal, *Fiber-optic communication systems*, 4th ed. Wiley, 2010.
- [47] E. Agrell, J. Lassing, E. G. Ström, and T. Ottosson, "On the optimality of the binary reflected Gray code," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3170–3182, Dec. 2004.
- [48] 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.
- [49] M. P. Fitz and J. P. Seymour, "On the bit error probability of QAM modulation," *Int. J. Wireless Inf. Netw.*, vol. 1, no. 2, pp. 131–139, Apr. 1994.
- [50] M. K. Simon, S. M. Hinedi, and W. C. Lindsey, *Digital Communication Techniques: Signal Design and Detection*. Englewood Cliffs, NJ: Prentice-Hall, 1995.
- [51] 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.
- [52] A. Demir, "Nonlinear phase noise in optical-fiber-communication systems," *J. Lightw. Technol.*, vol. 25, no. 8, pp. 2002–2032, Aug. 2007.
- [53] 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.
- [54] K.-T. Fang, S. Kotz, and K. W. Ng, *Symmetric Multivariate and Related Distributions*. Springer, 1990.
- [55] S. Kotz and S. Nadarajah, *Multivariate t Distributions and Their Applications*. Cambridge University Press, 2004.
- [56] N. M. Blachman, "A comparison of the informational capacities of amplitude- and phase-modulation communication systems," *Proc. I.R.E.*, vol. 41, no. 6, pp. 748–759, June 1953.
- [57] K.-P. Ho and J. M. Kahn, "Channel capacity of WDM systems using constant-intensity modulation formats," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, Anaheim, CA, Mar. 2002.
- [58] E. Agrell, "Conditions for a monotonic channel capacity," 2012, preprint. [Online]. Available: <http://arxiv.org/abs/1209.2820>
- [59] G. R. Grimmett and D. R. Stirzaker, *Probability and Random Processes*, 3rd ed. Oxford University Press, 2001.
- [60] A. Lapidath and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2426–2467, Oct. 2003.

**Erik Agrell** (M'99–SM'02) received the Ph.D. degree in information theory in 1997 from Chalmers University of Technology, Sweden.

From 1997 to 1999, he was a Postdoctoral Researcher with the University of California, San Diego and the University of Illinois at Urbana-Champaign. In 1999, he joined the faculty of Chalmers University of Technology, where he is a Professor in Communication Systems since 2009. In 2010, he cofounded the Fiber-Optic Communications Research Center (FORCE) at Chalmers, where he leads the signals and systems research area. His research interests belong to the fields of information theory, coding theory, and digital communications, and his favorite applications are found in optical communications.

Prof. Agrell served as Publications Editor for the IEEE Transactions on Information Theory from 1999 to 2002 and is an Associate Editor for the IEEE Transactions on Communications since 2012. He is a recipient of the 1990 John Ericsson Medal, the 2009 ITW Best Poster Award, the 2011 GlobeCom Best Paper Award, the 2013 CTW Best Poster Award, and the 2013 Chalmers Supervisor of the Year Award.

**Alex Alvarado** (S'06–M'11) was born in 1982 in Quellón, on the island of Chiloé, Chile. He received his Electronics Engineer degree (Ingeniero Civil Electrónico) and his M.Sc. degree (Magíster en Ciencias de la Ingeniería Electrónica) from Universidad Técnica Federico Santa María, Valparaíso, Chile, in 2003 and 2005, respectively. He obtained the degree of Licentiate of Engineering (Teknologie Licentiatexamen) in 2008 and his PhD degree in 2011, both of them from Chalmers University of Technology, Gothenburg, Sweden.

Dr. Alvarado is currently a Senior Research Associate at the Optical Networks Group, University College London, United Kingdom. In 2012–2014 he was a Marie Curie Intra-European Fellow at the University of Cambridge, United Kingdom, and during 2011–2012 he was a Newton International Fellow at the same institution. In 2008, he was holder of the Merit Scholarship Program for Foreign Students, granted by the Ministère de l'Éducation, du Loisir et du Sports du Québec. His general research interests are in the areas of digital communications, coding, and information theory.

**Giuseppe Durisi** (S'02–M'06–SM'12) received the Laurea degree summa cum laude and the Doctor degree both from Politecnico di Torino, Italy, in 2001 and 2006, respectively. From 2002 to 2006, he was with Istituto Superiore Mario Boella, Torino, Italy. From 2006 to 2010 he was a post-doctoral researcher at ETH Zurich, Switzerland. Since 2010 he has been with Chalmers University of Technology, Gothenburg, Sweden, where is now associate professor. He held visiting researcher positions at IMST, Germany, University of Pisa, Italy, ETH Zurich, Switzerland, and Vienna University of Technology, Austria.

Dr. Durisi is a senior member of the IEEE. He is the recipient of the 2013 IEEE ComSoc Best Young Researcher Award for the Europe, Middle East, and Africa Region, and is co-author of a paper that won a student paper award at the 2012 International Symposium on Information Theory, and of a paper that won the 2013 IEEE Sweden VT-COM-IT joint chapter best student conference paper award. He served as TPC member in several IEEE conferences, and is currently publications editor of the IEEE Transactions on Information Theory. His research interests are in the areas of communication and information theory.

**Magnus Karlsson** received his Ph. D in 1994 from Chalmers University of Technology, Gothenburg, Sweden. Since 1995, he has been with the Photonics Laboratory at Chalmers, first as assistant professor and since 2003 as professor in photonics. He has authored or co-authored around 300 scientific journal and conference contributions in the areas of nonlinear optics and fiber optic transmission. He co-founded the Chalmers fiber-optic communication research center FORCE in 2010.

He has served in the technical committee for the Optical Fiber Communication Conference (OFC), and currently serves in the technical program committees for the European Conference of Optical Communication (ECOC), and the Asia Communications and Photonics Conference (ACP). He is associate editor for Optics Express since 2010, and has served as guest editor for Journal of Lightwave Technology.

He contributed in the CELTIC 100-GET and EO-Net projects that were both awarded the CELTIC excellence award. He supervised students receiving the best paper awards at GlobeCom 2011 and at ECOC 2012. He is senior member of IEEE and was appointed Fellow of the Optical Society of America in 2012.