



## **Accuracy Assessment of Nondispersive Optical Perturbative Models through Capacity Analysis**

Downloaded from: <https://research.chalmers.se>, 2024-04-25 08:50 UTC


Citation for the original published paper (version of record):

Keykhosravi, K., Durisi, G., Agrell, E. (2019). Accuracy Assessment of Nondispersive Optical Perturbative Models through Capacity Analysis. *Entropy*, 21(8). <http://dx.doi.org/10.3390/e21080760>

N.B. When citing this work, cite the original published paper.

## Article

# Accuracy Assessment of Nondispersive Optical Perturbative Models through Capacity Analysis

Kamran Keykhosravi , Giuseppe Durisi and Erik Agrell \*

Department of Electrical Engineering, Chalmers University of Technology, 41296 Gothenburg, Sweden

\* Correspondence: [agrell@chalmers.se](mailto:agrell@chalmers.se)

Received: 20 May 2019; Accepted: 28 July 2019; Published: 5 August 2019

**Abstract:** A number of simplified models, based on perturbation theory, have been proposed for the fiber-optical channel and have been extensively used in the literature. Although these models are mainly developed for the low-power regime, they are used at moderate or high powers as well. It remains unclear to what extent the capacity of these models is affected by the simplifying assumptions under which they are derived. In this paper, we consider single-channel data transmission based on three continuous-time optical models: (i) a regular perturbative channel, (ii) a logarithmic perturbative channel, and (iii) the stochastic nonlinear Schrödinger (NLS) channel. To obtain analytically tractable discrete-time models, we consider zero-dispersion fibers and a sampling receiver. We investigate the per-sample capacity of these models. Specifically, (i) we establish tight bounds on the capacity of the regular perturbative channel; (ii) we obtain the capacity of the logarithmic perturbative channel; and (iii) we present a novel upper bound on the capacity of the zero-dispersion NLS channel. Our results illustrate that the capacity of these models departs from each other at high powers because these models yield different capacity pre-logs. Since all three models are based on the same physical channel, our results highlight that care must be exercised in using simplified channel models in the high-power regime.

**Keywords:** achievable rate; channel capacity; information theory; nonlinear channel; optical fiber

## 1. Introduction

The vast majority of the global Internet traffic is conveyed through fiber-optical networks, which form the backbone of our information society. To cope with the growing data demand, the fiber-optical networks have evolved from regenerated direct-detection systems to coherent wavelength division multiplexing (WDM) ones. Newly emerging bandwidth-hungry services, like Internet-of-Things (IoT) applications and cloud processing, require even higher data rates. Motivated by this ever-growing demand, an increasing attention has been devoted in recent years to the analysis of the capacity of the fiber-optical channel.

Finding the capacity of the fiber-optical channel that is governed by the stochastic nonlinear Schrödinger (NLS) equation ([1], Equation (1)), which captures the effects of Kerr nonlinearity, chromatic dispersion, and amplification noise, remains an open problem. An information-theoretic analysis of the NLS channel is cumbersome because of a complicated signal–noise interaction caused by the interplay between the nonlinearity and the dispersion [2]. In general, capacity analyses of optical fibers are performed either by considering simplified channels, or by evaluating mismatched decoding lower bounds [3] via simulations (see [4] and ([5], Sec. I) for excellent literature reviews). Lower bounds based on the mismatch-decoding framework go to zero after reaching a maximum (see, for example, [2,6–9]). Capacity lower bounds with a similar behavior are also reported in [10]. In [11], it has been shown that the maximum value of a capacity lower bound can be increased by increasing fiber dispersion, which mitigates the effects of nonlinearity. To establish a capacity upper

bound, Kramer et al. [12] used the split-step Fourier (SSF) method, which is a standard approach to solve the NLS equation numerically ([13], Sec. 2.4.1), to derive a discrete-time channel model. They proved that the capacity of this discrete-time model is upper-bounded by that of an equivalent additive white Gaussian noise (AWGN) channel. In contrast to the available lower bounds, which fall to zero or saturate at high powers, this upper bound, which is the only one available for a realistic fiber channel model, grows unboundedly.

Since the information-theoretic analysis of the NLS channel is difficult, to approximate capacity, one can resort to simplified models, a number of which have been studied in the literature (see [14] and references therein for a recent review). Two approaches to obtain such models are to use the regular perturbation or the logarithmic perturbation methods. In the former, the effects of nonlinearity are captured by an additive perturbative term [15,16]. This approach yields a discrete-time channel with input–output relation  $\mathbf{y}_l = \mathbf{x}_l + \Delta\mathbf{x}_l + \mathbf{n}_l$  ([14], Equation (5)), where  $\mathbf{x}_l$  and  $\mathbf{y}_l$  are the transmitted and the received symbols, respectively;  $\mathbf{n}_l$  is the amplification noise; and  $\Delta\mathbf{x}_l$  is the perturbative nonlinear distortion. This model holds under the simplifying assumption that both the nonlinearity and the signal–noise interaction are weak, which is reasonable only at low power.

Regular perturbative fiber-optical channel models, with or without memory, have been extensively investigated in the literature. In [17], a first-order perturbative model for WDM systems with arbitrary filtering and sampling demodulation, and coherent detection is proposed. The accuracy of the model is assessed by comparing the value of a mismatch-decoding lower bound, which is derived analytically based on the perturbative model, with simulation results over a realistic fiber-optical channel. A good agreement at all power levels is observed. The capacity of a perturbative multiple-access channel is studied in [18]. It is shown that the nonlinear crosstalk between channels does not affect the capacity region when the information from all the channels is optimally used at each detector. However, if joint processing is not possible (it is typically computationally demanding [19]), the channel capacity is limited by the inter-channel distortion.

Another class of simplified models, which are equivalent to the regular perturbative ones up to a first-order linearization, is that of logarithmic perturbative models, where the nonlinear distortion term  $\Delta\mathbf{x}_l$  is modeled as a phase shift. This yields a discrete-time channel with input–output relation  $\mathbf{y}_l = \mathbf{x}_l e^{j\Delta\mathbf{x}_l} + \mathbf{n}_l$  ([14], Equation (7)). In [5], a single-span optical channel model for a two-user WDM transmission system is developed from a coupled NLS equation, neglecting the dispersion effects within the WDM bands. The channel model in [5] resembles the perturbative logarithmic models. The authors study the capacity region of this channel in the high-power regime. It is shown that the capacity pre-log pair (1,1) is achievable, where the capacity pre-log is defined as the asymptotic limit of  $\mathcal{C} / \log P$  for  $P \rightarrow \infty$ , where  $P$  is the input power and  $\mathcal{C}$  is the capacity.

Despite the fact that the aforementioned simplified channels are valid in the low-power regime, these models are often used also in the moderate- and high-power regimes. Currently, it is unclear to what extent the simplifications used to obtain these models influence the capacity at high powers. To find out, we study the capacity of two single-channel memoryless perturbative models, namely, a *regular perturbative channel (RPC)*, and a *logarithmic perturbative channel (LPC)*. To assess accuracy of these two perturbative models, we also investigate the per-sample capacity of a *memoryless NLS channel (MNC)*.

To enable an information-theoretic analysis of the fiber-optical channel, we deploy two common assumptions on the channel model. First, the dispersion is set to zero and second, a sampling receiver is used to obtain discrete-time models from continuous-time channels. These two assumptions were first applied to the NLS equation in [1] to obtain an analytically tractable channel model. This channel model was developed also in [20–22] using different methods. In this paper, we refer to this model as MNC.

In [21], a lower bound on the per-sample capacity of the memoryless NLS channel is derived, which proves that the capacity goes to infinity with power. In [22], the capacity of the same channel is evaluated numerically. Furthermore, it is shown that the capacity pre-log is 1/2. Approximations

of the capacity and optimal input distribution in the intermediate power range are derived in [23,24]. These results are extended to a channel with a more realistic receiver than the sampling one in [25]. The only known nonasymptotic upper bound on the capacity of this channel is  $\log(1 + \text{SNR})$  (bits per channel use) [12], where SNR is the signal-to-noise ratio. This upper bound holds also for the general case of nonzero dispersion.

The novel contributions of this paper are as follows. First, we tightly bound the capacity of the RPC model and prove that its capacity pre-log is 3. Second, the capacity of the LPC is readily shown to be the same as that of an AWGN channel with the same input and noise power. Hence, the capacity pre-log of the LPC is 1. Third, we establish a novel upper bound on the capacity of the MNC (first presented in the conference version of this manuscript [26]). Our upper bound improves the previously known upper bound [12] on the capacity of this channel significantly and, together with the proposed lower bound, allows one to characterize the capacity of the MNC accurately.

Although all three models represent the same physical optical channel, their capacities behave very differently in the high-power regime. This result highlights the profound impact of the simplifying assumptions on the capacity at high powers, and indicates that care should be taken in translating the results obtained based on these models into guidelines for system design.

The rest of this paper is organized as follows. In Section 2, we introduce the three channel models. In Section 3, we present upper and lower bounds on the capacity of these channels and establish the capacity pre-log of the perturbative models. Numerical results are provided in Section 4. We conclude the paper in Section 5. The proofs of all theorems are given in the appendices.

*Notation:* Random quantities are denoted by boldface letters. We use  $\mathcal{CN}(0, \sigma^2)$  to denote the complex zero-mean circularly symmetric Gaussian distribution with variance  $\sigma^2$ . We write  $\mathcal{R}(x)$ ,  $|x|$ , and  $\angle x$  to denote the real part, the absolute value, and the phase of a complex number  $x$ . All logarithms are in base two. The mutual information between two random variables  $\mathbf{x}$  and  $\mathbf{y}$  is denoted by  $I(\mathbf{x}; \mathbf{y})$ . The entropy and differential entropy are denoted by  $H(\cdot)$  and  $h(\cdot)$ , respectively. Finally, we use  $*$  for the convolution operator.

## 2. Channel Models

The fiber-optical channel is well-modeled by the NLS equation, which describes the propagation of a complex baseband electromagnetic field through a lossy single-mode fiber as

$$\frac{\partial \mathbf{a}}{\partial z} + \frac{\alpha - g}{2} \mathbf{a} + j \frac{\beta_2}{2} \frac{\partial^2 \mathbf{a}}{\partial t^2} - j \gamma |\mathbf{a}|^2 \mathbf{a} = \mathbf{n}. \quad (1)$$

Here,  $\mathbf{a} = \mathbf{a}(z, t)$  is the complex baseband signal at time  $t$  and location  $z$ . The parameter  $\gamma$  is the nonlinear coefficient,  $\beta_2$  is the group-velocity dispersion parameter,  $\alpha$  is the attenuation constant,  $g = g(z)$  is the gain profile of the amplifier, and  $\mathbf{n} = \mathbf{n}(z, t)$  is the Gaussian amplification noise, which is bandlimited because of the inline channel filters. The third term on the left-hand side of (1) is responsible for the channel memory and the fourth term for the channel nonlinearity.

To compensate for the fiber losses, two types of signal amplification can be deployed, namely, distributed and lumped amplification. The former method compensates for the fiber loss continuously along the fiber, whereas the latter method boosts the signal power by dividing the fiber into several spans and using an optical amplifier at the end of each span. With distributed amplification, which we focus on in this paper, the noise can be described by the autocorrelation function [2]

$$\mathbb{E}[\mathbf{n}(z, t) \mathbf{n}^*(z', t')] = \alpha n_{\text{sp}} h \nu \delta_{W_N}(t - t') \delta(z - z'). \quad (2)$$

Here,  $n_{\text{sp}}$  is the spontaneous emission factor,  $h$  is Planck's constant, and  $\nu$  is the optical carrier frequency. In addition,  $\delta(\cdot)$  is the Dirac delta function and  $\delta_{W_N}(x) = W_N \text{sinc}(W_N x)$ , where  $W_N$  is the noise bandwidth. In this paper, we shall focus on the ideal distributed-amplification case  $g(z) = \alpha$ .

We use a sampling receiver to go from continuous-time channels to discrete-time ones. A comprehensive description of the sampling receiver and of the induced discrete-time channel is provided in ([22], Section III). Here, we review some crucial elements of this description for completeness. Assume that a signal  $\mathbf{a}(0, t)$ , which is band-limited to  $W_0$  hertz, is transmitted through a zero-dispersion NLS channel ((1) with  $\beta_2 = 0$ ) in the time interval  $[0, T]$ . Because of nonlinearity, the bandwidth of the received signal  $\mathbf{a}(L, t)$  may be larger than that of  $\mathbf{a}(0, t)$ . To avoid signal distortion by the inline filters, we assume that  $W_0$  is set such that  $\mathbf{a}(z, t)$  is band-limited to  $W_N$  hertz for  $0 \leq z \leq L$ . Since  $W_0 \leq W_N$ , assuming  $W_N T \gg 1$ , both the transmitted and the received signal can be represented by  $2W_N T$  equispaced samples. The transmitter encodes data into subsets of these samples of cardinality  $2W_0 T$ , referred to as the principal samples. At the receiver, demodulation is performed by sampling  $\mathbf{a}(L, t)$  at instances corresponding to the principal samples. This results in  $2W_0 T$  parallel independent discrete-time channels that have the same input–output relation.

The sampling receiver has a number of shortcomings [27] and using it should be considered a simplification. The resulting discrete-time model is used extensively in the literature (see, for example, [1,20–22,28,29]), since it makes analytical calculation possible. In this paper, we apply the sampling receiver not only to the memoryless NLS channel but also to the memoryless perturbative models.

Next, we review two perturbative channel models that are used in the literature to approximate the solution of the NLS Equation (1). Among the multiple variations of perturbative models available in the literature, we use the ones proposed in [30]. For both perturbative models, first continuous-time dispersive models are introduced, and then memoryless discrete-time channels are developed by assuming that  $\beta_2 = 0$  and by using a sampling receiver. Finally, we introduce the MNC model, which is derived from (1) under the two above-mentioned assumptions.

*Regular perturbative channel (RPC):* Let  $\mathbf{a}^{\text{li}}(z, t)$  be the solution of the linear noiseless NLS equation (Equation (1) with  $\mathbf{n}(z, t) = 0$  and  $\gamma = 0$ ). It can be computed as  $\mathbf{a}^{\text{li}}(z, t) = \mathbf{a}(0, t) * h(z, t)$ , where  $h(z, t) = \mathcal{F}^{-1} \{ \exp(j\beta_2 \omega^2 z / 2) \}$  and  $\mathcal{F}^{-1}(\cdot)$  denotes the inverse Fourier transform. In the regular perturbation method, the output of the noiseless NLS channel (Equation (1) with  $\mathbf{n} = 0$ ) is approximated as

$$\mathbf{a}(L, t) = \mathbf{a}^{\text{li}}(L, t) + \Delta \mathbf{a}(L, t). \quad (3)$$

Here,  $L$  is the fiber length and  $\Delta \mathbf{a}(z, t)$  is the nonlinear perturbation term. If now the model is expanded to include amplification noise as an additive noise component, neglecting signal–noise interactions, then the accumulated amplification noise

$$\mathbf{w}(L, t) = \int_0^L \mathbf{n}(z, t) dz \quad (4)$$

can be added to the signal at the receiver to obtain the channel model ([14], Equation (5))

$$\mathbf{a}(L, t) = \mathbf{a}^{\text{li}}(L, t) + \Delta \mathbf{a}(L, t) + \mathbf{w}(L, t). \quad (5)$$

The first-order approximation of  $\Delta \mathbf{a}(L, t)$  is ([30], Equation (13))

$$\Delta \mathbf{a}(L, t) = j\gamma \int_0^L \left[ \left| \mathbf{a}^{\text{li}}(\zeta, t) \right|^2 \mathbf{a}^{\text{li}}(\zeta, t) \right] * h(L - \zeta, t) d\zeta, \quad (6)$$

where the convolution is over the time variable. (Using higher-order nonlinear terms improves the accuracy of the regular perturbative channels. However, in this paper, we focus only on the channel model based on the first-order approximation, which is commonly used in the literature.) Neglecting dispersion (i.e., setting  $\beta_2 = 0$ ), we have  $h(z, t) = \delta(t)$  and  $\mathbf{a}^{\text{li}}(\zeta, t) = \mathbf{a}(0, t)$ . Using this in (6), and then substituting (6) into (5), we obtain

$$\mathbf{a}(L, t) = \mathbf{a}(0, t) + jL\gamma |\mathbf{a}(0, t)|^2 \mathbf{a}(0, t) + \mathbf{w}(L, t). \quad (7)$$

Finally, by deploying sampling receiver, we obtain from (7) the discrete-time channel model

$$\mathbf{y} = \mathbf{x} + j\eta|\mathbf{x}|^2\mathbf{x} + \mathbf{n}. \quad (8)$$

Here,  $\mathbf{n} \sim \mathcal{CN}(0, P_N)$ ,

$$P_N = 2\alpha n_{\text{sp}} h\nu L W_N \quad (9)$$

is the total noise power, and

$$\eta = \gamma L. \quad (10)$$

We refer to (8) as the RPC.

*Logarithmic perturbative channel (LPC):* Another method for approximating the solution of the NLS Equation (1) is to use logarithmic perturbation. With this method, the output signal is approximated as ([14], Equation (7))

$$\mathbf{a}(L, t) = \mathbf{a}(0, t) \exp(j\Delta\theta(L, t)) + \mathbf{w}(L, t), \quad (11)$$

where  $\mathbf{w}(L, t)$  is the same noise term as in (5)–(4). The first-order approximation of  $\Delta\theta(L, t)$  is ([30], Equation (19))

$$\Delta\theta(L, t) = \frac{\gamma}{\mathbf{a}^{\text{li}}(L, t)} \int_0^L \left[ \left| \mathbf{a}^{\text{li}}(\zeta, t) \right|^2 \mathbf{a}^{\text{li}}(\zeta, t) \right] * h(L - \zeta, t) d\zeta. \quad (12)$$

Under the zero-dispersion assumption ( $\beta_2 = 0$ ), we have  $h(z, t) = \delta(t)$  and  $\mathbf{a}^{\text{li}}(\zeta, t) = \mathbf{a}(0, t)$ . Using this in (12), and then substituting (12) into (11), we obtain

$$\mathbf{a}(L, t) = \mathbf{a}(0, t) e^{j\gamma L |\mathbf{a}(0, t)|^2} + \mathbf{w}(L, t). \quad (13)$$

Finally, by sampling the output signal, the discrete-time channel

$$\mathbf{y} = \mathbf{x} e^{j\eta|\mathbf{x}|^2} + \mathbf{n} \quad (14)$$

is obtained, where  $\mathbf{n} \sim \mathcal{CN}(0, P_N)$ ,  $P_N$  is given in (9), and  $\eta$  is defined in (10). We note that the channels (8) and (14) are equal up to a first-order linearization, which is accurate in the low-power regime. Furthermore, one may also obtain the model in (13) by solving (1) for  $\beta_2 = 0$ ,  $\mathbf{n} = 0$ , and  $g = \alpha$  and by adding the noise at the receiver.

*Memoryless NLS Channel (MNC):* Here, we shall study the underlying NLS channel in (1) under the assumptions that  $\beta_2 = 0$  and that a sampling receiver is used to obtain a discrete-time channel. Let  $\mathbf{r}_0$  and  $\theta_0$  be the amplitude and the phase of a transmitted symbol  $\mathbf{x}$ , and let  $\mathbf{r}$  and  $\theta$  be those of the received samples  $\mathbf{y}$ . The discrete-time channel input–output relation can be described by the conditional probability density function (pdf) ([20], Ch. 5) (see also ([28], Sec. II))

$$f_{\mathbf{r}, \theta | \mathbf{r}_0, \theta_0}(r, \theta | r_0, \theta_0) = \frac{f_{\mathbf{r} | \mathbf{r}_0}(r | r_0)}{2\pi} + \frac{1}{\pi} \sum_{m=1}^{\infty} \mathcal{R} \left( C_m(r) e^{-jm(\theta - \theta_0)} \right). \quad (15)$$

The conditional pdf  $f_{\mathbf{r} | \mathbf{r}_0}(r | r_0)$  and the Fourier coefficients  $C_m(r)$  in (15) are given by

$$f_{\mathbf{r} | \mathbf{r}_0}(r | r_0) = \frac{2r}{P_N} \exp \left( -\frac{r^2 + r_0^2}{P_N} \right) I_0 \left( \frac{2rr_0}{P_N} \right), \quad (16)$$

$$C_m(r) = 2rv_m \exp \left( -\left( r^2 + r_0^2 \right) v_m \cos x_m \right) I_m(2rr_0 v_m). \quad (17)$$

Here,  $I_m(\cdot)$  denotes the  $m$ th order modified Bessel function of the first kind, and

$$x_m = \left( \frac{2jm\gamma r_0^2 P_N L}{2r_0^2 + P_N} \right)^{1/2}, \quad (18)$$

$$v_m = \frac{x_m}{P_N \sin x_m}. \quad (19)$$

The complex square root in (18) is a two-valued function, but both choices give the same values of  $v_m$  and  $C_m(r)$ .

In the next section, we study the capacity of the channel models given in (8), (14), and (15). Since all of these models are memoryless, their capacities under a power constraint  $P$  are given by

$$\mathcal{C} = \sup I(\mathbf{x}; \mathbf{y}), \quad (20)$$

where the supremum is over all complex probability distributions of  $\mathbf{x}$  that satisfy the average-power constraint

$$\mathbb{E}[|\mathbf{x}|^2] \leq P. \quad (21)$$

### 3. Analytical Results

In this section, we study the capacity of the RPC, the LPC, and the MNC models. All these models are based on the same fiber-optical channel and share the same set of parameters. Bounds on the capacity of the RPC in (8) are provided in Theorems 1–3. Specifically, in Theorem 1, we establish a closed-form lower bound, which, together with the upper bound provided in Theorem 2, tightly bounds capacity (see Section 4). A different upper bound is provided in Theorem 3. Numerical evidence suggests that this alternative bound is less tight than the one provided in Theorem 2 (see Section 4). However, this alternative bound has a simple analytical form, which makes it easier to characterize it asymptotically. By using the bounds derived in Theorems 1 and 3, we prove that the capacity pre-log of the RPC is 3. In Theorem 4, we derive the capacity of the LPC in (14) and show that it coincides with the capacity of an equivalent AWGN channel. Hence, the capacity pre-log is 1. Theorem 4 is rather trivial and we present it in this section only for completeness. Finally, in Theorem 5, we provide an upper bound on the capacity of the MNC in (15), which improves the previous known upper bound [12] significantly, and, together with a proposed capacity lower bound, yields a tight characterization of capacity (see Section 4).

#### 3.1. Capacity Analysis of the RPC

**Theorem 1.** The capacity  $\mathcal{C}_{\text{RPC}}$  of the RPC in (8) is lower-bounded by

$$\mathcal{C}_{\text{RPC}} \geq \mathcal{L}_{\text{RPC}}(P) = \max_{\lambda} \left\{ \log \left( \frac{\lambda^2 + 6\eta^2}{\lambda^3 P_N} e^{\frac{12\eta^2}{\lambda^2 + 6\eta^2}} + 1 \right) \right\}, \quad (22)$$

where  $\lambda$  is positive and satisfies the constraint

$$\frac{18\eta^2 + \lambda^2}{\lambda(6\eta^2 + \lambda^2)} \leq P. \quad (23)$$

Furthermore, the maximum in (22) is achieved by the unique real solution of the equation

$$P\lambda^3 - \lambda^2 + 6P\eta^2\lambda - 18\eta^2 = 0. \quad (24)$$

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



**Theorem 2.** The capacity of the RPC in (8) is upper-bounded by

$$\mathcal{C}_{\text{RPC}} \leq \mathcal{U}_{\text{RPC}}(P) = \min_{\mu > 0, \lambda > 0} \left\{ \log \frac{\mu^2 + 6\eta^2}{\mu^3 e P_N} + \lambda + \max_{s > 0} \left\{ \mu \mathbb{E} \left[ q(|\mathbf{y}|^2) \mid |\mathbf{x}|^2 = s \right] \log e - \lambda \frac{s + P_N}{P + P_N} \right\} \right\}. \quad (25)$$

Here,  $q(x) = g^{-1}(x)$ , where  $g(x) = x + \eta^2 x^3$ .

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

Note that, given  $|\mathbf{x}|^2 = s$ , the random variable  $2|\mathbf{y}|^2/P_N$  is conditionally distributed as a noncentral chi-squared random variable with two degrees of freedom and noncentrality parameter  $2(s + \eta^2 s^3)/P_N$ . This enables numerical computation of  $\mathcal{U}_{\text{RPC}}(P)$ .

**Theorem 3.** The capacity of the RPC in (8) is upper-bounded by

$$\mathcal{C}_{\text{RPC}} \leq \tilde{\mathcal{U}}_{\text{RPC}}(P) = \min_{\mu > 0} \left\{ \log \left( \frac{\mu^2 + 6\eta^2}{\mu^3 e P_N} \right) + \mu(P + B) \log e \right\}, \quad (26)$$

where

$$B = P_N + \frac{\sqrt{\pi P_N}}{12^{3/8} \sqrt{(\sqrt{3} - 1)\eta}}. \quad (27)$$

Furthermore, the minimum in (26) is achieved by the unique real solution of the equation

$$(P + B) \mu^3 - \mu^2 + 6\eta^2(P + B) \mu - 18\eta^2 = 0. \quad (28)$$

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

*Pre-log analysis:* By substituting  $\mu = 1/P$  into (26), we see that

$$\lim_{P \rightarrow \infty} [\mathcal{C}_{\text{RPC}} - 3 \log(P)] \leq \log \left( \frac{6\eta^2}{P_N} \right). \quad (29)$$

Furthermore, since

$$\frac{18\eta^2 + \lambda^2}{\lambda(6\eta^2 + \lambda^2)} \leq \frac{18\eta^2 + 3\lambda^2}{\lambda(6\eta^2 + \lambda^2)} \quad (30)$$

$$= \frac{3}{\lambda}, \quad (31)$$

we can obtain a valid lower bound on  $\mathcal{C}_{\text{RPC}}$  by substituting  $\lambda = 3/P$  into (22). Doing so, we obtain

$$\lim_{P \rightarrow \infty} [\mathcal{C}_{\text{RPC}} - 3 \log(P)] \geq \log \left( \frac{2\eta^2 e^2}{9P_N} \right). \quad (32)$$

It follows from (29) and (32) that the capacity pre-log of the RPC is 3.

### 3.2. Capacity Analysis of the LPC

**Theorem 4.** The capacity of the LPC in (14) is

$$\mathcal{C}_{\text{LPC}} = \log \left( 1 + \frac{P}{P_N} \right). \quad (33)$$



**Proof.** We use the maximum differential entropy lemma ([31], Sec. 2.2) to upper-bound  $\mathcal{C}_{\text{LPC}}$  by  $\log(1 + P/P_N)$ . Then, we note that we can achieve this upper bound by choosing  $\mathbf{x} \sim \mathcal{CN}(0, P)$ . Alternatively, this theorem can be readily proved by applying a preprocessing step  $\tilde{\mathbf{x}} = \mathbf{x} \exp(-j\eta|\mathbf{x}|^2)$  to the input and transferring the channel to a linear AWGN channel with the same power constraint, whose capacity is  $\log(1 + P/P_N)$ .  $\square$

### 3.3. Capacity Analysis of the MNC

A novel upper bound on the capacity of the MNC in (15) is presented in the following theorem [26].

**Theorem 5.** The capacity of the MNC in (15) is upper-bounded by

$$\mathcal{C}_{\text{MNC}} \leq U_{\text{MNC}}(P) \quad (34)$$

$$= \min_{\lambda > 0, \alpha > 0} \left\{ \alpha \log \left( \frac{P + P_N}{\alpha} \right) + \log(\pi \Gamma(\alpha)) + \lambda + \max_{r_0 > 0} \{g_{\lambda, \alpha}(r_0, P)\} \right\}, \quad (35)$$

where  $\Gamma(\cdot)$  denotes the Gamma function and

$$\begin{aligned} g_{\lambda, \alpha}(r_0, P) = & (\alpha \log e - \lambda) \frac{r_0^2 + P_N}{P + P_N} + (1 - 2\alpha) \mathbb{E}[\log(\mathbf{r}) \mid \mathbf{r}_0 = r_0] \\ & - h(\mathbf{r} \mid \mathbf{r}_0 = r_0) - h(\boldsymbol{\theta} \mid \mathbf{r}, \mathbf{r}_0 = r_0, \boldsymbol{\theta}_0 = 0). \end{aligned} \quad (36)$$

The upper bound  $U_{\text{MNC}}(P)$  can be calculated numerically using the expression for the conditional pdf  $f_{\mathbf{r}, \boldsymbol{\theta} \mid \mathbf{r}_0, \boldsymbol{\theta}_0}(r, \boldsymbol{\theta} \mid r_0, \boldsymbol{\theta}_0)$  given in (15).

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

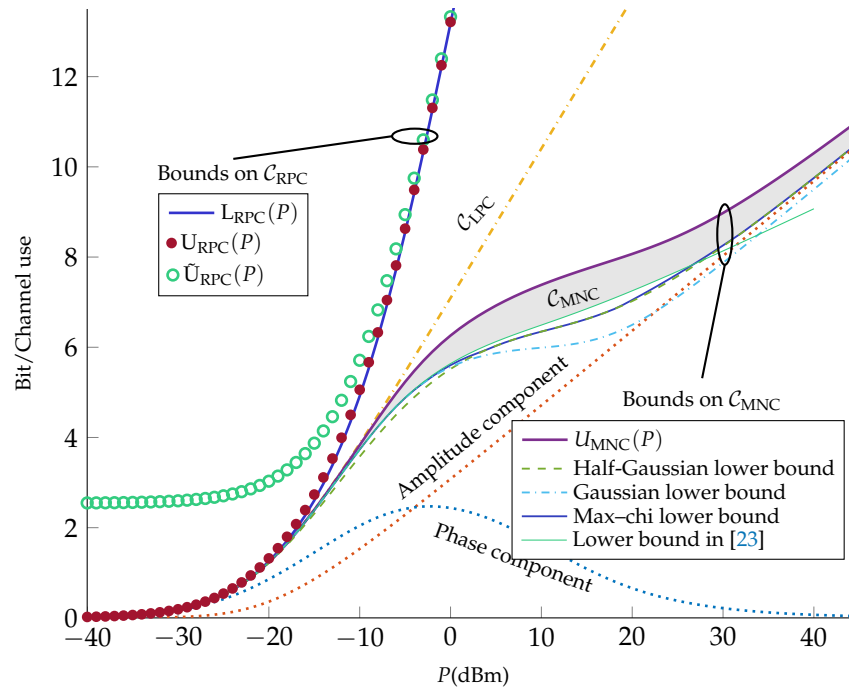
## 4. Numerical Examples

In Figure 1, we evaluate the bounds derived in Section 3 for a fiber-optical channel whose parameters are listed in Table 1. (The channel parameters are the same as in ([22], Table I).) Using (10), we obtain  $\eta = 6350 \text{ W}^{-1}$ .

As can be seen from Figure 1, the capacity of the RPC is tightly bounded between the upper bound  $U_{\text{RPC}}(P)$  in (25) and the lower bound  $L_{\text{RPC}}(P)$  in (22). Furthermore, one can observe that although the alternative upper bound  $\tilde{U}_{\text{RPC}}(P)$  in (26) is loose at low powers, it becomes tight in the moderate- and high-power regimes.

**Table 1.** Channel parameters.

Parameter	Symbol	Value
Attenuation	$\alpha$	0.2 dB/km
Nonlinearity	$\gamma$	$1.27 \text{ (W} \cdot \text{km)}^{-1}$
Fiber length	$L$	5000 km
Maximum bandwidth	$W_N$	125 GHz
Emission factor	$n_{\text{sp}}$	1
Photon energy	$h\nu$	$1.28 \times 10^{-19} \text{ J}$
Noise variance	$P_N$	−21.3 dBm



**Figure 1.** Capacity bounds for the RPC in (8) and the MNC in (15), together with the capacity of the LPC in (14). The amplitude and the phase components of the half-Gaussian lower bound for the MNC are also plotted.

We also plot the upper bound  $U_{\text{MNC}}(P)$  on the capacity of the MNC. It can be seen that  $U_{\text{MNC}}(P)$  improves substantially on the upper bound given in [12], i.e., the capacity of the corresponding AWGN channel (33) (which coincides with  $C_{\text{LPC}}$ ). As a lower bound on the MNC capacity, we propose the mutual information in (20) with an input  $\mathbf{x}$  with uniform phase and amplitude  $r_0$  following a chi distribution with  $k$  degrees of freedom. Specifically, we set

$$f_{r_0}(r_0) = \frac{2r_0^{k-1}}{\Gamma(k/2)} \left(\frac{k}{2P}\right)^{k/2} \exp\left(-\frac{kr_0^2}{2P}\right), \quad (37)$$

where  $\Gamma(\cdot)$  denotes the gamma function. The parameter  $k$  is optimized for each power. (Due to the computational complexity, we only considered  $k$  values from 0.5 to 2.5 in steps of 0.5.) We calculated the bound numerically and include it in Figure 1 (referred to as max-chi lower bound). We also include two lower bounds corresponding to  $k = 1$  (with half-Gaussian amplitude distribution, first presented in [22]) and  $k = 2$  (with Rayleigh-distributed amplitude, or equivalently, a complex Gaussian input  $\mathbf{x}$ , first presented in [26]). The max-chi lower bound coincides with these two lower bounds at asymptotically low and high power, and improves slightly thereon at intermediate powers (around 0 dBm), similarly to the numerical bound in [32]. Specifically, at asymptotically low powers,  $k = 2$  (Gaussian lower bound) is optimal. This is expected, since the channel is essentially linear at low powers. At high powers, on the other hand, the optimal  $k$  value approaches 1 (half-Gaussian lower bound), which is consistent with [22], where it has been shown that half-Gaussian amplitude distribution is capacity-achieving for the MNC in the high-power regime. Based on our numerical evaluations, we observed that  $k = 0.5$  maximizes the max-chi lower bound (among the set of considered values of  $k$ ) in the power range  $18 \leq P \leq 32$  dBm. Finally, in Figure 1, we plot the lower bound based on the input distribution ([23], Equation (45)). As can be seen, based on our numerical evaluation, this lower bound almost coincides with the max-chi lower bound at low powers ( $P < -5$ ) and improves on it in the intermediate power range ( $-5 \leq P < 27$  dBm); however, it is suboptimal at high powers ( $P \geq 27$  dBm).

Figure 1 suggests that  $C_{\text{MNC}}$  experiences changes in slope at about 0 and 30 dBm (corresponding to the inflection points at about  $-10$  dBm and 20 dBm). To explain this behavior, we evaluate the phase

and the amplitude components of the half-Gaussian lower bound. Specifically, we split the mutual information into two parts as

$$I(\mathbf{x}; \mathbf{y}) = I(\mathbf{r}_0, \boldsymbol{\theta}_0; \mathbf{r}, \boldsymbol{\theta}) \quad (38)$$

$$= I(\mathbf{r}_0, \boldsymbol{\theta}_0; \mathbf{r}) + I(\mathbf{r}_0, \boldsymbol{\theta}_0; \boldsymbol{\theta} \mid \mathbf{r}). \quad (39)$$

The first term in (39) is the amplitude component and the second term is the phase component of the mutual information. These two components are evaluated for the half-Gaussian amplitude distribution and plotted in Figure 1. It can be seen from Figure 1 that the amplitude component is monotonically increasing with power while the phase component goes to zero with power after reaching a maximum. Indeed, at high powers, the phase of the received signal becomes uniformly distributed over  $[0, 2\pi]$  and independent of the transmitted signal ([33], Lem. 5). By adding these two components, one obtains a capacity lower bound that changes concavity at two points. The reduction of the capacity slope at intermediate powers is consistent with [23,24], where it is shown that the capacity grows according to  $\log \log P$  in this regime.

As a final observation, we note that  $\mathcal{C}_{\text{RPC}}$  diverges from  $\mathcal{C}_{\text{MNC}}$  at about  $-15$  dBm, whereas  $\mathcal{C}_{\text{LPC}}$  diverges from  $\mathcal{C}_{\text{MNC}}$  at about  $-5$  dBm. Since the MNC describes the nondispersive NLS channel more accurately than the other two channels, this result suggests that the perturbative models are grossly inaccurate in the high-power regime.

## 5. Discussion

The capacity of three optical models, namely, the RPC, the LPC, and the MNC, were investigated. All three models are developed under two assumptions: channel memory is ignored and a sampling receiver is applied. Furthermore, two of these models, i.e., the RPC and the LPC, are based on perturbation theory and ignore signal–noise interaction, which makes them accurate only in the low-power regime. By tightly bounding the capacity of the RPC, by characterizing the capacity of the LPC, and by developing a tight upper bound on the capacity of the MNC, we showed that the capacity of these models, for the same underlying physical channel, behave very differently at high powers. Since the MNC is a more accurate channel model than the other two, one may conclude that the perturbative models become grossly inaccurate at high powers in terms of capacity calculation.

Note that the LPC model can be obtained from the MNC by neglecting the signal–noise interaction. Comparing the capacity of these two channels allows us to conclude that the impact of neglecting the signal–noise interaction on capacity is significant. Observe also that the capacity of the LPC model grows quickly with power because of the large capacity pre-log. Such a behavior is caused by the additive model used for the nonlinear distortion, which causes an artificial power increase at high SNR. A more accurate model than the RPC may be obtained by performing a normalization that conserves the signal power. Future work should consider more realistic channel models with nonzero dispersion and with more practical receivers.

**Author Contributions:** K.K. derived the theorems and proofs, carried out the numerical calculations, and wrote the paper. E.A. formulated the problem and contributed to the analysis. G.D. contributed to the information-theoretic analysis. All authors reviewed and revised the paper.

**Funding:** This work was supported by the Swedish Research Council (VR) under Grant 2013-5271.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A. Proof of Theorem 1

The capacity of the regular perturbative channel can be written as

$$\mathcal{C}_{\text{RPC}} = \sup I(\mathbf{x}; \mathbf{y}), \quad (\text{A1})$$

where the supremum is over all the probability distributions on  $\mathbf{x}$  that satisfy the power constraint (21). Let

$$\mathbf{w} = \mathbf{x} + j\eta|\mathbf{x}|^2\mathbf{x}. \quad (\text{A2})$$

We have that

$$I(\mathbf{x}; \mathbf{y}) = h(\mathbf{y}) - h(\mathbf{y} | \mathbf{x}) \quad (\text{A3})$$

$$= h(\mathbf{w} + \mathbf{n}) - h(\mathbf{w} + \mathbf{n} | \mathbf{x}) \quad (\text{A4})$$

$$= h(\mathbf{w} + \mathbf{n}) - h(\mathbf{n}). \quad (\text{A5})$$

Using the entropy power inequality ([31], Sec. 2.2) and the Gaussian entropy formula ([34], Th. 8.4.1), we conclude that

$$h(\mathbf{w} + \mathbf{n}) \geq \log(2^{h(\mathbf{w})} + 2^{h(\mathbf{n})}) \quad (\text{A6})$$

$$= \log(2^{h(\mathbf{w})} + \pi e P_N). \quad (\text{A7})$$

Substituting (A7) into (A5), and using again the Gaussian entropy formula ([34], Th. 8.4.1), we obtain

$$I(\mathbf{x}; \mathbf{y}) \geq \log(2^{h(\mathbf{w})} + \pi e P_N) - \log(\pi e P_N). \quad (\text{A8})$$

We take  $\mathbf{x}$  circularly symmetric. It follows from (A2) that  $\mathbf{w}$  is also circularly symmetric. Using ([35], Equation (320)) to compute  $h(\mathbf{w})$ , we obtain

$$h(\mathbf{w}) = h(|\mathbf{w}|^2) + \log \pi. \quad (\text{A9})$$

Substituting (A9) into (A8), we get

$$I(\mathbf{x}; \mathbf{y}) \geq \log\left(\frac{2^{h(|\mathbf{w}|^2)}}{e P_N} + 1\right). \quad (\text{A10})$$

Next, to evaluate the right-hand side (RHS) of (A10), we choose the following distribution for the amplitude square  $s = |\mathbf{x}|^2$  of  $\mathbf{x}$ :

$$f_s(s) = \zeta (3\eta^2 s^2 + 1) e^{-\lambda s}, \quad s \geq 0. \quad (\text{A11})$$

The parameters  $\lambda > 0$  and  $\zeta > 0$  are chosen so that (A11) is a pdf and so that the power constraint (21) is satisfied. We prove in Appendix A.1 that by choosing these two parameters so that

$$\zeta = \frac{\lambda^3}{\lambda^2 + 6\eta^2} \quad (\text{A12})$$

and so that (24) holds, both constraints are met. In Appendix A.2, we then prove that

$$h(|\mathbf{w}|^2) = -\log \zeta + \zeta \left( \frac{1}{\lambda} + \frac{18\eta^2}{\lambda^3} \right) \log e. \quad (\text{A13})$$

Substituting (A13) and (A12) into (A10), we obtain (22). Although not necessary for the proof, in Appendix A.3, we justify the choice of the pdf in (A11) by showing that it maximizes  $h(\mathbf{w})$ . (Based on (A5), the distribution that maximizes  $h(\mathbf{w})$  achieves capacity at high powers, where  $h(\mathbf{w} + \mathbf{n}) \approx h(\mathbf{w})$ .)

### Appendix A.1. Choosing $\zeta$ and $\lambda$

We choose the coefficients  $\zeta$  and  $\lambda$  so that (A11) is a valid pdf and  $E[s] \leq P$ . Note that

$$\int_0^{\infty} f_s(s) ds = \int_0^{\infty} \zeta (3\eta^2 s^2 + 1) e^{-\lambda s} ds \quad (\text{A14})$$

$$= \zeta \frac{\lambda^2 + 6\eta^2}{\lambda^3}. \quad (\text{A15})$$

Therefore, choosing  $\zeta$  according to (A12) guarantees that  $f_s(s)$  integrates to 1. We next compute  $E[s]$ :

$$E[s] = \int_0^{\infty} s f_s(s) ds \quad (\text{A16})$$

$$= \int_0^{\infty} s \zeta (3\eta^2 s^2 + 1) e^{-\lambda s} ds \quad (\text{A17})$$

$$= \zeta \left( \frac{18\eta^2}{\lambda^4} + \frac{1}{\lambda^2} \right). \quad (\text{A18})$$

Substituting (A12) into (A18), we obtain

$$E[s] = \frac{18\eta^2 + \lambda^2}{\lambda(\lambda^2 + 6\eta^2)}. \quad (\text{A19})$$

We see now that imposing  $E[s] \leq P$  is equivalent to (23). Observe that the RHS of (A19) and the objective function on the RHS of (22) are decreasing functions of  $\lambda$ . Therefore, setting the RHS of (A19) equal to  $P$ , which yields (24), maximizes the objective function in (22).

Finally, we prove that (24) has a single positive root. We have

$$f(\lambda) = P\lambda^3 - \lambda^2 + 6P\eta^2\lambda - 18\eta^2 \quad (\text{A20})$$

$$= (\lambda^2 + 6\eta^2)(P\lambda - 1) - 12\eta^2. \quad (\text{A21})$$

Note that  $f(\lambda) \rightarrow \infty$  as  $\lambda \rightarrow \infty$  and that the RHS of (A21) is negative when  $\lambda < 1/P$ . Furthermore,  $f(\lambda)$  is monotonically increasing in the interval  $[1/P, \infty)$ . Indeed, when  $\lambda \geq 1/P$ ,

$$\frac{d}{d\lambda} f(\lambda) = 3P\lambda^2 - 2\lambda + 6P\eta^2 \quad (\text{A22})$$

$$\geq 3\lambda - 2\lambda + 6P\eta^2 \quad (\text{A23})$$

$$> 0. \quad (\text{A24})$$

This yields the desired result.

### Appendix A.2. Proof of (A13)

To compute the differential entropy of  $\mathbf{t} = |\mathbf{w}|^2$ , we first determine the pdf of  $\mathbf{t}$ . By definition,

$$\mathbf{t} = \mathbf{s} + \eta^2 \mathbf{s}^3. \quad (\text{A25})$$

Let now  $g(x) = x + \eta^2 x^3$ . Since

$$\frac{d}{dx} g(x) = 1 + 3\eta^2 x^2, \quad (\text{A26})$$

we conclude that  $g(x)$  is monotonically increasing for  $x \geq 0$ . Hence,  $g(x)$  is one-to-one when  $x \geq 0$  and its inverse

$$q(x) = g^{-1}(x) \quad (\text{A27})$$

is well defined. Thus, the pdf of  $\mathbf{t}$  is given by ([36], Ch. 5)

$$f_{\mathbf{t}}(t) = \frac{f_{\mathbf{s}}(q(t))}{g'(q(t))} \quad (\text{A28})$$

$$= \frac{\zeta(3\eta^2 q^2(t) + 1) e^{-\lambda q(t)}}{3\eta^2 q^2(t) + 1} \quad (\text{A29})$$

$$= \zeta e^{-\lambda q(t)}, \quad t \geq 0. \quad (\text{A30})$$

Here, (A29) holds because of (A11). Using (A30), we can now compute  $h(\mathbf{t})$  as

$$h(\mathbf{t}) = - \int_0^\infty f_{\mathbf{t}}(t) \log(f_{\mathbf{t}}(t)) dt \quad (\text{A31})$$

$$= -\log \zeta + \lambda \zeta (\log e) \int_0^\infty q(t) e^{-\lambda q(t)} dt \quad (\text{A32})$$

$$= -\log \zeta + \lambda \zeta (\log e) \int_0^\infty r e^{-\lambda r} (1 + 3\eta^2 r^2) dr \quad (\text{A33})$$

$$= -\log \zeta + \lambda \zeta \left( \frac{1}{\lambda^2} + \frac{18\eta^2}{\lambda^4} \right) \log e, \quad (\text{A34})$$

where in (A33) we used the change of variables  $r = q(t)$ . This proves (A13).

#### Appendix A.3. $f_{\mathbf{s}}(s)$ Maximizes $h(\mathbf{w})$

We shall prove that the pdf  $f_{\mathbf{s}}(s) = \zeta (3\eta^2 s^2 + 1) e^{-\lambda s}$ ,  $s \geq 0$ , maximizes  $h(\mathbf{w})$ . It follows from (A9) that, to maximize  $h(\mathbf{w})$ , we need to maximize  $h(\mathbf{t})$ . We assume that the power constraint is fulfilled with equality, i.e., that

$$\mathbb{E}[\mathbf{s}] = \int_0^\infty s f_{\mathbf{s}}(s) ds = P. \quad (\text{A35})$$

Using the change of variables  $s = q(t)$ , where  $q(t)$  was defined in (A27), we obtain

$$\int_0^\infty q(t) f_{\mathbf{s}}(q(t)) q'(t) dt = P. \quad (\text{A36})$$

Substituting (A28) into (A36) and using that  $q'(t) = 1/g'(q(t))$ , we obtain

$$\int_0^\infty q(t) f_{\mathbf{t}}(t) dt = P. \quad (\text{A37})$$

It follows now from ([34], Th. 12.1.1) that the pdf that maximizes  $h(\mathbf{t})$  is of the form  $f_{\mathbf{t}}(t) = e^{\lambda_0 + \lambda_1 q(t)}$ ,  $t \geq 0$ , where  $\lambda_0$  and  $\lambda_1$  need to be chosen so that (A37) is satisfied and  $f_{\mathbf{t}}(t)$  integrates to one. Using (A28), we get

$$f_{\mathbf{s}}(s) = f_{\mathbf{t}}(g(s)) g'(s) \quad (\text{A38})$$

$$= (1 + 3\eta^2 s^2) e^{\lambda_0 + \lambda_1 s}, \quad s \geq 0. \quad (\text{A39})$$

By setting  $\zeta = e^{\lambda_0}$  and  $\lambda = \lambda_1$ , we obtain (A11).

## Appendix B. Proof of Theorem 2

Fix  $\lambda \geq 0$ . It follows from (20) and (21) that

$$\mathcal{C}_{\text{MNC}}(P) \leq \sup \left\{ I(\mathbf{x}; \mathbf{y}) + \lambda \left( 1 - \frac{\mathbb{E}[|\mathbf{x}|^2] + P_N}{P + P_N} \right) \right\}, \quad (\text{A40})$$

where the supremum is over the set of probability distributions that satisfy the power constraint (21). Next, we upper-bound the mutual information  $I(\mathbf{x}; \mathbf{y})$  as

$$I(\mathbf{x}; \mathbf{y}) = h(\mathbf{y}) - h(\mathbf{y} | \mathbf{x}) \quad (\text{A41})$$

$$= h(\mathbf{y}) - h(\mathbf{n}) \quad (\text{A42})$$

$$= h(|\mathbf{y}|) + h(\underline{\mathbf{y}} | |\mathbf{y}|) + \mathbb{E}(\log |\mathbf{y}|) - h(\mathbf{n}) \quad (\text{A43})$$

$$= h(|\mathbf{y}|^2) + h(\underline{\mathbf{y}} | |\mathbf{y}|) - \log 2 - h(\mathbf{n}) \quad (\text{A44})$$

$$\leq h(|\mathbf{y}|^2) + \log(\pi) - h(\mathbf{n}) \quad (\text{A45})$$

$$= h(|\mathbf{y}|^2) - \log(eP_N), \quad (\text{A46})$$

where in (A43) we used ([35], Lemma 6.16) and in (A44) we used ([35], Lemma 6.15). We fix now an arbitrary input pdf  $f_{\mathbf{x}}(\cdot)$  that satisfies the power constraint and define the random variables  $\mathbf{v} = |\mathbf{y}|^2$  and  $\mathbf{w} = \mathbf{x} + j\eta|\mathbf{x}|^2\mathbf{x}$ . Next, we shall obtain an upper bound on  $h(\mathbf{v})$  that is valid for all  $f_{\mathbf{x}}(\cdot)$ . Let

$$\tilde{f}_{\mathbf{v}}(v) = \kappa e^{-\mu q(v)}, \quad v \geq 0 \quad (\text{A47})$$

for some parameters  $\kappa > 0$  and  $\mu > 0$ . The function  $q(\cdot)$  is defined in (A27). We next choose  $\kappa$  so that  $\tilde{f}_{\mathbf{v}}(v)$  is a valid pdf. To do so, we set  $z = q(v)$ , which implies that  $g(z) = v$ , and that

$$(1 + 3\eta^2 z^2) dz = dv. \quad (\text{A48})$$

Therefore, integrating  $\tilde{f}(v)$  in (A47), we obtain

$$\int_0^\infty \kappa e^{-\mu q(v)} dv = \kappa \int_0^\infty e^{-\mu z} (1 + 3\eta^2 z^2) dz \quad (\text{A49})$$

$$= \kappa \left( \frac{\mu^2 + 6\eta^2}{\mu^3} \right). \quad (\text{A50})$$

We see from (A50) that the choice

$$\kappa = \frac{\mu^3}{\mu^2 + 6\eta^2} \quad (\text{A51})$$

makes  $\tilde{f}_{\mathbf{v}}(v)$  a valid pdf. Using the definition of the relative entropy, we have

$$D(f_{\mathbf{v}}(v) \| \tilde{f}_{\mathbf{v}}(v)) = \int_{-\infty}^{+\infty} f_{\mathbf{v}}(v) \log \left( \frac{f_{\mathbf{v}}(v)}{\tilde{f}_{\mathbf{v}}(v)} \right) dv \quad (\text{A52})$$

$$= -h(\mathbf{v}) - \mathbb{E}[\log(\tilde{f}_{\mathbf{v}}(v))] . \quad (\text{A53})$$

Since the relative entropy is nonnegative ([34], Thm. 8.6.1), we obtain

$$h(\mathbf{v}) \leq -\mathbb{E}_{\mathbf{v}}(\log(\tilde{f}_{\mathbf{v}}(\mathbf{v}))) \quad (\text{A54})$$

$$= -\log \kappa + \mu \mathbb{E}[q(\mathbf{v})] \log e. \quad (\text{A55})$$



Substituting (A46) and (A55) into (A40), we obtain

$$\mathcal{C}_{\text{MNC}}(P) \leq -\log(eP_N) - \log \kappa + \lambda + \sup \left\{ \mu E[q(\mathbf{v})] \log e - \lambda \frac{E[|\mathbf{x}|^2] + P_N}{P + P_N} \right\} \quad (\text{A56})$$

$$\leq -\log(eP_N) - \log \kappa + \lambda + \max_{s>0} \left\{ \mu E[q(\mathbf{v}) \mid |\mathbf{x}|^2 = s] \log e - \lambda \frac{s + P_N}{P + P_N} \right\}. \quad (\text{A57})$$

The final upper bound (25) is obtained by minimizing (A57) over all  $\lambda \geq 0$  and  $\mu \geq 0$ .

### Appendix C. Proof of Theorem 3

It follows from (A55) and (A46) that

$$I(\mathbf{x}; \mathbf{y}) \leq -\log(\kappa) + \mu E[q(\mathbf{v})] \log e - \log(eP_N), \quad (\text{A58})$$

where  $\mathbf{v} = |\mathbf{y}|^2$ . Moreover,

$$E[q(\mathbf{v})] = E[q(|\mathbf{w} + \mathbf{n}|^2)] \quad (\text{A59})$$

$$= E[q(|\mathbf{w}|^2 + |\mathbf{n}|^2 + 2\mathcal{R}(\mathbf{w}\mathbf{n}^*))]. \quad (\text{A60})$$

Next, we analyze the function  $q(x)$ . We have

$$q'(x) = \frac{1}{g'(q(x))} \quad (\text{A61})$$

$$= \frac{1}{1 + 3\eta^2 q^2(x)}. \quad (\text{A62})$$

Furthermore,

$$q''(x) = -\frac{6\eta^2 q(x)}{(1 + 3\eta^2 q^2(x))^3} \leq 0. \quad (\text{A63})$$

Therefore,  $q(x)$  is a nonnegative concave function on  $[0, \infty)$ . Thus, for every real numbers  $x \geq 0$  and  $y \geq -x$ ,

$$q(x + y) \leq q(x) + q'(x)y. \quad (\text{A64})$$

Using (A64) in (A60), with  $x = |\mathbf{w}|^2 + |\mathbf{n}|^2$  and  $y = 2\mathcal{R}(\mathbf{w}\mathbf{n}^*)$ , we get

$$E[q(\mathbf{v})] \leq E[q(|\mathbf{w}|^2 + |\mathbf{n}|^2) + 2q'(|\mathbf{w}|^2 + |\mathbf{n}|^2) \mathcal{R}(\mathbf{w}\mathbf{n}^*)]. \quad (\text{A65})$$

Using (A64) once more with  $x = |\mathbf{w}|^2$  and  $y = |\mathbf{n}|^2$ , we obtain

$$E[q(\mathbf{v})] \leq E[q(|\mathbf{w}|^2) + q'(|\mathbf{w}|^2) |\mathbf{n}|^2 + 2q'(|\mathbf{w}|^2 + |\mathbf{n}|^2) \mathcal{R}(\mathbf{w}\mathbf{n}^*)] \quad (\text{A66})$$

$$= E[q(|\mathbf{w}|^2)] + P_N E[q'(|\mathbf{w}|^2)] + 2E[q'(|\mathbf{w}|^2 + |\mathbf{n}|^2) \mathcal{R}(\mathbf{w}\mathbf{n}^*)]. \quad (\text{A67})$$

We shall now bound each expectation in (A67) separately. Since  $|\mathbf{w}|^2 = g(|\mathbf{x}|^2)$ , we have that

$$E[q(|\mathbf{w}|^2)] = E[q(|\mathbf{x}|^2)] \quad (\text{A68})$$

$$\leq P, \quad (\text{A69})$$

where the last inequality follows from (21). It also follows from (A62) that

$$q'(|\mathbf{w}|^2) \leq 1. \quad (\text{A70})$$

Furthermore, (A62) and (A63) imply that the function  $q'(x)$  is positive and decreasing in the interval  $x \geq 0$ . Therefore,

$$\mathbb{E}\left[q'(|\mathbf{w}|^2 + |\mathbf{n}|^2) \mathcal{R}(\mathbf{w}\mathbf{n}^*)\right] \leq \mathbb{E}\left[q'(|\mathbf{w}|^2 + |\mathbf{n}|^2) |\mathcal{R}(\mathbf{w}\mathbf{n}^*)|\right] \quad (\text{A71})$$

$$\leq \mathbb{E}\left[q'(|\mathbf{w}|^2) |\mathcal{R}(\mathbf{w}\mathbf{n}^*)|\right] \quad (\text{A72})$$

$$\leq \mathbb{E}\left[q'(|\mathbf{w}|^2) |\mathbf{w}| \cdot |\mathbf{n}|\right] \quad (\text{A73})$$

$$= \frac{\sqrt{\pi P_N}}{2} \mathbb{E}\left[q'(|\mathbf{w}|^2) |\mathbf{w}|\right] \quad (\text{A74})$$

$$\leq \frac{\sqrt{\pi P_N}}{2} \max_{t \geq 0} \left\{ t q'(t^2) \right\} \quad (\text{A75})$$

$$= \frac{\sqrt{\pi P_N}}{2} \max_{t \geq 0} \left\{ \frac{t}{1 + 3\eta^2 q^2(t^2)} \right\}, \quad (\text{A76})$$

where (A74) holds because  $\mathbb{E}[|\mathbf{n}|] = \sqrt{\pi P_N/4}$  and the last equality follows from (A62). To calculate the maximum in (A76), we use the change of variables  $t^2 = x + \eta^2 x^3$  to obtain

$$\max_{t \geq 0} \left\{ \frac{t}{1 + 3\eta^2 q^2(t^2)} \right\} = \max_{x \geq 0} \left\{ \frac{\sqrt{x + \eta^2 x^3}}{1 + 3\eta^2 x^2} \right\} \quad (\text{A77})$$

$$= \frac{1}{12^{3/8} \sqrt{(\sqrt{3} - 1) \eta}}, \quad (\text{A78})$$

where the last step follows by some standard algebraic manipulations that involve finding the roots of the derivative of the objective function on the RHS of (A77). Substituting (A78) into (A76), we obtain

$$\mathbb{E}\left[q'(|\mathbf{w}|^2 + |\mathbf{n}|^2) \mathcal{R}(\mathbf{w}\mathbf{n}^*)\right] \leq \frac{\sqrt{\pi P_N}}{2 \times 12^{3/8} \sqrt{(\sqrt{3} - 1) \eta}}. \quad (\text{A79})$$

Substituting (A69), (A70), and (A79) into (A67), and the result into (A58), we obtain

$$I(\mathbf{x}; \mathbf{y}) \leq -\log \kappa + \mu (\log e) \left\{ P + P_N + \frac{\sqrt{\pi P_N}}{12^{3/8} \sqrt{(\sqrt{3} - 1) \eta}} \right\} - \log(e P_N). \quad (\text{A80})$$

Finally, we obtain (25) by substituting (A51) into (A80). Since the upper bound (A80) on mutual information holds for every input distribution that satisfies the power constraint, it is also an upper bound on capacity for every  $\mu > 0$ . To find the optimal  $\mu$ , we need to minimize

$$\log\left(\frac{\mu^2 + 6\eta^2}{\mu^3}\right) + \mu (P + B) \log e = \log\left(\frac{\exp(\mu(P + B)) (\mu^2 + 6\eta^2)}{\mu^3}\right), \quad (\text{A81})$$

where  $B$  was defined in (28). Observe now that the function inside logarithm on the RHS of (A81) goes to infinity when  $\mu \rightarrow 0$  and when  $\mu \rightarrow \infty$ . Therefore, since this function is positive, it must have a minimum in the interval  $[0, \infty)$ . To find this minimum, we set its derivative equal to zero and get (28).

Note finally that, since (24) has exactly one real root, which was proved in Appendix A.1, (28) also has exactly one real root.

#### Appendix D. Proof of Theorem 5

The proof uses similar steps as in ([37], Sec. III-C). We upper-bound the mutual information between the  $\mathbf{x}$  and  $\mathbf{y}$  expressed in polar coordinates as

$$I(\mathbf{x}; \mathbf{y}) = I(\mathbf{r}_0, \theta_0; \mathbf{r}, \theta) \quad (\text{A82})$$

$$= I(\mathbf{r}_0, \theta_0; \mathbf{r}) + I(\mathbf{r}_0, \theta_0; \theta | \mathbf{r}) \quad (\text{A83})$$

$$= h(\mathbf{r}) - h(\mathbf{r} | \mathbf{r}_0, \theta_0) + h(\theta | \mathbf{r}) - h(\theta | \mathbf{r}, \mathbf{r}_0, \theta_0) \quad (\text{A84})$$

$$\leq h(\mathbf{r}^2) - E[\log(\mathbf{r})] - \log 2 - h(\mathbf{r} | \mathbf{r}_0, \theta_0) + \log(2\pi) - h(\theta | \mathbf{r}, \mathbf{r}_0, \theta_0). \quad (\text{A85})$$

In (A85), we used ([35], Equation (317)) and that  $h(\theta | \mathbf{r}) \leq \log(2\pi)$ . Let now  $\tilde{f}_{\mathbf{r}^2}(\cdot)$  denote an arbitrary pdf for  $\mathbf{r}^2$ . Following the same calculations as in (A52)–(A54), we obtain

$$h(\mathbf{r}^2) \leq -E_{\tilde{f}_{\mathbf{r}^2}}[\log(\tilde{f}_{\mathbf{r}^2}(\mathbf{r}^2))] \quad (\text{A86})$$

We shall take  $\tilde{f}_{\mathbf{r}^2}(\cdot)$  to be a Gamma distribution with parameters  $\alpha > 0$  and  $\beta = (P + P_N)/\alpha$ , i.e.,

$$\tilde{f}_{\mathbf{r}^2}(z) = \frac{z^{\alpha-1} e^{-z/\beta}}{\beta^\alpha \Gamma(\alpha)}, \quad z \geq 0. \quad (\text{A87})$$

Here,  $\Gamma(\cdot)$  denotes the Gamma function. Substituting (A87) into (A86), we obtain

$$E[\log(\tilde{f}_{\mathbf{r}^2}(\mathbf{r}^2))] \quad (\text{A88})$$

$$= 2(\alpha - 1)E[\log(\mathbf{r})] - \alpha \frac{E[\mathbf{r}_0^2] + P_N}{P + P_N} \log e - \alpha \log\left(\frac{P + P_N}{\alpha}\right) - \log(\Gamma(\alpha)). \quad (\text{A89})$$

It follows from (15) that the random variables  $\mathbf{r}$  and  $\theta_0$  are conditionally independent given  $\mathbf{r}_0$ . Therefore,

$$h(\mathbf{r} | \mathbf{r}_0, \theta_0) = h(\mathbf{r} | \mathbf{r}_0). \quad (\text{A90})$$

Next, we study the term  $h(\theta | \mathbf{r}, \mathbf{r}_0, \theta_0)$  in (A85). From Bayes' theorem and (15) it follows that, for every  $\theta' \in [0, 2\pi)$ ,

$$f_{\theta|\mathbf{r}, \mathbf{r}_0, \theta_0}(\theta | r, r_0, \theta_0) = f_{\theta|\mathbf{r}, \mathbf{r}_0, \theta_0}(\theta - \theta' | r, r_0, \theta_0 - \theta'). \quad (\text{A91})$$

Therefore,

$$h(\theta | \mathbf{r}, \mathbf{r}_0, \theta_0) = \int_0^{2\pi} f_{\theta_0}(\theta_0) h(\theta | \mathbf{r}, \mathbf{r}_0, \theta_0 = \theta_0) d\theta_0 \quad (\text{A92})$$

$$= \int_0^{2\pi} f_{\theta_0}(\theta_0) h(\theta - \theta_0 | \mathbf{r}, \mathbf{r}_0, \theta_0 = 0) d\theta_0 \quad (\text{A93})$$

$$= \int_0^{2\pi} f_{\theta_0}(\theta_0) h(\theta | \mathbf{r}_0, \theta_0 = 0) d\theta_0 \quad (\text{A94})$$

$$= h(\theta | \mathbf{r}, \mathbf{r}_0, \theta_0 = 0). \quad (\text{A95})$$

Here, (A94) follows because differential entropy is invariant to translations ([34], Th. 8.6.3). Substituting (A89), (A86), (A90), and (A95) into (A85), we obtain

$$\begin{aligned} I(\mathbf{r}_0, \theta_0; \mathbf{r}, \theta) &\leq \alpha \log\left(\frac{P + P_N}{\alpha}\right) + \log(\Gamma(\alpha)) + \log(\pi) + \alpha \frac{E[\mathbf{r}_0^2] + P_N}{P + P_N} \log e \\ &\quad + (1 - 2\alpha)E[\log(\mathbf{r})] - h(\mathbf{r} | \mathbf{r}_0) - h(\theta | \mathbf{r}, \mathbf{r}_0, \theta_0 = 0). \end{aligned} \quad (\text{A96})$$

Fix  $\lambda \geq 0$ . We next upper bound  $\mathcal{C}_{\text{MNC}}$  using (A96) as

$$\mathcal{C}_{\text{MNC}}(P) \leq \sup \left\{ I(\mathbf{r}_0, \boldsymbol{\theta}_0; \mathbf{r}, \boldsymbol{\theta}) + \lambda \left( 1 - \frac{\mathbb{E}[\mathbf{r}_0^2] + P_N}{P + P_N} \right) \right\} \quad (\text{A97})$$

$$\begin{aligned} &\leq \alpha \log \left( \frac{P + P_N}{\alpha} \right) + \log(\Gamma(\alpha)) + \log(\pi) + \lambda \\ &\quad + \sup \left\{ (\alpha \log e - \lambda) \frac{\mathbb{E}[\mathbf{r}_0^2] + P_N}{P + P_N} + (1 - 2\alpha) \mathbb{E}[\log(\mathbf{r})] - h(\mathbf{r} | \mathbf{r}_0) \right. \\ &\quad \left. - h(\boldsymbol{\theta} | \mathbf{r}, \mathbf{r}_0, \boldsymbol{\theta}_0 = 0) \right\}, \end{aligned} \quad (\text{A98})$$

where the supremum is over the set of input probability distributions that satisfy (21). We complete the proof by noting that the supremum in (A98) is less than or equal to  $\max_{r_0 > 0} \{g_{\lambda, \alpha}(r_0, P)\}$ , where  $g_{\lambda, \alpha}(r_0, P)$  is defined in (36).

## References

1. Mecozzi, A. Limits to long-haul coherent transmission set by the Kerr nonlinearity and noise of the in-line amplifiers. *J. Lightw. Technol.* **1994**, *12*, 1993–2000. [\[CrossRef\]](#)
2. Essiambre, R.J.; Kramer, G.; Winzer, P.J.; Foschini, G.J.; Goebel, B. Capacity Limits of Optical Fiber Networks. *J. Lightw. Technol.* **2010**, *28*, 662–701. [\[CrossRef\]](#)
3. Merhav, N.; Kaplan, G.; Lapidoth, A.; Shamai (Shitz), S. On information rates for mismatched decoders. *IEEE Trans. Inform. Theory* **1994**, *40*, 1953–1967. [\[CrossRef\]](#)
4. Secondini, M.; Forestieri, E. Scope and limitations of the nonlinear Shannon limit. *J. Lightw. Technol.* **2017**, *35*, 893–902. [\[CrossRef\]](#)
5. Ghozlan, H.; Kramer, G. Models and information rates for multiuser optical fiber channels with nonlinearity and dispersion. *IEEE Trans. Inform. Theory* **2017**, *63*, 6440–6456. [\[CrossRef\]](#)
6. Secondini, M.; Forestieri, E.; Prati, G. Achievable Information Rate in Nonlinear WDM Fiber-Optic Systems With Arbitrary Modulation Formats and Dispersion Maps. *J. Lightw. Technol.* **2013**, *31*, 3839–3852. [\[CrossRef\]](#)
7. Fehenberger, T.; Alvarado, A.; Bayvel, P.; Hanik, N. On achievable rates for long-haul fiber-optic communications. *Opt. Express* **2015**, *23*, 9183–9191. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Mitra, P.P.; Stark, J.B. Nonlinear Limits to the Information Capacity of Optical Fibre Communications. *Nature* **2001**, *411*, 1027–1030. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Ellis, A.D.; Zhao, J.; Cotter, D. Approaching the Nonlinear Shannon Limit. *J. Lightw. Technol.* **2010**, *28*, 423–433. [\[CrossRef\]](#)
10. Dar, R.; Shtaif, M.; Feder, M. New Bounds on the Capacity of the Nonlinear Fiber-Optic Channel. *Opt. Lett.* **2014**, *39*, 398–401. [\[CrossRef\]](#)
11. Splett, A.; Kurzke, C.; Petermann, K. Ultimate Transmission Capacity of Amplified Optical Fiber Communication Systems taking into account Fiber Nonlinearities. In Proceedings of the European Conference on Optical Communication (ECOC), Montreux, Switzerland, 12–16 September 1993; p. MoC2.4.
12. Kramer, G.; Yousefi, M.I.; Kschischang, F.R. Upper Bound on the Capacity of a Cascade of Nonlinear and Noisy Channels. In Proceedings of the IEEE Information Theory Workshop (ITW), Rotorua, New Zealand, 29 August–1 September 2015.
13. Agrawal, G.P. *Nonlinear Fiber Optics*, 4th ed.; Elsevier: San Diego, CA, USA, 2006.
14. Agrell, E.; Durisi, G.; Johannisson, P. Information-theory-friendly models for fiber-optic channels: A primer. In Proceedings of the IEEE Information Theory Workshop (ITW), Rotorua, New Zealand, 29 August–1 September 2015.
15. Peddanarappagari, K.V.; Brandt-Pearce, M. Volterra series transfer function of single-mode fibers. *J. Lightw. Technol.* **1997**, *15*, 2232–2241. [\[CrossRef\]](#)
16. Mecozzi, A.; Clausen, C.B.; Shtaif, M. Analysis of Intrachannel Nonlinear Effects in Highly Dispersed Optical Pulse Transmission. *IEEE Photon. Technol. Lett.* **2000**, *12*, 392–394. [\[CrossRef\]](#)

17. Meccozzi, A.; Essiambre, R.J. Nonlinear Shannon Limit in Pseudolinear Coherent Systems. *J. Lightw. Technol.* **2012**, *30*, 2011–2024. [\[CrossRef\]](#)
18. Taghavi, M.H.; Papen, G.C.; Siegel, P.H. On the Multiuser Capacity of WDM in a Nonlinear Optical Fiber: Coherent Communication. *IEEE Trans. Inform. Theory* **2006**, *52*, 5008–5022. [\[CrossRef\]](#)
19. Dar, R.; Feder, M.; Mecozzi, A.; Shtaif, M. Accumulation of nonlinear interference noise in fiber-optic systems. *Opt. Express* **2014**, *22*, 14199–14211. [\[CrossRef\]](#)
20. Ho, K.P. *Phase-Modulated Optical Communication Systems*; Springer: New York, NY, USA, 2005.
21. Turitsyn, K.S.; Derevyanko, S.A.; Yurkevich, I.; Turitsyn, S.K. Information Capacity of Optical Fiber Channels with Zero Average Dispersion. *Phys. Rev. Lett.* **2003**, *91*, 203901–1–203901–4. [\[CrossRef\]](#)
22. Yousefi, M.I.; Kschischang, F.R. On the per-sample capacity of nondispersive optical fibers. *IEEE Trans. Inform. Theory* **2011**, *57*, 7522–7541. [\[CrossRef\]](#)
23. Terekhov, I.; Reznichenko, A.; Kharkov, Y.A.; Turitsyn, S. Log-log growth of channel capacity for nondispersive nonlinear optical fiber channel in intermediate power range. *Phys. Rev. E* **2017**, *95*, 062133. [\[CrossRef\]](#)
24. Panarin, A.; Reznichenko, A.; Terekhov, I. Next-to-leading-order corrections to capacity for a nondispersive nonlinear optical fiber channel in the intermediate power region. *Phys. Rev. E* **2017**, *95*, 012127. [\[CrossRef\]](#)
25. Reznichenko, A.; Chernykh, A.; Smirnov, S.; Terekhov, I. Log-log growth of channel capacity for nondispersive nonlinear optical fiber channel in intermediate power range: Extension of the model. *Phys. Rev. E* **2019**, *99*, 012133. [\[CrossRef\]](#)
26. Keykhosravi, K.; Durisi, G.; Agrell, E. A Tighter Upper Bound on the Capacity of the Nondispersive Optical Fiber Channel. In Proceedings of the European Conference on Optical Communication (ECOC), Gothenburg, Sweden, 17–21 September 2017.
27. Kramer, G. Autocorrelation function for dispersion-free fiber channels with distributed amplification. *IEEE Trans. Inform. Theory* **2018**, *64*, 5131–5155. [\[CrossRef\]](#)
28. Lau, A.P.T.; Kahn, J.M. Signal design and detection in the presence of nonlinear. *J. Lightw. Technol.* **2007**, *25*, 3008–3016. [\[CrossRef\]](#)
29. Tavana, M.; Keykhosravi, K.; Aref, V.; Agrell, E. A Low-Complexity Near-Optimal Detector for Multispan Zero-Dispersion Fiber-Optic Channels. In Proceedings of the European Conference on Optical Communication (ECOC), Roma, Italy, 23–27 September 2018.
30. Forestieri, E.; Secondini, M. Solving the nonlinear Schrödinger equation. In *Optical Communication Theory and Techniques*; Forestieri, E., Ed.; Springer: Boston, MA, USA, 2005; pp. 3–11.
31. Gamal, A.E.; Kim, Y.H. *Network Information Theory*; Cambridge University Press: Cambridge, UK, 2011.
32. Li, S.; Häger, C.; Garcia, N.; Wymeersch, H. Achievable information rates for nonlinear fiber communication via end-to-end autoencoder learning. In Proceedings of the European Conference on Optical Communication (ECOC), Roma, Italy, 23–27 September 2018.
33. Yousefi, M.I. The asymptotic capacity of the optical fiber. *arXiv* **2016**, arXiv:1610.06458.
34. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; Wiley: Hoboken, NJ, USA, 2006.
35. Lapidot, A.; Moser, S.M. Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels. *IEEE Trans. Inform. Theory* **2003**, *49*, 2426–2467. [\[CrossRef\]](#)
36. Papoulis, A.; Pillai, S.U. *Probability, Random Variables, and Stochastic Processes*, 4th ed.; McGraw-Hill Education: New York, NY, USA, 2002.
37. Durisi, G. On the capacity of the block-memoryless phase-noise channel. *IEEE Commun. Lett.* **2012**, *16*, 1157–1160. [\[CrossRef\]](#)

