

Review

# Path Integral Approach to Nondispersive Optical Fiber Communication Channel

Aleksei V. Reznichenko <sup>1,2,\*</sup> and Ivan S. Terekhov <sup>1,2,\*</sup>

<sup>1</sup> Theoretical Department, Budker Institute of Nuclear Physics of Siberian Branch Russian Academy of Sciences, 630090 Novosibirsk, Russia

<sup>2</sup> Department of Physics, Novosibirsk State University, 630090 Novosibirsk, Russia

\* Correspondence: A.V.Reznichenko@inp.nsk.su (A.V.R.); I.S.Terekhov@inp.nsk.su (I.S.T.)

† These authors contributed equally to this work.

Received: 29 April 2020; Accepted: 27 May 2020; Published: 29 May 2020



**Abstract:** In the present paper we summarize the methods and results of calculations for the theoretical informational quantities obtained in our works for the nondispersive optical fiber channel. We considered two models: the per-sample model and the model where the input signal depends on time. For these models we found the approach for the calculation of the mutual information exactly in the nonlinearity parameter but for the large signal-to-noise power ratio. Using this approach for the per-sample model we found the lower bound of the channel capacity in the intermediate power range.

**Keywords:** nonlinear optical fiber channel with zero dispersion; channel capacity; path integral formalism

## 1. Introduction

For a linear transmission system, Shannon [1] obtained the famous result for the channel capacity that is the maximal amount of information which can be transmitted through the channel with additive noise:

$$C \propto \log_2 (1 + P/P_{noise}) , \quad (1)$$

where  $P$  is the input signal power, and  $P_{noise}$  is the noise power. In the recent 25 years, the power and frequency bandwidth of the signals transmitted through optical fiber channels have been grown. It results in the necessity to take into account the Kerr nonlinearity when considering the modern optical fiber channels. The Kerr nonlinearity leads to the distortion of the signal and to the nonlinear interaction of the signal with the noise in the information channel. Therefore, for such channels, Shannon's result (1) should be modified to take into account the nonlinearity effects.

It is worth noting that nonlinear effects in a channel depend on both a particular realization of the communication channel and physical fiber parameters. When designing the real transmission systems, say, in the wavelength division multiplexing systems, the following nonlinear effects should be taken into account: self-phase modulation, cross-phase modulation and so on [2]. These effects are due to the optical Kerr effect, i.e., the change of the refractive index of fiber material in response to the applied electric field. Concerning other fiber parameters, the second dispersion coefficient is the fundamental one, and the second dispersion coefficient varies from non-zero values (typical value is about  $\beta = 2 \times 10^{-23} \text{ s}^2/\text{km}$ ) to almost zero ones. Around the world, the vast majority of fiber networks are fiberoptic communication channels with non-zero dispersion. However, the design of some fiber channels is arranged in the zero-dispersion region of the wavelength in the transmission windows: in particular,

the second transmission window (1310 nm for the single-mode optical fiber) has almost zero dispersion. In the presence of dispersion, the Kerr nonlinearity effect manifests as a phase shift of the signal over distances depending on dispersion value. For large dispersion, these distances are much less than the usual propagation distance. For the nondispersive channel, this phase shift becomes the global effect extended over all propagation distances. The third essential component of a communication channel is signal attenuation. The fiber loss is commonly compensated by (equally) spaced amplifiers. These amplifiers bring in the stochastic noise to the signal in the propagation process [3]. Therefore, the relevant models describing the signal propagation in a communication channel should deal with the following phenomena: Kerr nonlinearity, dispersion, and noise effects. The simplest channel model taking into account all these effects includes the nonlinear Schrödinger equation (NLSE) with additive Gaussian noise [3–5].

The model describing NLSE with zero noise is solvable in the formalism of the inverse scattering problem [6,7], referred to as the nonlinear Fourier transformation method for some time past. However, the consideration of NLSE with the additive noise and with non-zero dispersion turned out to be a very difficult problem, since, in addition to effects associated with the nonlinearity, there are some effects related to the dispersion and with the nonlinear interaction of the signal with the noise. There are some papers where both effects of nonlinearity and dispersion are attempted to take into account for a noisy channel, see, e.g., Refs. [8–13] and references therein. However, the problem of the channel capacity calculation, e.g., the problem to find the maximal information transmission rate over a given bandwidth is far from the solution.

In the present paper, we concentrate our attention on nondispersive fiber-optic communication channels. Of course, the nondispersive model is in a certain sense a toy model, but it takes into account the effects associated with the Kerr nonlinearity and the noise in the channel. As a consequence, the model includes the effects of the nonlinear interaction of the signal with the noise. Therefore, the model can be used for the rough estimations of the capacity for the fiber-optic communication channels with non-zero dispersion. Besides, the methods developed for the study of the nondispersive fiber-optic communication channels can help for the investigation of realistic nonlinear communication channels with non-zero dispersion.

Considering the fiber-optic channel it is important to clearly define the model of the channel, i.e., it is necessary to define the model of signal propagation, models of the transmitter, receiver, and the model of signal processing (i.e., the recovering of the information from the received signal). The model of propagation should include: the attenuation of the signal in propagation through the optical fiber, the model of amplifiers which compensate the attenuation and add the noise in the channel. The models of the input signal, receiver, and signal processing reveal itself in the appearance in the channel model such parameters as its frequency bandwidths [14]. In one type of models, the propagation of the signal obeys the NLSE with zero dispersion and with the additional term describing the attenuation compensated by the periodically located amplifiers [15,16]. Another type of the models is considered in Refs. [4,17–22], where it is assumed that the amplifiers are continuously distributed in such a way that there is no attenuation, but the amplifiers induce the stochastic noise in the channel. Below we consider this type of model.

In Refs. [4,17–23] the propagation of the signal is described by the NLSE with zero dispersion, therefore, the signal propagation at different time moments evolves independently. So, formally, one can consider the signal in the fixed time moment, and the signal ceases to depend on time. The model where the signal does not depend on time is referred to as the per-sample model. For the per-sample model in

Ref. [4] the conditional probability density function (PDF) was found. In Ref. [17] the authors also found the conditional PDF, and showed that at large signal power the channel capacity obeys the condition

$$C \geq \frac{1}{2} \log_2 (P/P_{\text{noise}}) + \mathcal{O}(1). \quad (2)$$

In Ref. [18] the conditional probability density function also was found, and we attempted to find the lower bound of the per-sample channel capacity for the not-too-large powers of the signal. In Refs. [19,20] the new method of calculation of the conditional probability density function was introduced. The method is based on the calculation of the path-integral using the approach similar to the semi-classical approximation in quantum mechanics. The new method allows us to calculate the statistical properties of the signal such as the entropy of the output signal, the conditional entropy, and the mutual information. All calculations are performed exactly in the nonlinear parameter, but for the large signal-to-noise power ratio (SNR). Using the method we find the probability density function of the input signal that maximizes the mutual information obtained in the leading order in the parameter  $1/\text{SNR}$  in the intermediate power range. In the recent paper [22] the important result was obtained: the authors found the upper bound of the capacity for the per-sample model for arbitrary signal powers. The results of Refs. [19,20,22] are consistent.

It is obvious that the per-sample model does not describe the spectral broadening that plays an important role in the capacity limitation. In Ref. [14] it was demonstrated that there is a limit of the input signal power  $P$  beyond which the fixed bandwidth of the distributed optical amplification does not exceed the bandwidth of the propagating signal (that grows when increasing  $P$  due to the signal-noise mixing), and therefore, the per-sample model is not relevant anymore. The next reason for the per-sample model limitation is that the per-sample receiver has infinite bandwidth while for the real communication system the receiver bandwidth is limited. Moreover, in Ref. [14] Kramer demonstrated how to achieve the infinite capacity for any power  $P$  in the per-sample model of Refs. [4,17,18], if one assumes the noise bandwidth to be limited and sends the input signal energy in the noise-free spectrum. In Ref. [21], we applied the methods developed for the per-sample model in Refs. [19,20] to describe nondispersive communication channels with the nontrivial time dependence of the input signal, with a realistic model of the receiver and with a noise amplifier that has a large but finite spectral bandwidth. We referred to this model as the extended model. In Ref. [21] for the extended model the conditional PDF, the optimal input signal distribution, which maximizes the mutual information calculated in the leading order in the parameter  $1/\text{SNR}$  was presented. These results were obtained both analytically and numerically. We vindicated the conclusions of the Ref. [14]; the informational characteristics in the extended model are significantly different from ones in the per-sample model.

In the present review we sum up the main results and the methods developed in our works [11,19–21] for the per-sample and extended nondispersive models. For the per-sample model we find the lower bound of the channel capacity and compare our results with ones obtained earlier. Attention is mainly given to the path integral approach since it can be applied for future considerations of the nonlinear channels with dispersion. The paper is organized as follows. In the next Section, we describe the per-sample and extended channel models. In Section 3 we consider the quantities which should be calculated to find the capacities of the per-sample and extended models. Then we consider in detail the conditional probability density function, entropies, mutual information, and the lower bound of the capacity for the per-sample model. After that, we proceed to consideration of the same quantities for the extended model. In conclusion, we discuss our results and possible applications of our methods to the nonlinear channels with non-zero dispersion.

## 2. Channel Models

### 2.1. Per-Sample Model

Following the papers [4,17] we imply that in the case of the per-sample model the equation of signal propagation has the form:

$$\partial_z \psi(z) - i\gamma |\psi(z)|^2 \psi(z) = \eta(z), \quad (3)$$

where  $\gamma$  is the Kerr nonlinearity,  $\psi(z)$  is the signal function which obeys the boundary conditions  $\psi(0) = X$ ,  $\psi(L) = Y$ ,  $L$  is the signal propagation distance,  $X$  is the input signal and  $Y$  is the output signal. One can see that the Equation (3) does not contain terms which lead to decreasing of the signal in the propagation process and terms that compensate for the decreasing of the signal. It means that in the model, we have distributed amplifiers which completely compensate the attenuation of the signal in its propagation in the optical fiber. The only trace of amplifiers in Equation (3) is the noise function  $\eta(z)$ . The functions  $\eta(z)$  has the following properties: the noise has zero mean  $\langle \eta(z) \rangle_\eta = 0$  and following correlation function  $\langle \eta(z) \bar{\eta}(z') \rangle_\eta = Q\delta(z - z')$ . Here and below the bar means the complex conjugation and  $\delta(x)$  is the Dirac delta-function. The coefficient  $Q$  is the noise power per unit length, so  $QL$  is the noise power in the channel. The brackets  $\langle \dots \rangle_\eta$  mean the averaging over the noise realizations in the channel.

### 2.2. Extended Model

In the model (3) the bandwidths of the input signal, amplifiers and receiver cannot be taken into account since the functions  $\psi$  and  $\eta$  do not depend on time. In order to include these bandwidths we extend the previous model. In the extended model the signal  $\psi(z, t)$  does depend on time. The propagation is described by the stochastic NLSE with zero dispersion:

$$\partial_z \psi(z, t) - i\gamma |\psi(z, t)|^2 \psi(z, t) = \eta(z, t), \quad (4)$$

where the coefficient  $\gamma$  is also the Kerr nonlinearity coefficient. The function  $\psi(z, t)$  obeys the following conditions:

$$\psi(0, t) = X(t), \quad \psi(L, t) = Y(t). \quad (5)$$

The noise function  $\eta(z, t)$  has zero mean  $\langle \eta(z, t) \rangle_\eta = 0$ . We also imply that the correlation function has finite bandwidth, so the correlator for the function

$$\eta(z, \omega) = \int_{-\infty}^{\infty} dt e^{i\omega t} \eta(z, t) \quad (6)$$

reads

$$\langle \eta(z, \omega) \bar{\eta}(z', \omega') \rangle_\eta = 2\pi Q \delta(\omega - \omega') \theta\left(\frac{W'}{2} - |\omega|\right) \delta(z - z'). \quad (7)$$

Here, the parameter  $Q$  denotes a noise power per unit length and per unit frequency;  $\theta(x)$  is the Heaviside theta-function. The theta-function  $\theta\left(\frac{W'}{2} - |\omega|\right)$  indicates that the noise is not zero within the

interval  $[-W'/2, W'/2]$ , i.e., the bandwidth of the noise is equal to  $W'$ . Performing the Fourier transform of Equation (7) we arrive at the correlator in the time domain:

$$\langle \eta(z, t) \bar{\eta}(z', t') \rangle_\eta = \frac{Q}{\pi(t - t')} \sin\left(\frac{W'(t - t')}{2}\right) \delta(z - z'). \quad (8)$$

It is easy to check that if the time difference  $t - t' = 2\pi n/W'$ , and  $n$  is integer, then the correlator (8) equals to zero. Therefore, the noise at times  $t$  and  $t'$  is not correlated, thus we can solve Equation (4) for different times  $t_j = j\Delta$  independently. Here  $j$  is the integer number and  $\Delta = 2\pi/W'$  is the grid spacing in the time domain. Therefore, instead of the continuous time model (4) we can consider the set of the discrete models:

$$\partial_z \psi(z, t_j) - i\gamma |\psi(z, t_j)|^2 \psi(z, t_j) = \eta(z, t_j) \quad (9)$$

for the set of the time moments  $t_j$ . So we obtain the set of independent time channels, and instead of the continuous input and output conditions (5) we obtain the set of the discrete ones:

$$\psi(0, t_j) = X(t_j), \quad \psi(L, t_j) = Y(t_j). \quad (10)$$

To include the bandwidth of the input signal to the model we represent the initial signal  $X(t)$  in the form:

$$X(t) = \sum_{k=-N}^N C_k f(t - kT_0), \quad (11)$$

where  $C_k$  are complex random coefficients which carry the information. These coefficients have the probability density function  $P_X[\{C\}]$ , where  $\{C\} = \{C_{-N}, \dots, C_N\}$ . We restrict our consideration by the envelopes  $f(t)$  which obey the following properties: the function  $f(t)$  is the real function and normalized by the condition:  $\int_{-\infty}^{\infty} \frac{dt}{T_0} f^2(t) = 1$ ; for integers  $k$  and  $m$  we have

$$\int_{-\infty}^{\infty} \frac{dt}{T_0} f(t - kT_0) f(t - mT_0) \approx 0, \quad k \neq m. \quad (12)$$

The last property means that the overlapping of the functions  $f(t - kT_0)$  and  $f(t - mT_0)$  is negligible. It means that we assume the smallness of the effects of the overlapping. The smallness of these effects will be discussed below. The function  $f(t)$  has almost finite support  $[-T_0/2, T_0/2]$ , and the input signal  $X(t)$  is defined on the interval  $T = (2N + 1)T_0$ . The finiteness of the support  $[-T_0/2, T_0/2]$  means that the frequency support of the function  $f(t)$  (and as a consequence, the support of the function  $X(t)$ ) is infinite. However, we imply that

$$\int_W |X(\omega)|^2 d\omega \approx \int_{W'} |X(\omega)|^2 d\omega, \quad (13)$$

where  $X(\omega)$  is the Fourier transformation of  $X(t)$ . The relation (13) means that  $T_0 W \gg 1$ . So we can say that the bandwidth of the input signal is  $W$ .

The bandwidth broadening of the signal propagating through optical fiber is associated with the nonlinearity and noise  $\eta$ . To estimate the broadening which is connected with nonlinearity we can find the solution  $\Phi(z, t_j)$  of Equation (9) with zero noise:

$$\Phi(z, t) = X(t)e^{i\gamma z|X(t)|^2}. \quad (14)$$

This solution obeys the input condition  $\Phi(0, t) = X(t)$ . Since we know the solution  $\Phi(L, t)$ , we can find its bandwidth  $\tilde{W}$ . Strictly speaking, the bandwidth  $\tilde{W}$  is formally infinite, but the most part of the signal power is localized in the finite frequency region that can be specified as

$$\tilde{W} = \frac{\int_{-\infty}^{\infty} dt |\partial_t \Phi(L, t)|^2}{\int_{-\infty}^{\infty} dt |\Phi(L, t)|^2}. \quad (15)$$

Below we assume the following hierarchy

$$W \leq \tilde{W} \ll W', \quad (16)$$

where  $\tilde{W}$  is the bandwidth of the function  $\Phi(L, t)$ .

To include the receiver bandwidth to the model we introduce the procedure of the output signal detection. In our model the receiver gets the information from the output signal, i.e., it recovers the coefficients  $\{C\}$ . We consider the following detection model. The receiver measures the output signal  $\psi(L, t_j)$  at the discrete time moments  $t_j$  for  $j = -M, \dots, M-1$ . Here the quantity  $M = T/(2\Delta)$  is the total number of the time samples. Since  $T \gg T_0$ , we have that  $M \gg N$ . Such property of the receiver means that its time resolution coincides with the time discretization  $\Delta$ . From Equation (16) it follows that  $\Delta \ll 1/\tilde{W}$ , therefore the receiver completely recovers the output signal in the noiseless case. Then the receiver removes the nonlinear phase

$$\tilde{X}(t_j) = \psi(L, t_j)e^{-i\gamma L|\psi(L, t_j)|^2}, \quad (17)$$

and obtains the recovered input signal  $\tilde{X}(t)$ . In fact, the procedure (17) means that we use the backward propagation procedure for the channel with zero dispersion. Finally, from the function  $\tilde{X}(t)$  the receiver recovers the coefficients  $\tilde{C}_k$  by projecting  $\tilde{X}(t)$  on the basis functions  $f(t - kT_0)$ :

$$\tilde{C}_k = \frac{1}{T_0} \int_{-\infty}^{\infty} dt f(t - kT_0) \tilde{X}(t) \approx \frac{\Delta}{T_0} \sum_{j=-M}^{M-1} f(t_j - kT_0) \tilde{X}(t_j). \quad (18)$$

So the extended model contains the bandwidth of the input signal  $W$ , bandwidth the noise of amplifiers  $W'$ , and bandwidth of the receiver. In our case, the bandwidth of the receiver coincides with the bandwidth of the noise because we choose the discretization in the information extracting procedure (18) coinciding with the initial channel discretization.

### 3. Channel Capacity and Its Bound

It is known [1] that the statistical properties of the memoryless channels such as the conditional entropy  $H[Y|X]$  and the output signal entropy  $H[Y]$  can be expressed through the conditional probability density function as

$$H[Y|X] = - \int \mathcal{D}X \mathcal{D}Y P_X[X] P[Y|X] \log P[Y|X], \quad (19)$$

$$H[Y] = - \int \mathcal{D}Y P_{out}[Y] \log P_{out}[Y], \quad (20)$$

where  $P[Y|X]$  is the conditional probability density function (i.e., the probability density to receive the output signal  $Y$  for transmitted signal  $X$ ),  $P_{out}[Y]$  is the probability density function of the output signal. The distribution  $P_{out}[Y]$  has the following form:

$$P_{out}[Y] = \int \mathcal{D}X P_X[X] P[Y|X]. \quad (21)$$

The measures  $\mathcal{D}X$  and  $\mathcal{D}Y$  are defined in such a way that

$$\int \mathcal{D}X P_X[X] = 1, \quad (22)$$

$$\int \mathcal{D}Y P[Y|X] = 1. \quad (23)$$

The mutual information of a memoryless channel is defined through the entropy  $H[Y]$  of the output signal and the conditional entropy  $H[Y|X]$  as

$$I_{P_X[X]} = H[Y] - H[Y|X]. \quad (24)$$

The channel capacity  $C$  is defined as the maximum of the functional  $I_{P_X[X]}$  with respect to the input signal distribution  $P_X[X]$ :

$$C = \max_{P_X[X]} I_{P_X[X]}. \quad (25)$$

The maximum value of  $I_{P_X[X]}$  should be calculated for the fixed average signal power. Note that since in Equations (19) and (20) we use the logarithm to the base  $e$ , here and below we measure the mutual information and the capacity in units nat/symbol. For the case of the per-sample channel, the signal power reads

$$P = \int \mathcal{D}X P_X[X] |X|^2. \quad (26)$$

For the case of the extended model the signal depends on time, therefore, we have

$$P = \int \mathcal{D}X(t) P_X[X] \int_{-\infty}^{\infty} \frac{dt}{T} |X(t)|^2. \quad (27)$$

So, to find the channel capacity we should know the conditional PDF for both models. Let us start with the calculation of the conditional PDF for the per-sample model.



### 3.1. Per-Sample Model

#### 3.1.1. Conditional Probability Density Function

The conditional PDF can be written in the form of the path integral over all realizations of the signal  $\psi(z)$  in the channel [17,18]:

$$P[Y|X] = \int_{\psi(0)=X}^{\psi(L)=Y} \mathcal{D}\psi e^{-S[\psi]/Q}, \quad (28)$$

where the effective action  $S[\psi]$  reads as the integral of the squared left-hand side of Equation (3) over the variable  $z$ :  $S[\psi] = \int_0^L dz \left| \partial_z \psi - i\gamma |\psi|^2 \psi \right|^2$ . To calculate the path integral, we use the retarded discretization scheme which reflects the physics of the propagation process. The retarded discretization assumes that the derivation means  $(\partial_z \psi)(z_n) = (\psi(z_n) - \psi(z_{n-1}))/\Delta_z$ , where  $z_n = n\Delta_z$ ,  $\Delta_z = L/N_z$  ( $z_0 = 0$ ,  $z_{N_z} = L$ ), and any integral  $\int_0^L dz f(z)$  means  $\Delta_z \sum_{n=0}^{N_z} f(z_n)$ . The general approach for the derivation of the representation (28) and the argumentation for the retarded scheme one can find in Ref. [24].

For the first time, the conditional PDF  $P[Y|X]$  for the per-sample model was obtained in the form of infinite series in Ref. [4]:

$$P[Y|X] = \frac{1}{\pi Q} \sum_{m=-\infty}^{+\infty} e^{im(\phi^{(Y)} - \phi^{(X)} - \mu)} \frac{\exp\{-\frac{\rho^2 + \rho'^2}{Q} k_m \coth(k_m L)\}}{\sinh(k_m L)} k_m I_{|m|} \left( \frac{2k_m \rho \rho'}{Q \sinh(k_m L)} \right), \quad (29)$$

where the input signal  $X = \rho e^{i\phi^{(X)}}$ , the output signal  $Y = \rho' e^{i\phi^{(Y)}}$ ,  $\mu = \gamma L |X|^2$  is a dimensionless nonlinear parameter,  $k_m = \sqrt{Q m \gamma} e^{i\frac{\pi}{4}}$ , and  $I_{|m|}(z)$  is the modified Bessel function of the index  $|m|$ . The representation (29) was rederived using the path integral representation (28) in Refs. [17,18]. In Ref. [18] the authors applied two various methods: the recursive derivation based on the discretizing of the nondispersive NLSE and the properties of Markov chains (Chapman–Kolmogorov equation); and the derivation of the conditional PDF via the stochastic approach and Ito calculus [25].

Using the result (29), the lower bound of the capacity for the large signal power was found in Ref. [17]:

$$C \geq \frac{\log(\text{SNR})}{2} + \mathcal{O}(1). \quad (30)$$

Here the signal-to-noise ratio has the form

$$\text{SNR} = \frac{P}{QL}. \quad (31)$$

Recall that the noise power for the per-sample model is  $QL$ . To obtain the result (30) the authors demonstrate that at large-signal power the phase of the output signal occupies the entire phase interval  $[0, 2\pi]$  due to the interaction of the signal with the noise. As a result, the phase does not carry the information, see also [18]. Therefore, to find the lower bound for the large signal power it is necessary to take only the term with  $m = 0$  in Equation (29).

At the so-called intermediate power  $P$  range:

$$QL \ll P \ll (Q\gamma^2 L^3)^{-1}, \quad (32)$$



it is necessary to take into account terms with  $m \neq 0$  in the Equation (29). In Ref. [18], using Equation (29), the attempt was made to find the lower bound for the capacity in the intermediate power range. However due to the inconvenience of using of Equation (29) for analytical calculation this attempt was unsuccessful.

To calculate the mutual information in the intermediate power range we have to find the conditional probability density function (29) in a more convenient form. In Ref. [19] we found the method of the conditional probability density function calculation in the form of expansion in the parameter  $1/\text{SNR}$ . Using the developed method we found the conditional PDF in the convenient for further calculation form. The expansion in  $1/\text{SNR}$ , or in the small parameter  $Q$  similar to the semi-classical approximation (expansion in small Planck's constant  $\hbar$ ) in quantum mechanics.

Using the “semi-classical” method, see Ref. [26], we perform the change of integration variables (the simple shift with the Jacobian equals to unity) in Equation (28):

$$\psi(z) = \Psi_{cl}(z) + \tilde{\psi}(z), \quad (33)$$

and represent the conditional PDF (28) in the form:

$$P[Y|X] = \Lambda e^{-S[\Psi_{cl}(z)]/Q}, \quad (34)$$

where the normalization factor  $\Lambda$  has the form

$$\Lambda = \int_{\tilde{\psi}(0)=0}^{\tilde{\psi}(L)=0} \mathcal{D}\tilde{\psi} \exp \left\{ -\frac{S[\Psi_{cl}(z) + \tilde{\psi}(z)] - S[\Psi_{cl}(z)]}{Q} \right\}, \quad (35)$$

the function  $\Psi_{cl}(z)$  is the “classical” solution of the equation  $\delta S[\Psi_{cl}] = 0$  (here  $\delta S$  is the variation of the action  $S[\psi]$ ). The measure  $\mathcal{D}\tilde{\psi}$  for new variables  $\tilde{\psi}$  in Equation (34) is defined as

$$\mathcal{D}\tilde{\psi} = \left( \frac{1}{\Delta_z \pi Q} \right)^{N_z} \prod_{i=1}^{N_z-1} d\text{Re}\tilde{\psi}_i d\text{Im}\tilde{\psi}_i, \quad (36)$$

here  $\tilde{\psi}_i = \tilde{\psi}(z_i)$ ,  $\Delta_z = \frac{L}{N_z}$  is the grid spacing.

The Euler–Lagrange equation  $\delta S[\Psi_{cl}] = 0$  has the following explicit form

$$\frac{d^2 \Psi_{cl}}{dz^2} - 4i\gamma |\Psi_{cl}|^2 \frac{d\Psi_{cl}}{dz} - 3\gamma^2 |\Psi_{cl}|^4 \Psi_{cl} = 0. \quad (37)$$

The boundary conditions for the function  $\Psi_{cl}(z)$  are as follows

$$\Psi_{cl}(0) = X, \quad \Psi_{cl}(L) = Y. \quad (38)$$

Equation (37) is of great concern and we present two different approaches to its solution.

One can find the analytical solution of Equation (37) in the polar coordinate system [19]:  $\Psi_{cl}(z) = \rho(\zeta)e^{i\theta(\zeta)}$ ,  $\zeta = z/L$ . The solution depends on the following real integration constants:  $E$ ,  $\tilde{\mu}$ ,  $\zeta_0$  and  $\theta_0$ . These constants should be expressed from two boundary conditions (38). There are two different types of the solution. The first and the second type corresponds to the cases  $E \geq 0$  and  $E \leq 0$ , respectively. In the first case we have the solution:

$$\rho^2(\zeta) = \frac{1}{2L\gamma} \left( \tilde{\mu} + \sqrt{\tilde{\mu}^2 - k^2 \cos[2k(\zeta - \zeta_0)]} \right), \quad (39)$$

$$\begin{aligned}\theta(\zeta) = & \frac{\tilde{\mu}}{2}(\zeta - \zeta_0) + \sqrt{\tilde{\mu}^2 - k^2} \frac{\sin[2k(\zeta - \zeta_0)]}{4k} + \\ & \arctan \left[ (\tilde{\mu} - \sqrt{\tilde{\mu}^2 - k^2}) \frac{\tan[k(\zeta - \zeta_0)]}{k} \right] + \theta_0,\end{aligned}\quad (40)$$

where  $k = \sqrt{2E}$ . The solution (39), (40) is obtained under conditions  $\tilde{\mu} \geq k \geq 0$ . The integration constants  $\tilde{\mu}$ ,  $k$  and  $\zeta_0$  must be found from the boundary conditions:

$$|X|^2 = \rho^2(0) = \frac{\tilde{\mu} + \sqrt{\tilde{\mu}^2 - k^2} \cos[2k\zeta_0]}{2L\gamma}, \quad (41)$$

$$|Y|^2 = \rho^2(1) = \frac{\tilde{\mu} + \sqrt{\tilde{\mu}^2 - k^2} \cos[2k(1 - \zeta_0)]}{2L\gamma}, \quad (42)$$

$$\begin{aligned}\phi^{(X)} = \theta(0) = & -\frac{\tilde{\mu}}{2}\zeta_0 - \sqrt{\tilde{\mu}^2 - k^2} \frac{\sin[2k\zeta_0]}{4k} - \\ & \arctan \left[ (\tilde{\mu} - \sqrt{\tilde{\mu}^2 - k^2}) \frac{\tan[k\zeta_0]}{k} \right] + \theta_0,\end{aligned}\quad (43)$$

$$\begin{aligned}\phi^{(Y)} = \theta(1) = & \frac{\tilde{\mu}}{2}(1 - \zeta_0) + \sqrt{\tilde{\mu}^2 - k^2} \frac{\sin[2k(1 - \zeta_0)]}{4k} + \\ & \arctan \left[ (\tilde{\mu} - \sqrt{\tilde{\mu}^2 - k^2}) \frac{\tan[k(1 - \zeta_0)]}{k} \right] + \theta_0.\end{aligned}\quad (44)$$

If the integration constants are found then we can express the action in the form:

$$S[\Psi_{cl}(z; E, \tilde{\mu}, \zeta_0, \theta_0)] = \frac{k^2}{2\gamma L} \left( \tilde{\mu} - \sqrt{\tilde{\mu}^2 - k^2} \frac{\sin[2k(1 - \zeta_0)] + \sin[2k\zeta_0]}{2k} \right). \quad (45)$$

In the second case ( $E \leq 0$ ) the solution has the form

$$\rho^2(\zeta) = \frac{-\tilde{\mu} + \sqrt{\tilde{\mu}^2 + k^2} \cosh[2k(\zeta - \zeta_0)]}{2L\gamma}, \quad (46)$$

$$\begin{aligned}\theta(\zeta) = & -\frac{\tilde{\mu}}{2}(\zeta - \zeta_0) + \sqrt{\tilde{\mu}^2 + k^2} \frac{\sinh[2k(\zeta - \zeta_0)]}{4k} - \\ & \arctan \left[ (\tilde{\mu} + \sqrt{\tilde{\mu}^2 + k^2}) \frac{\tanh[k(\zeta - \zeta_0)]}{k} \right] + \theta_0,\end{aligned}\quad (47)$$

where  $k = \sqrt{-2E}$ . The integration parameters  $\tilde{\mu}$ ,  $k$ ,  $\zeta_0$ , and  $\theta_0$  are derived from the same procedure as in the first case. The action has the form

$$S[\Psi_{cl}(z; E, \tilde{\mu}, \zeta_0, \theta_0)] = \frac{k^2}{2\gamma L} \left( \tilde{\mu} + \sqrt{\tilde{\mu}^2 + k^2} \frac{\sinh[2k(1 - \zeta_0)] + \sinh[2k\zeta_0]}{2k} \right). \quad (48)$$

One can see that to obtain the solution  $\rho(\zeta)$ ,  $\theta(\zeta)$ , it is necessary to solve the system of nonlinear equations Equations (41)–(44) for integration constants. Of course, the system (41)–(44) can be solved numerically, but for the analytical calculation of the mutual information, it is necessary to develop the method which allows us to find the solution  $\Psi_{cl}(z)$  and action  $S[\Psi_{cl}(z)]$  analytically as a functional of the input  $X$  and output  $Y$  signals.

In Ref. [19] we have proposed the method based on the expansion of the semi-classical solution in the vicinity of the solution of the Equation (3) with zero noise. This makes sense when the noise power is much less than the signal power.

To demonstrate the approach, we find the solution of (37) in the leading order in the parameter  $1/\text{SNR}$ , linearizing Equation (37) in the vicinity of the solution  $\Psi_0(z)$ . The function  $\Psi_0(z)$  is the solution of the Equation (3) with zero noise. It obeys the input boundary condition  $\Psi_0(0) = X = \rho e^{i\phi^{(X)}}$ ,  $\rho = |X|$ . Note that we do not assume smallness of the nonlinearity. The function  $\Psi_0(z)$  reads

$$\Psi_0(z) = \rho \exp \left\{ i\mu \frac{z}{L} + i\phi^{(X)} \right\}, \quad (49)$$

where  $\mu = \gamma L \rho^2$  is the nonlinear dimensionless parameter. Let us represent the “classical” solution  $\Psi_{cl}(z)$  in the form

$$\Psi_{cl}(z) = \Psi_0(z) + \delta\Psi(z). \quad (50)$$

Here

$$\delta\Psi(z) = \varkappa(z) \exp \left\{ i\mu \frac{z}{L} + i\phi^{(X)} \right\}, \quad (51)$$

where the function  $\varkappa(z)$  is assumed to be much less than the  $\rho$ :  $|\varkappa(z)| \ll \rho$ , i.e.,  $\Psi_{cl}(z)$  is close to  $\Psi_0(z)$ . Note that the function  $\varkappa(z)$  depends on the output boundary conditions, therefore in a general case, the ratio  $|\varkappa(z)|/\rho$  can be of order of unity. The boundary conditions for the function  $\varkappa(z)$  are as follows:

$$\varkappa(0) = 0, \quad \varkappa(L) = Y e^{-i\phi^{(X)} - i\mu} - \rho \equiv x_0 + iy_0, \quad (52)$$

where  $x_0 = \text{Re}\{\varkappa(L)\}$  and  $y_0 = \text{Im}\{\varkappa(L)\}$ . It is important, that the configurations of  $\varkappa(z)$  at which  $\Psi_{cl}(z)$  significantly deviates from  $\Psi_0(z)$  are statistically irrelevant.

One can check that since the action achieves the absolute minimum ( $S[\Psi_0(z)] = 0$ ) on the solution  $\Psi_0(z)$ , the expansion of the action  $S[\Psi_0(z) + \delta\Psi(z)]$  starts from the quadratic terms for small  $\varkappa(z)$ :

$$S[\Psi_0(z) + \delta\Psi(z)] \propto \varkappa^2(z).$$

Therefore, the exponent  $e^{-S[\Psi_{cl}(z)]/Q}$  and, as a result, the conditional PDF  $P[Y|X]$  decreases exponentially when  $\varkappa(z) \gg \sqrt{QL}$ .

The next step in evaluation of the conditional probability  $P[Y|X]$  is the calculation of the path integral (35). To calculate the integral (35) in the leading order in parameter  $1/\sqrt{\text{SNR}}$  one should retain only terms quadratic in the function  $\tilde{\psi}$  in the integrand. Any extra powers of  $\tilde{\psi}$  or  $\varkappa$  lead to suppression in the multiplicative parameter  $\sqrt{QL}$ , since for small  $Q$  the dominant contribution to the path integral comes from region where  $\tilde{\psi} \sim \sqrt{QL}$ . Since we calculate the the integral (35) in the leading order in the parameter  $Q$ , the function  $\Psi_{cl}(z)$  can be replaced by the function  $\Psi_0(z)$  in the exponent  $\exp \{ -(S[\Psi_{cl}(z) + \tilde{\psi}(z)] - S[\Psi_{cl}(z)]) / Q \}$ .

To find the next-to-leading order corrections in the parameter  $1/\sqrt{\text{SNR}}$  to the conditional PDF  $P[Y|X]$  one should keep both  $\varkappa(z)$  in  $\Psi_{cl}(z)$  and higher powers of  $\tilde{\psi}$  in the exponent in Equation (35). Details of

the path integral calculation in the leading and next-to-leading orders in  $1/\sqrt{\text{SNR}}$  one can find in Ref. [19]. Here we present only the result obtained in the leading order in the parameter  $1/\sqrt{\text{SNR}}$ :

$$P[Y|X] = \frac{\exp\left\{-\frac{(1+4\mu^2/3)x_0^2 - 2\mu x_0 y_0 + y_0^2}{QL(1+\mu^2/3)}\right\}}{\pi QL\sqrt{1+\mu^2/3}}, \quad (53)$$

where  $x_0 = \text{Re}\{Ye^{-i\phi^{(X)}} - i\mu - \rho\}$  and  $y_0 = \text{Im}\{Ye^{-i\phi^{(X)}} - i\mu - \rho\}$ . One can see that the expression (53) is much simpler than the exact result (29). Note that the distribution (53) has the following property

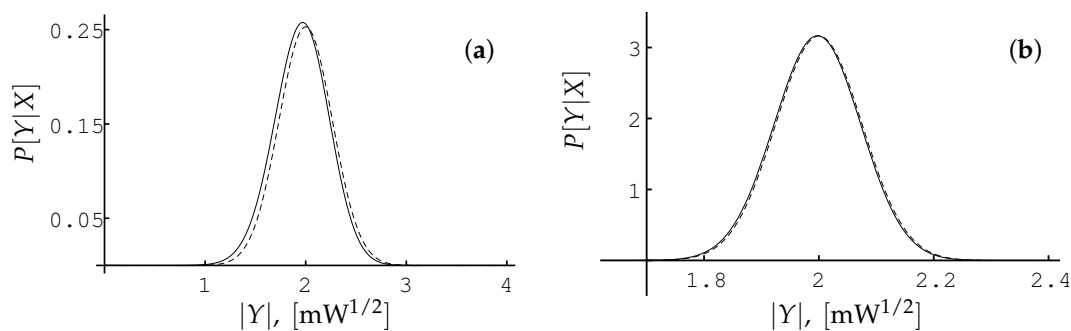
$$\lim_{Q \rightarrow 0} P[Y|X] = \delta(Y - \Psi_0(L)). \quad (54)$$

The limit (54) is the deterministic limit of  $P[Y|X]$  in the absence of noise. Also Equation (53) has the correct limit for small  $\gamma$ :

$$\lim_{\gamma \rightarrow 0} P[Y|X] = \frac{e^{-|Y-X|^2/QL}}{\pi QL}, \quad (55)$$

where the right-hand-side is the conditional PDF for the linear nondispersive channel with the additive noise.

Let us compare the result obtained in the leading in  $1/\sqrt{\text{SNR}}$  order (53) with exact result (29). In Figure 1 we plot the PDF  $P[Y|X]$  as the function of  $|Y|$  for  $X = 2 \text{ mW}^{1/2}$ ,  $\arg(Y) = \mu$ ,  $\mu = \gamma L|X|^2 = 4$ , so we choose  $\gamma L = \text{mW}^{-1}$ , and for two values of parameter  $QL$ :  $QL = 1/2 \text{ mW}$ ,  $QL = 1/25 \text{ mW}$  (it corresponds to  $\text{SNR} = 8$  and  $\text{SNR} = 100$ , respectively).



**Figure 1.** The dependence of the probability density function  $P[Y|X]$  on  $|Y|$  for  $X = 2 \text{ mW}^{1/2}$  and the argument  $\arg(Y) = \mu = 4$ . The plot (a) corresponds to signal-to-noise power ratio ( $\text{SNR}$ ) = 8, the plot (b) corresponds to  $\text{SNR} = 100$ . The solid line and dashed line correspond to the exact expression (29) and approximation (53), correspondingly.

One can see the good agreement between the exact result (29) and the approximation (53) even for  $\text{SNR} = 8$ . For  $\text{SNR} = 100$  the approximation almost coincides with the exact result. In the case when the  $\text{SNR} \approx 10^2$ – $10^4$ , which corresponds to optical fiber channels, the difference between the approximation and the exact result is of the order of  $1/\sqrt{\text{SNR}}$ . To decrease this difference we should calculate the corrections of the order of  $1/\sqrt{\text{SNR}}$  and  $1/\text{SNR}$ , see the Section 3.1.3.

### 3.1.2. Probability Density Function $P_{out}[Y]$

Let us consider the integral (21) which defines the the probability density function of the output signal:

$$P_{out}[Y] = \int \mathcal{D}X P[Y|X] P_X[X], \quad (56)$$

where the distribution  $P_X[X]$  is a smooth function. Since the input signal power is  $P$ , we can expect that the function  $P_X[X]$  changes on the scale  $X \sim \sqrt{P}$  which is assumed to be much greater than  $\sqrt{QL}$ . For this case we can use the Laplace's method [27] to calculate the integral (56) up to the terms proportional to the noise power  $QL$ , for details of the calculation see Appendix C in Ref. [19]. The idea of the integral (56) calculation is based on the fact that the function  $P[Y|X]$  is much narrower than the function  $P_X[X]$  (the function  $P[Y|X]$  is almost Dirac delta-function for the function  $P_X[X]$ ). Therefore, the calculation of the integral is simple and the result has the form [19]:

$$P_{out}[Y] = \int \mathcal{D}X P[Y|X] P_X[X] \approx P_X \left[ Y e^{-i\gamma|Y|^2 L} \right]. \quad (57)$$

When obtaining the result (57) it is not required to pass to the limit  $Q \rightarrow 0$  but only the relation  $P \gg QL$  between the scales  $P$  and  $QL$  is used.

For the class of the distributions  $P_X[X]$  depending only on absolute value  $|X|$  we have  $P_{out}[Y] = P_X[|Y|]$ . For such distributions we can calculate corrections to (57) in the parameter  $QL$  in any order in  $QL$ .

Let us restrict our consideration in the remainder of this sub-subsection to the case of the distributions  $P_X[X]$  depending only on  $|X|$ . We can use conditional PDF  $P[Y|X]$  (29) found in Ref. [17]. In this case the function  $P_{out}[Y]$  depends only on  $|Y| = \rho'$ :

$$P_{out}[\rho'] = \frac{2e^{-\rho'^2/(QL)}}{QL} \int_0^\infty d\rho \rho e^{-\rho^2/(QL)} I_0 \left( \frac{2\rho\rho'}{QL} \right) P_X[\rho]. \quad (58)$$

Using this formula one can obtain the simple relation for  $P_{out}[\rho']$  within the perturbation theory in the parameter  $QL$ . Performing the zero order Hankel transformation [27]:

$$\hat{P}[k] = \int_0^\infty d\rho \rho J_0(k\rho) P_X[\rho] \quad (59)$$

for both sides of Equation (58), and using the standard integral with the Bessel  $J_\nu(x)$  and modified Bessel functions [28]

$$\int_0^\infty dz z e^{-pz^2} J_\nu(bz) I_\nu(cz) = \frac{1}{2p} J_\nu \left( \frac{bc}{2p} \right) e^{\frac{c^2-b^2}{4p}},$$

we arrive at the relation between the Hankel images of the output and input signal PDFs:

$$\hat{P}_{out}[k] = e^{-k^2 \frac{QL}{4}} \hat{P}[k]. \quad (60)$$

Then, we perform the inverse Hankel transformation

$$P_X[\rho] = \int_0^\infty dk k J_0(k\rho) \hat{P}[k], \quad (61)$$

and obtain the following important result

$$P_{out}[\rho] = e^{\frac{QL}{4}\Delta_\rho} P_X[\rho], \quad (62)$$

where  $\Delta_\rho = \frac{d^2}{d\rho^2} + \frac{1}{\rho} \frac{d}{d\rho}$  is the two-dimensional radial Laplace operator. The relation (62) allows us to find the corrections of orders of  $(QL)^n$  to  $P_{out}[\rho]$  by the expansion of the exponent and calculation of the action of the differential operator  $\Delta_\rho^n$  on the input PDF  $P_X[\rho]$ .

Let us consider the widely used example of the modified Gaussian distribution with one parameter  $\beta$ :

$$P_X^{(\beta)}[X] = \frac{\exp\{-\beta|X|^2/(2P)\} |X|^{\beta-2}}{\pi\Gamma(\beta/2) (2P/\beta)^{\beta/2}}, \quad (63)$$

where  $\Gamma(x)$  is the Euler gamma function. In the case of  $\beta > 0$  the distribution  $P_X^{(\beta)}[X]$  obeys the conditions:  $2\pi \int_0^\infty d\rho \rho P_X^{(\beta)}[\rho] = 1$ ,  $2\pi \int_0^\infty d\rho \rho^3 P_X^{(\beta)}[\rho] = P$ . The last one means that the average power of the input signal is equal to  $P$ . The generalized distribution  $P_X^{(\beta)}[X]$  for  $\beta = 1$  is called as the half-Gaussian distribution

$$P_X^{(1)}[X] = \frac{\exp\{-|X|^2/(2P)\}}{\pi\sqrt{2\pi P}|X|}, \quad (64)$$

and the Gaussian distribution for  $\beta = 2$ :

$$P_X^{(2)}[X] = \frac{e^{-|X|^2/P}}{\pi P}. \quad (65)$$

Inserting (63) into Equation (58) we arrive at the standard integral, see [28], and obtain:

$$P_{out}^{(\beta)}[Y] = {}_1F_1\left(\frac{\beta}{2}; 1; \frac{|Y|^2 2P}{QL(2P + \beta QL)}\right) \frac{\exp\{-|Y|^2/QL\}}{\pi QL} \left(\frac{\beta QL}{2P + \beta QL}\right)^{\beta/2}, \quad (66)$$

where  ${}_1F_1(\beta/2; 1; z)$  is the confluent hypergeometric function. The function reduces to  $e^z$  for the case of the Gaussian distribution, and to the expression  $e^{z/2} I_0(z/2)$  for the case of the half-Gaussian one:

$$P_{out}^{(2)}[Y] = \frac{e^{-|Y|^2/(P+QL)}}{\pi(P+QL)}, \quad (67)$$

$$P_{out}^{(1)}[Y] = \frac{1}{\pi\sqrt{QL(2P+QL)}} I_0\left(\frac{|Y|^2 P}{QL(2P+QL)}\right) \exp\left\{-\frac{|Y|^2(P+QL)}{QL(2P+QL)}\right\}. \quad (68)$$

The result (57) can be obtained from Equation (68) in the case  $QL \ll |Y|^2 \sim P$ . Expanding the right-hand side of the Equation (67) in the parameter  $QL/P$  we obtain

$$P_{out}^{(1)}[Y] \approx P_X^{(1)}[|Y|] \quad (69)$$

with accuracy  $\mathcal{O}(QL)$ . The result (69) coincides with the general relation (57).

### 3.1.3. Lower Bound for the Channel Capacity

The estimates for the capacity of the per-sample model in the regime of very large SNR were obtained in Ref. [17]. To specify, it was considered the case  $P \gg \left(P_{\text{noise}}\gamma^2 L^2\right)^{-1}$ , where  $P_{\text{noise}} = QL$  is the noise power. In Ref. [17] the lower bound for the capacity of the per-sample channel was found. Using the trial half-Gaussian input signal PDF (64) authors obtained the following result:

$$C \geq \frac{\log(\text{SNR})}{2} + \frac{1 + \gamma_E - \log(4\pi)}{2} + \mathcal{O}\left(\frac{\log(\text{SNR})}{\text{SNR}}\right), \quad (70)$$

where  $\gamma_E \approx 0.5772$  is the Euler constant. The second term on the right-hand side of Equation (30) was presented as  $\mathcal{O}(1)$  in Ref. [17] but it can be found using Equations (23) and (24) of Ref. [17]. Comparing the result (70) with the Shannon result (1) it is worth noting that the pre-logarithmic factor  $1/2$  differs from the unity in the Shannon result. The physical meaning of the difference is that the signal's phase does not carry information when the power is very large:  $P \gg \left(P_{\text{noise}}\gamma^2 L^2\right)^{-1}$ , see Ref. [17].

The most interesting power regime for the per-sample model is so-called intermediate power range defined in (32). In the regime we have on the one hand the parameter SNR is large,  $P_{\text{noise}} \ll P$ , and on the other hand the signal-dependent phase does not yet occupy the entire phase interval  $[0, 2\pi]$  due to the signal-to-noise interaction, i.e., the phase still carries information.

Capacity estimates in the intermediate power range (32) were presented in Ref. [18]. For such a power  $P$  the authors of this paper used the half-Gaussian input signal PDF (64) for the estimate of the lower bound for the capacity as well, see inequality (40) in Ref. [18]. But there were some flaws in the derivation of this inequality in Ref. [18], see the discussion in the Introduction of Ref. [19]. In our approach presented in Ref. [19] we solved a variational problem for the mutual information (24) with the normalization (22) and power restriction (26): we found both the optimal input signal distribution  $P_X[X]$  maximizing the mutual information (24) in the leading and next-to-leading orders in  $1/\sqrt{\text{SNR}}$  in the intermediate power range. Let us proceed with this calculation.

To begin with, when the parameter  $\text{SNR} \gg 1$  we can calculate the output signal entropy  $H[Y]$  by substituting  $P_X[Y \exp\{-i\gamma|Y|^2 L\}]$  to Equation (20) instead of  $P_{\text{out}}[Y]$  due to the relation (57), and then performing the change of the integration variable  $\phi = \phi^{(Y)} + \gamma|Y|^2 L$  we obtain

$$H[Y] \approx - \int_0^{2\pi} d\phi \int_0^\infty d\rho' \rho' P_X[\rho' e^{i\phi}] \log P_X[\rho' e^{i\phi}]. \quad (71)$$

Note that the output signal entropy in the form (71) coincides with the input signal entropy  $H[X]$  with the accuracy  $\mathcal{O}(QL)$ .

Secondly, to obtain the conditional entropy  $H[Y|X]$  we substitute the conditional PDF  $P[Y|X]$  in the form of Equation (53) into Equation (19). Then we change the integration variables  $\mathcal{D}Y \equiv d\text{Re}Y d\text{Im}Y$  to  $dx_0 dy_0$ , and perform the integration over  $x_0, y_0$ . The result has the form:

$$H[Y|X] = \log(e\pi QL) + \int_0^{2\pi} d\phi \int_0^\infty d\rho \rho P_X[\rho e^{i\phi}] \log \sqrt{1 + \gamma^2 L^2 \rho^4 / 3}. \quad (72)$$



Thirdly, to find the optimal input signal distribution  $P_X^{opt}[X]$  we solve the variational problem for the functional  $J[P_X, \lambda_1, \lambda_2]$ :

$$J[P_X, \lambda_1, \lambda_2] = H[Y] - H[Y|X] - \lambda_1 \left( \int \mathcal{D}X P_X[X] - 1 \right) - \lambda_2 \left( \int \mathcal{D}X P_X[X] |X|^2 - P \right), \quad (73)$$

where  $\lambda_{1,2}$  are Lagrange multipliers which corresponds to the normalization condition (22) and the condition (26) of the fixed average signal power  $P$ . The solution  $P_X^{opt}[X]$  of the corresponding Euler–Lagrange equations for (73) referred to as the “optimal” distribution:

$$P_X^{opt}[X] = N_0(P) \frac{e^{-\lambda_0(P)|X|^2}}{\sqrt{1 + \gamma^2 L^2 |X|^4/3}}, \quad (74)$$

where coefficients  $N_0(P)$  and  $\lambda_0(P)$  are determined from the conditions:

$$\int \mathcal{D}X P_X^{opt}[X] = 2\pi N_0(P) \int_0^\infty \frac{d\rho \rho e^{-\lambda_0(P)\rho^2}}{\sqrt{1 + \gamma^2 L^2 \rho^4/3}} = 1, \quad (75)$$

$$\int \mathcal{D}X P_X^{opt}[X] |X|^2 = 2\pi N_0(P) \int_0^\infty \frac{d\rho \rho^3 e^{-\lambda_0(P)\rho^2}}{\sqrt{1 + \gamma^2 L^2 \rho^4/3}} = P. \quad (76)$$

Note that in the leading order in the parameter  $1/\text{SNR}$  the function  $P_X^{opt}[X]$  depends only on  $|X|$ . In the next-to-leading the order in  $1/\text{SNR}$  this property holds true as well [20].

In the parametric form the power dependance of the parameters  $\lambda_0$  and  $N_0$  reads

$$\lambda_0(P) = \frac{\gamma L}{\sqrt{3}} \alpha, \quad N_0(P) = \frac{\gamma L}{\pi \sqrt{3} G(\alpha)}, \quad (77)$$

where  $G(\alpha) = \frac{\pi}{2} (H_0(\alpha) - Y_0(\alpha))$ ,  $Y_0(\alpha)$  and  $H_0(\alpha)$  are Neumann and Struve functions of zero order, respectively. The parameter  $\alpha(P)$  is the real solution of the equation

$$\frac{d}{d\alpha} \log G(\alpha) = -\gamma L P / \sqrt{3}.$$

Note that the optimal input signal distribution  $P_X^{opt}[X]$  (74) differs from the half-Gaussian distribution (64).

For sufficiently large values of the power  $P$ ,  $\log(\gamma P L) \gg 1$ , we use the asymptotics of  $Y_0(\alpha)$  and  $H_0(\alpha)$  at small  $\alpha$  and arrive at the following result for  $\lambda_0(P)$  and  $N_0(P)$ :

$$\lambda_0(P) \approx \frac{1 - \log \log(B\tilde{\gamma}) / \log(B\tilde{\gamma})}{P \log(B\tilde{\gamma})}, \quad N_0(P) \approx \frac{\tilde{\gamma}}{\pi P} \log^{-1} [B\tilde{\gamma} / (P\lambda_0(P))]. \quad (78)$$

Here and below we use the notation

$$B = 2e^{-\gamma E}, \quad (79)$$

$$\tilde{\gamma} = \gamma L P / \sqrt{3} \quad (80)$$

is the convenient dimensionless nonlinear parameter. At small  $P$ , such that the nonlinearity parameter  $\tilde{\gamma} \ll 1$ , the solutions of the Equations (75) and (76) have the form:

$$\lambda_0(P) \approx \frac{1 - 2\tilde{\gamma}^2}{P}, \quad N_0(P) \approx \frac{1 - \tilde{\gamma}^2}{\pi P}. \quad (81)$$

Note that at  $\tilde{\gamma} \rightarrow 0$  the optimal distribution (74) passes to the Gaussian distribution (65). It is known that this distribution is optimal for a linear channel [1]. In Ref. [20] using the same method we found the first correction to  $P_X^{opt}[X]$  proportional to  $QL$ .

The fourth, to calculate the mutual information we substitute the expression (74) for  $P_X^{opt}[X]$  in Equations (71) and (72) and using the definition (24) we obtain

$$I_{P_X^{opt}[X]} = C_0 = P\lambda_0(P) - \log N_0(P) - \log(\pi eQL). \quad (82)$$

The last equation gives the mutual information  $I_{P_X^{opt}[X]}$  with the accuracy  $\mathcal{O}(QL)$ .

At small  $\tilde{\gamma}$  we obtain

$$I_{P_X^{opt}[X]} \approx \log(\text{SNR}) - \tilde{\gamma}^2. \quad (83)$$

This result is the Shannon capacity  $\log(1 + \text{SNR})$  at large SNR of the linear channel (1) with the first nonlinear correction.

In the high power sub-interval  $(\gamma L)^{-1} \ll P \ll (QL^3\gamma^2)^{-1}$  using Equation (78) one can obtain the following asymptotics for the mutual information in the case of very large nonlinearity parameter ( $\log(\tilde{\gamma}) \gg 1$ ):

$$I_{P_X^{opt}[X]} = \log \log(B\gamma LP/\sqrt{3}) - \log(QL^2\gamma e/\sqrt{3}) + \frac{1}{\log(B\gamma LP/\sqrt{3})} \left[ \log \log(B\gamma LP/\sqrt{3}) + 1 - \frac{\log \log(B\gamma LP/\sqrt{3})}{\log(B\gamma LP/\sqrt{3})} \right]. \quad (84)$$

This expression is obtained with the accuracy  $1/\log^2(\tilde{\gamma})$ . One can see that the first term on right-hand side of the Equation (84) grows as  $\log \log P$ . It means that the mutual information  $I_{P_X^{opt}[X]}$  also grows as  $\log \log P$  at large enough  $P$ .

Note that the results (74), (82), and asymptotics (78), (81), (83), (84) are obtained in the leading and next-to-leading order in the parameter  $1/\sqrt{\text{SNR}}$ . Therefore, the results (74), (82) are calculated with the accuracy  $\mathcal{O}(1/\text{SNR})$ . However, in the literature the bounds of the capacity rather than the asymptotic estimates are on the carpet. Therefore, to find the lower bound of the channel capacity for the per-sample model it is necessary to calculate the corrections of the order of  $1/\text{SNR}$  and define their signs. Moreover, to find the applicability region of the result (82) we have to know the corrections of the order of  $1/\text{SNR}$  as well. The applicability region will be defined by the condition that the corrections of the order of  $1/\text{SNR}$  are much less than obtained results (74), (82). In Ref. [20] we have calculated these corrections using the approach described in Section 3.1. The calculation of these corrections is straightforward but cumbersome, therefore, here we present only the idea of the calculation and demonstrate the results for these corrections for the mutual information. The detailed calculation can be found in Ref. [20].

To calculate the correction to the mutual information we should know the corrections to the conditional probability density function (53). Therefore, we should calculate the corrections both to the action  $S$  and to the normalization factor  $\Lambda$  in Equation (34). To find these corrections we have to calculate the function

$\varkappa(z)$ , see Equation (51), in the leading, next-to-leading, and next-to-next-to-leading orders in the parameter  $1/\sqrt{\text{SNR}}$ . Then we should substitute the found corrections for the action  $S$  to the path integral (35), and calculate the path integral up to the terms which are proportional to the parameter  $Q$ . After that, we expand the product of the exponent and the normalization factor  $\Lambda$  up to the terms of order of  $1/\text{SNR}$ . This expression is cumbersome, and therefore, we do not present it here, but it can be found in Ref. [20]. To calculate the corrections to the mutual information we substitute the obtained result for  $P[Y|X]$  to the Equations (19)–(21), (24). Then, using the method described above, we perform maximization of the mutual information calculated with the accuracy  $1/\text{SNR}$ , and obtain the following result for it:

$$I_{P_X^{opt}[X]}^{(1)} = C_0 + \Delta C, \quad (85)$$

where  $C_0$  is defined in (82). The correction  $\Delta C$  has the form:

$$\Delta C = \frac{1}{\text{SNR}} \left\{ \pi N_0 P \left[ \frac{214}{375} - \frac{8}{375} \left( \frac{\lambda_0 P}{\tilde{\gamma}} \right)^2 \right] + \lambda_0 P \left[ \frac{137}{150} + \frac{8}{375} \left( \frac{\lambda_0 P}{\tilde{\gamma}} \right)^2 \right] - \frac{347}{750} (\lambda_0 P)^2 \right\}. \quad (86)$$

The quantity  $\Delta C$  corresponds to the first non-vanishing correction to the mutual information. One can verify that for the small parameter  $\gamma L^2 Q \ll 1$ , the correction (86) is always small with respect to  $C_0$ . Indeed, the ratio of the expression in the curly brackets in Equation (86) and  $\tilde{\gamma}$  is the bounded function for all values of  $\tilde{\gamma}$ .

We do not have the explicit analytical result for the correction  $\Delta C$  for the arbitrary parameter  $\tilde{\gamma}$ , since the quantities  $\lambda_0$  and  $N_0$  are the solutions of the nonlinear equations. But the correction  $\Delta C$  can be calculated numerically for any parameter  $\tilde{\gamma}$ . However, for the small and large parameters  $\tilde{\gamma}$  we can calculate the asymptotics of the quantity  $\Delta C$  analytically.

For the small nonlinearity  $\tilde{\gamma} \ll 1$  we substitute the parameters  $\lambda_0$  and  $N_0$  in the form (81) to the Equation (86) and obtain:

$$\Delta C \approx \frac{1}{\text{SNR}} - \frac{1}{\text{SNR}} \frac{\tilde{\gamma}^2}{3}. \quad (87)$$

Using this result and the asymptotics (83) we obtain the mutual information within our accuracy in the form:

$$I_{P_X^{opt}[X]}^{(1)} \approx \log(1 + \text{SNR}) - \tilde{\gamma}^2 - \frac{1}{\text{SNR}} \frac{\tilde{\gamma}^2}{3}. \quad (88)$$

One can see that the first term in the right-hand side is the the capacity of the linear channel, the second and the third ones correspond to the nonlinear corrections. The nonlinear corrections in (88) are negative and they decrease the mutual information  $I_{P_X^{opt}[X]}^{(1)}$  of the channel.

Let us consider the behavior of the correction  $\Delta C$  at large power  $P$ . For the case  $\log(\gamma LP) \gg 1$  (but  $P \ll (\gamma^2 QL^3)^{-1}$  to be within the intermediate power range) we obtain the simple result:

$$\Delta C \approx \frac{1}{\text{SNR}} \frac{214}{375} \pi N_0 P. \quad (89)$$

Using the asymptotic (78) for quantity  $N_0$  we arrive at the expression

$$\Delta C \approx \frac{\gamma L^2 Q}{\sqrt{3}} \frac{214}{375} \left( \log(B\tilde{\gamma} \log(B\tilde{\gamma})) + \frac{\log \log(B\tilde{\gamma})}{\log(B\tilde{\gamma})} \right)^{-1}. \quad (90)$$

Note that this correction is suppressed parametrically as  $\gamma L^2 Q$  instead of  $1/\text{SNR} = QL/P$ . One can see the correction decreases at large  $\tilde{\gamma}$  as  $1/\log \tilde{\gamma}$ . It is interesting that at large  $\tilde{\gamma}$  the correction  $\Delta C$  is positive, therefore, it slightly enhances the mutual information  $I_{P_X^{opt}[X]}^{(1)}$ . Since the correction  $\Delta C$  is positive in the region defined as  $\log(\gamma LP) \gg 1$ ,  $P \ll (\gamma^2 QL^3)^{-1}$ , and the next-to-leading corrections to the mutual information are suppressed parametrically, the quantity  $C_0$  is the lower bound of the per-sample channel capacity:

$$C \geq C_0. \quad (91)$$

Since there are no corrections of order of  $\gamma^2 L^3 QP$  at large  $P$ , see Equation (90), we expect the next correction containing the power  $P$  to be of order of  $(\gamma^2 L^3 QP)^2$ , see Ref. [19]. Therefore, the applicability region at large  $P$  for the quantity  $C_0$  is determined by the condition  $(\gamma^2 L^3 QP)^2 \ll 1$ . For the given small parameter  $\gamma^2 L^3 QP$  this condition extends the applicability region for the lower bound of the channel capacity  $C_0$ . For realistic channel parameters presented in the Table 1.

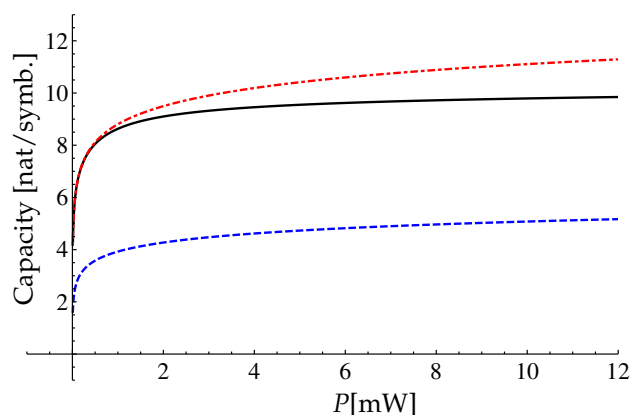
**Table 1.** Channel parameters for the per-sample model.

$\gamma$ [ (km × mW) <sup>−1</sup> ]	$Q$ [ mW/(km) ]	$L$ [ km ]
$10^{-3}$	$1.5 \times 10^{-7}$	$10^3$

We have the following intermediate power range:

$$1.5 \times 10^{-4} \text{ mW} \ll P \ll 0.66 \times 10^4 \text{ mW}. \quad (92)$$

One can see that the range is very wide. For the presented parameters we have numerically calculated the lower bound  $C_0$  using Equations (77) and (82). The result of the calculation is presented in Figure 2. Also in Figure 2 we present the comparison of the approximation (85) with the Shannon capacity of a linear channel and with the asymptotic capacity bound (70).

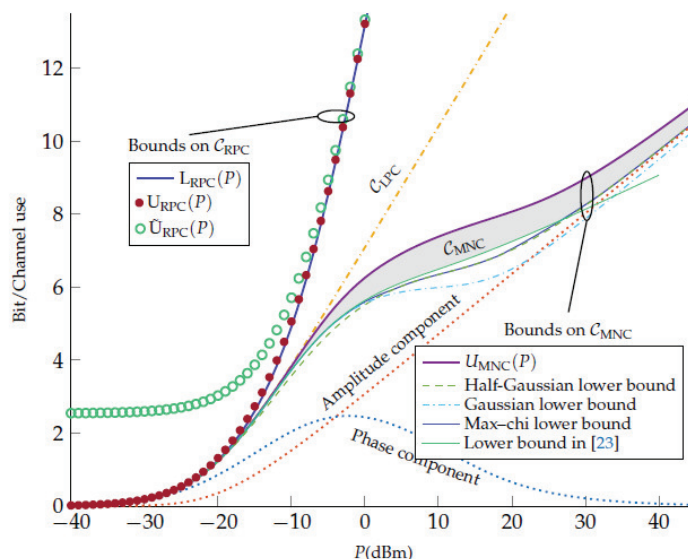


**Figure 2.** Shannon capacity, the lower bound  $C_0$ , and the asymptotic capacity bound (70) for the channel parameters from the Table 1. The red dashed-dotted line corresponds to the Shannon limit  $\log(1 + \text{SNR})$ , the black solid line corresponds to the lower bound  $C_0$ , see Equation (82), the blue dashed line corresponds to the bound (70).

One can see that the Shannon capacity of the linear channel with the additive noise (the red dashed-dotted line in the Figure 2) is always greater than the lower bound (82) for the nondispersive

nonlinear fiber channel (the black solid line in the Figure 2) for the considered range of  $P$ . But the lower bound (82) is greater than the asymptotic capacity bound (70) in the intermediate power range.

In Ref. [22] the comparison of our result (82) for the NLSE per-sample model (authors referred to our model as the memoryless NLS channel (MNC)) with two other models, more precisely, the regular perturbative channel (RPC) and the logarithmic perturbative channel (LPC), was performed. The comparison was illustrated in Figure 1 of Ref. [22], and we present this figure in the Figure 3.



**Figure 3.** [This figure is taken from Ref. [22]]. Capacity bounds for the RPC and LPC models of Ref. [22], together with the capacity of the per-sample NLSE model. The channel parameters are as follows:  $\gamma = 1.27 \text{ W}^{-1} \text{ km}^{-1}$ ,  $L = 5000 \text{ km}$ ,  $QL = 7.36 \times 10^{-3} \text{ mW}$ . Our result (82) is presented by the green solid line (“lower bound in [23]”). The intermediate power range goes out from  $P \sim -20 \text{ dBm}$  to  $P \sim 6 \text{ dBm}$ .

The authors claimed that they established a novel upper bound on the capacity of the NLSE per-sample model (the violet curve  $\mathcal{U}_{MNC}(P)$  in Figure 3). In addition, the authors considered various input signal distributions within MNC: half-Gaussian (64), Gaussian (65), and modified Gaussian distribution (63) optimized in parameter  $\beta$  (it is denoted as “Max-chi” in Figure 3). They used the following channel parameters:  $\gamma = 1.27 \text{ W}^{-1} \text{ km}^{-1}$ ,  $L = 5000 \text{ km}$ ,  $QL = 7.36 \times 10^{-3} \text{ mW}$ . For these parameters the upper limit of the intermediate power range (we choose it as  $P_{max} = 6\pi^2(Q\gamma^2L^3)^{-1}$ , see [18]) is estimated as  $P_{max} = 0.2 \text{ W} = 23 \text{ dBm}$ . It is obvious in Figure 3, that up to this power  $P_{max}$  our result (green solid line) is consistent with the capacity upper bound of the per-sample model (the violet curve  $\mathcal{U}_{MNC}$ ) and it exceeds other mutual information curves for other input signal distributions in all intermediate power range. Of course, for large powers ( $P \gtrsim P_{max}$ ) our input signal distribution (74) is not the optimal any more, and the mutual information (85) underestimates the real capacity. In the strict sense, the corrections to (85) are small only up to  $P \sim 5.5 \text{ dBm}$ , and when  $P > 5.5 \text{ dBm}$  we have the transition range from the intermediate power range with the optimal input signal distribution (74) to the large power semirange where the optimal distribution is believed to be the half-Gaussian one, see [17,18]. The large extension of the transition range (up to  $P_{max} \sim 23 \text{ dBm}$ ) can be explained by the smallness of the next-to-leading order corrections (86) and its decreasing when increasing the power, see Equation (90).

To finalize the per-sample channel consideration we emphasize the main results. We developed the path integral approach to the calculation of the conditional PDF  $P[Y|X]$  for large SNR. We demonstrated

that for the nonlinear nondispersive channel the lower bound  $C_0$  of the capacity increases only as  $\log \log P$  at large signal power  $P$  instead of the behavior  $\log P$  that is specific for the channel with zero nonlinearity. To determine the applicability region and the accuracy of the found quantity  $C_0$  we calculated the first non-zero correction  $\Delta C$  proportional to the noise power  $QL$  in the intermediate power range  $QL \ll P \ll (\gamma^2 L^3 Q)^{-1}$ . We demonstrated that the quantity  $\Delta C$  is small in the intermediate power range, and it is the positive decreasing function at large signal power  $P$ :  $(\gamma L)^{-1} \ll P \ll (\gamma^2 L^3 Q)^{-1}$ . The found result is in agreement with recent results of Ref. [22].

### 3.2. Extended Model: Considerations of the Time Dependent Input Signals

Let us start this section from the consideration of the conditional probability density function  $P[Y|X]$ . In the case of the nontrivial time dependance of the input  $X(t)$  and output  $Y(t)$  signals the conditional PDF reads Ref. [11]:

$$P[Y(t)|X(t)] = \int_{\psi(0,t)=X(t)}^{\psi(L,t)=Y(t)} \mathcal{D}\psi e^{-S[\psi]/Q}, \quad (93)$$

where the effective action  $S[\psi]$  has the form:  $S[\psi] = \int_0^L dz \int dt \left| \partial_z \psi(z,t) - i\gamma |\psi(z,t)|^2 \psi(z,t) \right|^2$ ; the integration measure  $\mathcal{D}\psi(z,t)$  depends on both  $z$ - and  $t$ - discretization scheme:

$$\mathcal{D}\psi(z,t) = \left( \frac{\Delta}{\Delta_z \pi Q} \right)^{2M} \prod_{j=-M}^{M-1} \prod_{i=1}^{N_z-1} \left\{ \frac{\Delta}{\Delta_z \pi Q} d\text{Re}\psi(z_i, t_j) d\text{Im}\psi(z_i, t_j) \right\}, \quad (94)$$

where  $\Delta = T/(2M) = (2N+1)T_0/(2M)$  is the time grid spacing and  $\Delta_z = L/N_z$  is the  $z$ -coordinate grid spacing. Of course, the expression (93) contains all information about the transmitted signal and its interaction with the noise, but as strange as it sounds, the expression (93) contains redundant information (i.e., degrees of freedom which cannot be detected). The point is that in the realistic communication channel the receiver has the finite bandwidth, it means that it reduces somehow the bandwidth of the received signal  $Y(t)$ . After that, there is a procedure of the extraction of the information from the signal measured by receiver. This detection procedure should be implemented in the function  $P[Y(t)|X(t)]$ .

To demonstrate the point, let us consider the input signal  $X(t)$  in the form (11). In such form, the number of the degrees of freedom in the path integral (93) is infinite, if the function  $X(t)$  is continuous, or  $2M$  degrees of freedom if we set the function  $X(t)$  in the discrete-time moments  $t_i$ , see the text after Equation (17). However, the transmitted information is carried only by  $2N+1$  coefficients  $C_n$  ( $N \ll M$ ), see Equation (11). It means that to obtain the conditional probability density function which describes the transmission of the information carried by the set of the coefficients  $C_n$  we have to integrate over the redundant degrees of freedom. The approach based on the path integral representation (93) is general and it can be used for any signal model, any receiver and projecting procedure (18). For our signal model described in the Section 2.2, we can use a simpler method of the calculation of the conditional PDF  $P[\{\tilde{C}\}|\{C\}]$ . Below we describe the method.

In our model, the signal propagation for different time moments  $t_j$  is independent because the dispersion is zero and the noise is not correlated for different time moments  $t_i \neq t_j$ . Therefore, the conditional PDF  $P[Y(t)|X(t)]$  can be presented in the factorized form:

$$P[Y(t)|X(t)] = \prod_{j=-M}^{M-1} P_j[Y_j|X_j], \quad (95)$$

where  $X_j = X(t_j)$ ,  $Y_j = Y(t_j)$ , and  $P_j[Y_j|X_j]$  is per-sample conditional PDF described in the previous Section 3.1.1, where we should replace  $Y \rightarrow Y_j$ ,  $X \rightarrow X_j$ ,  $Q \rightarrow Q/\Delta$ .

Our goal is to find the PDF  $P[\{\tilde{C}\}|\{C\}]$  in the leading order in the parameter  $1/\text{SNR}$ . Instead of calculation of the path integral, we build the PDF which reproduces all possible correlators of  $\tilde{C}_k$ :  $\langle \tilde{C}_{k_1} \rangle$ ,  $\langle \tilde{C}_{k_1} \tilde{C}_{k_2} \rangle$ ,  $\langle \tilde{C}_{k_1} \dots \tilde{C}_{k_n} \rangle$  for the fixed input set  $\{C\}$  in the leading order in the parameter  $Q$ . These correlators read as

$$\langle \tilde{C}_{k_1} \dots \tilde{C}_{k_n} \rangle = \int \prod_{j=-M}^{M-1} d^2 Y_j P[Y(t)|X(t)] \tilde{C}_{k_1} \dots \tilde{C}_{k_n}, \quad (96)$$

where  $d^2 Y_j = d\text{Re}Y_j d\text{Im}Y_j$ , and  $\tilde{C}_k$  is defined in Equation (18). After substitution of Equations (95), (53), and (18) into Equation (96) and performing the integration we obtain in the leading order in the noise parameter  $Q$ :

$$\langle \tilde{C}_k \rangle = C_k - \frac{iC_k Q L^2 \gamma}{\Delta} \left( 1 - \frac{i\gamma L |C_k|^2 n_4}{3} \right), \quad (97)$$

$$\langle (\tilde{C}_m - \langle \tilde{C}_m \rangle) (\tilde{C}_n - \langle \tilde{C}_n \rangle) \rangle = -i\delta_{m,n} \frac{C_m^2 Q L^2 \gamma}{T_0} \left( n_4 - \frac{2in_6}{3} \gamma L |C_m|^2 \right), \quad (98)$$

$$\langle (\tilde{C}_m - \langle \tilde{C}_m \rangle) \overline{(\tilde{C}_n - \langle \tilde{C}_n \rangle)} \rangle = \delta_{m,n} \frac{QL}{T_0} \left( 1 + \frac{2n_6}{3} \gamma^2 L^2 |C_m|^4 \right), \quad (99)$$

where  $\delta_{m,n}$  is the Kronecker symbol and we have introduced the following notation for the integral of the  $s$ -th power of the pulse envelope function  $f(t)$ :

$$n_s = \int_{-T_0/2}^{T_0/2} \frac{dt}{T_0} f^s(t). \quad (100)$$

We remind that  $f(t)$  is assumed to be normalized by the condition  $n_2 = 1$ . Note that the correlator  $\langle \tilde{C}_k - C_k \rangle$  is proportional to  $QL/\Delta = QLW'/(2\pi)$ , i.e., it is proportional to the total noise power in the whole bandwidth  $W'$ . The reason for that is the bandwidth of the receiver coincides with the bandwidth of the noise. The correlators (98) and (99) are proportional to  $QL/T_0$  and do not depend on the discretization parameter  $\Delta$ . It means that these correlators depend only on the bandwidth of the envelope function  $f(t)$ . So we obtain that in the leading order in the parameter  $Q$  the shift of the mean value  $\tilde{C}_k$  due to the signal-noise interaction is proportional to the total noise power in the channel ( $W' = 2\pi/\Delta$ ), whereas the spread around the average value, see (98), (99), is proportional to the noise power containing in the bandwidth  $W$  of the pulse envelope (bandwidth of the pulse envelope coincides with the bandwidth of the signal  $X(t)$ ). Note that the higher-order corrections in parameter  $Q$  to the correlators are more complicated and contain the noise bandwidth, see details in Appendix A of Ref. [21]. The correlators of higher orders in  $\tilde{C}$  can be calculated in the leading order in the parameter  $Q$  using Equations (97)–(99).

To verify the analytical results (97)–(99) the numerical simulations of pulse propagation through a nonlinear nondispersive optical fiber were performed in Ref. [21]. After that the numerical results for correlators (97)–(99) were obtained. To find these correlators the Equation (4) was solved numerically for the fixed input signal  $X(t)$  and for different realizations of the noise  $\eta(z, t)$ . After that the detection procedure described by Equations (17) and (18) was applied. Finally, the averaging procedure over noise



realizations for the coefficients  $\tilde{C}_k$  was performed. Two numerical methods of the solution of Equation (4) were used: the split-step Fourier method and the Runge–Kutta method of the fourth order. It was shown that the numerical results do not depend on the numerical method and these results are in consistent with analytical ones for different realizations of the input pulse envelope  $f(t)$  and the noise bandwidth  $W'$ . Below we present the comparison of the numerical and analytical results obtained in Ref. [21]. The numerical simulation was done for the following channel parameters, see the Table 2.

**Table 2.** Channel parameters for the extended model.

$\gamma$ [ (km×W) <sup>−1</sup> ]	$\mathcal{Q}$ [ W/(km×Hz)]	$L$ [ km]
1.25	$5.94 \times 10^{-21}$	800

We choose the duration of one pulse as  $T_0 = 10^{-10}$  s. Simulations were performed for different  $t$ -meshes, i.e., for different time grid spacing  $\Delta$ , and for different pulse envelopes. Different grid spacings  $\Delta$  correspond to the different noise bandwidths  $W' = 2\pi/\Delta$  for the fixed noise parameter  $\mathcal{Q}$ . The numerical calculations were performed for the different  $\Delta$  presented in the Table 3.

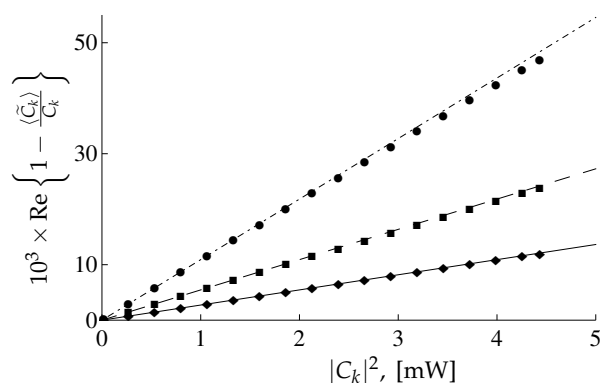
**Table 3.** Time grid spacings.

$\Delta_1$ [s]	$\Delta_2$ [s]	$\Delta_3$ [s]
$9.77 \times 10^{-14}$	$1.95 \times 10^{-13}$	$3.91 \times 10^{-13}$

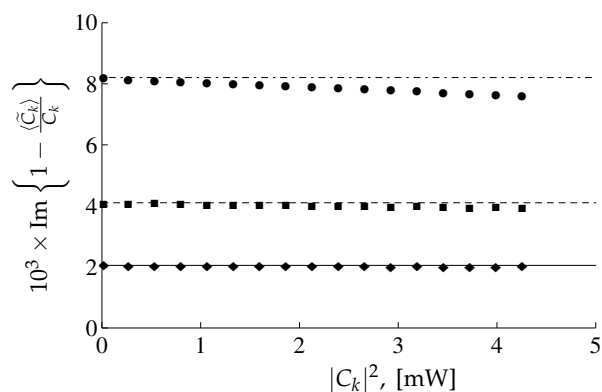
The different grid spacings determine the different widths of the conjugated  $\omega$ -meshes in the frequency domain:  $1/\Delta_1 = 10.26$  THz,  $1/\Delta_2 = 5.12$  THz and  $1/\Delta_3 = 2.56$  THz. In Ref. [21] the different envelopes  $f(t)$  were considered as well. Here we present results only for the Gaussian envelope:

$$f(t) = \sqrt{\frac{T_0}{T_1\sqrt{\pi}}} \exp\left(-\frac{t^2}{2T_1^2}\right), \quad (101)$$

where  $T_1 = T_0/10 = 10^{-11}$  s stands for the characteristic time scale of the function  $f(t)$ . Such relation between  $T_0$  and  $T_1$  means that the overlapping between different pulses is negligible. For pulses with envelope (101) the coefficients  $n_s$  defined in Equation (100) are  $n_4 = \frac{T_0/T_1}{\sqrt{2\pi}} \approx 3.989$ ,  $n_6 = \frac{(T_0/T_1)^2}{\pi\sqrt{3}} \approx 18.38$ ,  $n_8 = \frac{(T_0/T_1)^3}{2\pi\sqrt{\pi}} \approx 89.79$ . In Figures 4 and 5 the real and imaginary parts of the quantity  $(C_k - \langle \tilde{C}_k \rangle) / C_k$  as a function of  $|C_k|^2$  are depicted for different values of grid spacing  $\Delta$ .



**Figure 4.** [This figure is taken from Ref. [21]]. The real part of the relative difference of the coefficient  $C_k$  and the correlator (97) in units  $10^{-3}$  as a function of  $|C_k|^2$ , see [21]. Dashed-dotted, dashed, and solid lines correspond to an analytic representation (97) for time grid spacings  $\Delta_1$ ,  $\Delta_2$ ,  $\Delta_3$ , respectively. Circles, squares, and diamonds correspond to numerical results for time grid spacings  $\Delta_1$ ,  $\Delta_2$ ,  $\Delta_3$ , respectively.



**Figure 5.** [This figure is taken from Ref. [21]]. The imaginary part of the relative difference of the coefficient  $C_k$  and the correlator (97) in units  $10^{-3}$  as a function of  $|C_k|^2$ , see [21]. Dashed dotted, dashed, and solid lines correspond to analytic representation (97) for time grid spacings  $\Delta_1$ ,  $\Delta_2$ ,  $\Delta_3$ , respectively. Circles, squares, and diamonds correspond to numerical results for time grid spacings  $\Delta_1$ ,  $\Delta_2$ ,  $\Delta_3$ , respectively.

One can see the good agreement between analytical and numerical results depicted in Figure 4. There is some difference in the imaginary part of analytical and numerical results corresponding to grid spacing  $\Delta_1$  at large  $|C_k|$ , see Figure 5. The reason is that, for the analytical results corresponding to  $\Delta_1$  it is necessary to take into account the next corrections in the parameter  $Q$  Ref. [21]. The numerical and analytical results for the correlators (98), (99) are also in a good agreement, for details see Ref. [21]. In this paper, it was demonstrated that the relative importance of the next-to-leading order corrections for the correlators (98) and (99) is governed by the dimensionless parameter  $\left(\frac{QL}{\Delta}\gamma L\right)^2 \gamma LP$ , i.e., it increases linearly for increasing power  $P$ .

Now we can proceed to the search for the conditional probability density function. Using the correlators (97)–(99) we build the conditional PDF  $P[\{\tilde{C}\}|\{C\}]$  which reproduces all correlators of the coefficients  $\tilde{C}_m$  in the leading order in parameter  $Q$ . Thus, the conditional PDF has the form [21]:

$$P[\{\tilde{C}\}|\{C\}] = \prod_{m=-N}^N P_m[\tilde{C}_m|C_m], \quad (102)$$

where

$$P_m[\tilde{C}_m|C_m] \approx T_0 \frac{\exp \left[ -T_0 \frac{(1+4n_6\mu_m^2/3)x_m^2 + 2x_my_m\mu_m n_4 + y_m^2}{QL(1+\xi^2\mu_m^2/3)} \right]}{\pi QL \sqrt{1+\xi^2\mu_m^2/3}}, \quad (103)$$

$$x_m = \operatorname{Re} \left[ e^{-i\phi_m} \left\{ \tilde{C}_m - C_m + \frac{iC_m QL^2\gamma}{\Delta} \left( 1 - \frac{i\gamma L|C_m|^2 n_4}{3} \right) \right\} \right], \quad (104)$$

$$y_m = \operatorname{Im} \left[ e^{-i\phi_m} \left\{ \tilde{C}_m - C_m + \frac{iC_m QL^2\gamma}{\Delta} \left( 1 - \frac{i\gamma L|C_m|^2 n_4}{3} \right) \right\} \right], \quad (105)$$

where  $\phi_m = \arg C_m$ ,  $\mu_m = \gamma L|C_m|^2$ , and

$$\xi^2 = (4n_6 - 3n_4^2). \quad (106)$$

The parameter  $\xi^2$  obeys inequality  $\xi^2 \geq n_6 > 0$  due to Cauchy–Schwartz–Buniakowski inequality. For the Gaussian envelope (101) this parameter is

$$\xi^2 = \frac{1}{\pi} \left( \frac{T_0}{T_1} \right)^2 \left( \frac{4}{\sqrt{3}} - \frac{3}{2} \right) \approx 0.258 \left( \frac{T_0}{T_1} \right)^2, \quad (107)$$

and for the chosen parameters  $T_1 = T_0/10 = 10^{-11}$  sec one obtains  $\xi \approx 5.08$ .

Equation (102) means that our channel decomposes to the  $2N + 1$  independent information channels. Therefore, the function  $P_m[\tilde{C}_m|C_m]$  describes the channel corresponding to the  $m$ -th time slot. The function  $P_m[\tilde{C}_m|C_m]$  obeys the normalization condition

$$\int d^2\tilde{C}_m P_m[\tilde{C}_m|C_m] = 1. \quad (108)$$

Since there are  $2N + 1$  independent channels, we can choose the input signal distribution  $P_X[\{C_m\}]$  in the factorized form:

$$P_X[\{C\}] = \prod_{k=-N}^N P_X^{(k)}[C_k], \quad (109)$$

and we can consider only one channel, say  $m$ -th channel. One can see that the presentation of the conditional PDF (103) is close to the presentation (53) for the per-sample PDF. Also the function  $P_m[\tilde{C}_m|C_m]$  changes significantly when the variable  $C_m$  changes on the value of order of  $\sqrt{QL/T_0}$  for fixed value of  $\tilde{C}_m$ . Such behavior coincides with that for the function  $P[Y|X]$  for the per-sample model. Below, we imply that

$$P \gg QL/\Delta \gg QL/T_0, \quad (110)$$

where  $P$  is the mean power of the  $m$ -th pulse, i.e., the signal power is much greater than the noise power in the whole bandwidth  $W' = 2\pi/\Delta$  and in the input signal bandwidth  $W$ . For the extended model we define the signal-to-noise ratio as

$$\text{SNR} = \frac{PT_0}{QL}, \quad (111)$$

on the assumption (110) we have  $\text{SNR} \gg 1$ . We also imply that the PDF of the input signal  $P_X^{(m)}[C_m]$  is a smooth function that changes on a scale  $|C_m| \sim \sqrt{P}$ . Therefore, using a consideration similar to that for the per-sample model we obtain that the PDF of the output signal reads

$$P_{out}^{(m)}[\tilde{C}_m] = \int d^2 C_m P_m[\tilde{C}_m|C_m] P_X^{(m)}[C_m] \quad (112)$$

in the leading order in the parameter  $1/\text{SNR}$  it has the form:

$$P_{out}^{(m)}[\tilde{C}_m] \approx P_X^{(m)}[\tilde{C}_m]. \quad (113)$$

Just as a reminder, to obtain result (113) we perform the integration in Equation (112) using Laplace's method [27], see details in Ref. [19].

Since we know the conditional PDF  $P_m[\tilde{C}_m|C_m]$  and output signal PDF  $P_{out}^{(m)}[\tilde{C}_m]$  the entropies  $H[\tilde{C}_m]$ ,  $H[\tilde{C}_m|C_m]$  and then mutual information  $I_{P_X^{(m)}}$  can be calculated. The calculation is similar to that performed in the Section 3.1.3. After these calculations we solve the variational problem for the mutual information  $I_{P_X^{(m)}}$ , and find the optimal input signal PDF in the form:

$$P_{opt}^{(m)}[C_m] = \mathcal{N}_0 \frac{e^{-\lambda|C_m|^2}}{\sqrt{1 + \xi^2 \gamma^2 L^2 |C_m|^4/3}}. \quad (114)$$

Substituting the result (114) to the mutual information, we obtain the following result:

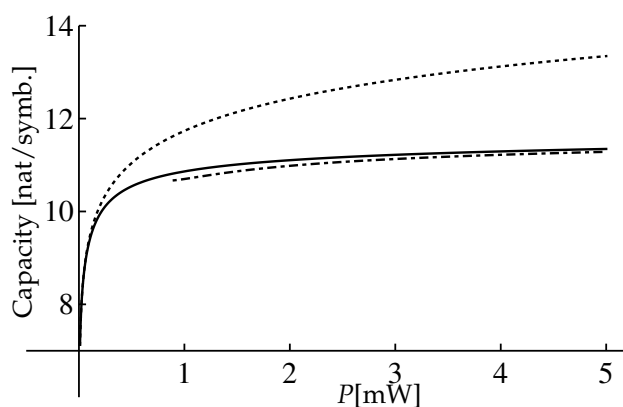
$$I_{P_{opt}^{(m)}} = \log \left( \frac{PT_0}{\pi e QL} \right) + P\lambda - \log [P\mathcal{N}_0], \quad (115)$$

where the parameters  $\mathcal{N}_0$  and  $\lambda$  are the solutions (as functions of power  $P$ ) of the normalization conditions for the function (114):

$$\int_0^\infty d\rho \frac{2\pi\mathcal{N}_0 \rho e^{-\lambda\rho^2}}{\sqrt{1 + \xi^2 \gamma^2 L^2 \rho^4/3}} = 1, \quad \int_0^\infty d\rho \frac{2\pi\mathcal{N}_0 \rho^3 e^{-\lambda\rho^2}}{\sqrt{1 + \xi^2 \gamma^2 L^2 \rho^4/3}} = P, \quad (116)$$

where we have performed the change of variables,  $|C_m|$  to  $\rho$ . Note that the results (115)–(116) are obtained in the leading order in the parameter  $1/\text{SNR}$ . Note that the expression (115) is obtained for one  $m$ -th pulse, therefore, to obtain the mutual information of the channel with the input signal (11) it is necessary to multiply the right-hand-side of Equation (115) by the number of the independent channels, i.e.,  $2N + 1$ .

One can see that relations in the Equation (116) coincide with ones in the Equation (76) after changing the parameter  $\gamma \rightarrow \gamma/\xi$ . Therefore, to obtain the results for the mutual information  $I_{P_{opt}^{(m)}}$  and its asymptotics for the extended model we can replace the parameter  $\gamma$  with  $\xi\gamma$  and parameter  $Q \rightarrow Q/T_0$  in Equations (82)–(84). So, by the rescaling of Figure 2 we obtain the mutual information  $I_{P_{opt}^{(m)}}$  of the extended model, see Figure 6.



**Figure 6.** [This figure is taken from Ref. [21]]. Shannon capacity and the mutual information  $I_{P_{opt}}^{(m)}$  for the parameters which are presented in the Table 2,  $T_0 = 10^{-10}$  sec, and for the Gaussian shape (101) of  $f(t)$ . The black dotted line corresponds to the Shannon limit  $\log\left(\frac{PT_0}{QL}\right)$ , the black solid line corresponds to  $I_{P_{opt}}^{(m)}$ , see Equation (115), the black dashed dotted line corresponds to the asymptotics (118) for large  $\gamma LP$ .

The asymptotics has the form:

$$I_{P_{opt}}^{(m)} \approx \log\left(\frac{PT_0}{QL}\right) - \frac{\xi^2 \gamma^2 L^2 P^2}{3}, \quad (117)$$

for  $\xi \gamma LP \ll 1$ , and for the case when power  $P$  obeys the conditions  $\log \xi \gamma LP \gg 1$ ,  $P \ll T_0 / (QL^3 \xi^2 \gamma^2)$  we have the asymptotics

$$I_{P_{opt}}^{(m)} \approx \log \log \left( B \xi \gamma LP / \sqrt{3} \right) - \log \left( QL^2 \xi \gamma e / (T_0 \sqrt{3}) \right) + \frac{1}{\log \left( B \xi \gamma LP / \sqrt{3} \right)} \left[ \log \log \left( B \xi \gamma LP / \sqrt{3} \right) + 1 - \frac{\log \log \left( B \xi \gamma LP / \sqrt{3} \right)}{\log \left( B \xi \gamma LP / \sqrt{3} \right)} \right]. \quad (118)$$

Note that the asymptotics (118) is obtained with accuracy  $1 / \log^2(\xi \gamma LP)$ .

The found mutual information (115) is calculated for the fixed shape of the pulse envelope  $f(t)$ . It is worth noting that the pulse shape  $f(t)$  is encoded only through one parameter  $\xi$ , see Equation (106). Therefore, strictly speaking, to find the capacity one should find the maximum over all forms of the pulses. On the one hand, we can consider the expression (115) as the estimation of the capacity of the channel whose model implies the specific given shape of the pulse envelope  $f(t)$ . On the other hand, using Equation (118) for fixed power  $P$  one can increase the value of  $I_{P_{opt}}^{(m)}$  raising the parameter  $\xi^2$ , say, decreasing the signal time width  $T_1$  for the Gaussian pulse envelope, see Equation (107). However, the limitation of the intermediate power range  $P \ll T_0 / (QL^3 \xi^2 \gamma^2)$  makes this parameter have an upper limit  $\xi_{\max}^2 \sim [\gamma LP (\gamma L^2 Q / T_0)]^{-1}$ , and the maximum value of the mutual information  $I_{P_{opt}}^{(m)}$  over pulse envelope parameters is attained for the argument  $B \xi_{\max} \gamma LP / \sqrt{3} \sim \sqrt{\frac{PT_0}{QL}} = \sqrt{\text{SNR}}$ .

Note that the mutual information  $I_{P_{opt}}^{(m)}$ , see Equation (115), cannot be considered as the lower bound of the channel capacity since we do not know the sign of the next-to-leading order corrections in the parameter  $Q$ . However, the quantity  $I_{P_{opt}}^{(m)}$  is the estimation of the capacity with the accuracy  $\mathcal{O}(Q)$ . The mutual information  $I_{P_{opt}}^{(m)}$  grows as  $\log \log P$ , see Equation (118), for sufficiently large average power  $P$ :  $(\xi \gamma L)^{-1} \ll P \ll T_0 / (QL^3 \xi^2 \gamma^2)$ . Note that we have obtained similar asymptotics for the per-sample

model. The time dependence of the pulse leads to same asymptotics behavior with modified nonlinearity parameter  $\gamma$  ( $\gamma \rightarrow \xi\gamma$ ).

It worth noting that all our calculations were performed under the assumption of the negligible overlapping of the pulses, see Equation (12). It means we imply that the corrections due to the overlapping effects are at least of the same order as the next-to-leading order corrections in the parameter  $QL$ . We can satisfy this condition by choosing appropriate pulse parameters  $T_0$  and  $T_1$ .

#### 4. Conclusions

In our review, we considered nondispersive nonlinear optical-fiber communication channel with the additive noise. We studied two different models of the channel: the per-sample model and the extended model. For these models, we present results of the calculation for the following theoretical information characteristics such as the output signal entropy, conditional entropy, mutual information in the leading order in the parameter  $1/\text{SNR}$ . To calculate these quantities two methods were developed for the calculation of the conditional probability density function  $P[Y|X]$  [19,21]. The first method which was used for the  $P[Y|X]$  calculation for the per-sample model is based on the path integral approach. In the approach, the path integral (28) was treated using the saddle point method, i.e., the expansion in the parameter  $1/\text{SNR}$ , see Refs. [19,20]. This method was used to obtain the expression for the conditional PDF  $P[Y|X]$  which is convenient for the analytical calculations of the output signal PDF  $P_{out}[Y]$ , entropies, the mutual information, and the optimal input signal distribution in the leading [19], next-to-leading [20] orders in the parameter  $1/\text{SNR}$ . These calculations allow us to find the lower bound of the channel capacity for the per-sample model in the intermediate power range. The second method was applied to the investigation of the extended model, which contains such characteristics as the bandwidths of the input signal, the noise, and the receiver, and also takes into account the projection procedure (18). The second method is based on the calculation of the correlators of the output signal for the fixed input signal in the leading and next-to-leading orders in the noise parameter  $Q$ , see Ref. [21]. Using these correlators the conditional PDF  $P[\{\tilde{C}\}|\{C\}]$  which reproduces all these correlators in the leading order in the parameter  $1/\text{SNR}$  was constructed. The knowledge of the function  $P[\{\tilde{C}\}|\{C\}]$  allowed us to find the informational characteristics of the extended model [21].

We compared the results of our calculations for the per-sample model with limitations on the capacity obtained by other authors [4,17,18,22]. We demonstrated that the conditional PDF  $P[Y|X]$  presented even in the leading order in the parameter  $1/\text{SNR}$  reproduces the exact result (29) with high accuracy, see Figure 1. Using the expression (53) in the intermediate power range, we found the lower bound (82) of the capacity which is consistent with the recent results obtained in Ref. [22].

For the extended model [21] we present results of the calculation of the output signal correlators for the fixed input signal and demonstrate that the difference in the average value for the recovered coefficient  $\tilde{C}_k$  and the input coefficient  $C_k$  is proportional to the noise power containing in the total noise bandwidth, see Equation (97). Whereas the covariances (98), (99) are proportional to the noise power containing in the input signal bandwidth. This behavior of the mean value  $\langle \tilde{C}_k \rangle$  is related to the model of the receiver (i.e., the bandwidth of the receiver coincides with the bandwidth of the noise). The obtained analytical results (97)–(99) were confirmed by the direct numerical calculations, see Figures 4 and 5. Therefore, the constructed conditional PDF  $P[\{\tilde{C}\}|\{C\}]$  contains information about the bandwidths of the signal, noise and receiver [21]. This result is in agreement with assertions made in Ref. [14]. Despite the dependance of the PDF  $P[\{\tilde{C}\}|\{C\}]$  on the noise bandwidth the mutual information calculated in the leading order in  $1/\text{SNR}$  for the extended model depends only on the noise power containing in the input signal bandwidth. Since we have not calculated the corrections in the parameter  $1/\text{SNR}$  to the mutual information (115), we can only to consider this quantity as the capacity estimation rather than the lower bound of the capacity.

The models considered in the present paper are widely-spaced from the modern communication systems where the coefficient of the second dispersion is not zero and the signal detection procedure differs from considered above. The effects related to the non-zero dispersion and properties of the receiver can significantly change the results for the mutual information obtained in our consideration. However, the methods described in the present paper may be useful for the consideration of real communication systems. The calculations performed in Refs. [29–33] is indicative of the possibility to use the presented methods for the capacity investigation of the nonlinear fiber-optical channels with non-zero dispersion.

**Author Contributions:** All authors performed the calculations together. The consideration of the per-sample model was prepared predominantly by I.S.T., whereas the consideration of the extended model was prepared predominantly by A.V.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was partly funded by Russian Science Foundation grant number 17-72-30006, and partly by the Ministry of Education and Science of the Russian Federation.

**Acknowledgments:** The work of I.S.T. Terekhov was supported by the Russian Science Foundation (Grant No. 17-72-30006). The work of A.V.R. Reznichenko was supported by the Ministry of Education and Science of the Russian Federation.

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

## Abbreviations

The following abbreviations are used in this manuscript:

PDF	Probability density function
NLSE	Nonlinear Schrödinger equation
SNR	Signal-to-noise power ratio
MNC	Memoryless nonlinear Schrödinger channel
RPC	Regular perturbative channel
LPC	Logarithmic perturbative channel

## References

- Shannon, C. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423; 27, 623–656. [\[CrossRef\]](#)
- Ilyas, M.; Mouftah, H.T. *The Handbook of Optical Communication Networks*; CRC Press LLC: Boca Raton, FL, USA, 2003.
- Haus, H.A. Quantum noise in a solitonlike repeater system. *J. Opt. Soc. Am. B* **1991**, *8*, 1122–1126. [\[CrossRef\]](#)
- Mecozzi, A. Limits to long-haul coherent transmission set by the Kerr nonlinearity and noise of the in-line amplifiers. *J. Light. Technol.* **1994**, *12*, 1993–2000. [\[CrossRef\]](#)
- Iannoe, E.; Matera, F.; Mecozzi, A.; Settembre, M. *Nonlinear Optical Communication Networks*; John Wiley & Sons: New York, NY, USA, 1998.
- Zakharov, V.E.; Shabat, A.B. Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media. *Soviet J. Exp. Theory Phys.* **1972**, *34*, 62–69.
- Novikov, S.; Manakov, S.V.; Pitaevskii, L.P.; Zakharov, V.E. *Theory of Solitons: The Inverse Scattering Method (Monographs in Contemporary Mathematics)*; Springer: Rumford, ME, USA, 1984.
- Essiambre, R.-J.; Foschini, G.J.; Kramer, G.; Winzer, P.J. Capacity Limits of Information Transport in Fiber-Optic Networks. *Phys. Rev. Lett.* **2008**, *101*, 163901. [\[CrossRef\]](#)
- Essiambre, R.-J.; Kramer, G.; Winzer, P.J.; Foschini, G.J.; Goebel, B. Capacity Limits of Optical Fiber Networks. *J. Light. Technol.* **2010**, *28*, 662–701. [\[CrossRef\]](#)
- Sorokina, M.A.; Turitsyn, S.K. Regeneration limit of classical Shannon capacity. *Nat. Commun.* **2014**, *5*, 3861. [\[CrossRef\]](#)
- Terekhov, I.S.; Vergeles, S.S.; Turitsyn, S.K. Conditional Probability Calculations for the Nonlinear Schrödinger Equation with Additive Noise. *Phys. Rev. Lett.* **2014**, *113*, 230602. [\[CrossRef\]](#)



12. Liga, G.; Xu, T.; Alvarado, A.; Killey, R.I.; Bayvel, P. On the performance of multichannel digital backpropagation in high-capacity long-haul optical transmission. *Opt. Express* **2014**, *22*, 30053–30062. [\[CrossRef\]](#)
13. Semrau, D.; Xu, T.; Shevchenko, N.; Paskov, M.; Alvarado, A.; Killey, R.; Bayvel, P. Achievable information rates estimates in optically amplified transmission systems using nonlinearity compensation and probabilistic shaping. *Opt. Lett.* **2017**, *42*, 121–124. [\[CrossRef\]](#)
14. Kramer, G. Autocorrelation Function for Dispersion-Free Fiber Channels with Distributed Amplification. *IEEE Tran. Inf. Theory* **2018**, *64*, 5131–5155. [\[CrossRef\]](#)
15. Tang, J. The Shannon channel capacity of dispersion-free nonlinear optical fiber transmission. *J. Light. Technol.* **2001**, *19*, 1104–1109. [\[CrossRef\]](#)
16. Tang, J. The multispan effects of Kerr nonlinearity and amplifier noises on Shannon channel capacity of a dispersion-free nonlinear optical fiber. *J. Light. Technol.* **2001**, *19*, 1110–1115. [\[CrossRef\]](#)
17. Turitsyn, K.S.; Derevyanko, S.A.; Yurkevich, I.V.; Turitsyn, S.K. Information Capacity of Optical Fiber Channels with Zero Average Dispersion. *Phys. Rev. Lett.* **2003**, *91*, 203901. [\[CrossRef\]](#)
18. Yousefi, M.I.; Kschischang, F.R. On the Per-Sample Capacity of Nondispersive Optical Fibers. *IEEE Trans. Inf. Theory* **2011**, *57*, 7522–7541. [\[CrossRef\]](#)
19. Terekhov, I.S.; Reznichenko, A.V.; Kharkov, Y.A.; Turitsyn, S.K. The log log growth of channel capacity for nondispersive nonlinear optical fiber channel in intermediate power range. *Phys. Rev. E* **2017**, *95*, 062133. [\[CrossRef\]](#)
20. Panarin, A.A.; Reznichenko, A.V.; Terekhov, I.S. Next-to-leading order corrections to capacity for nondispersive nonlinear optical fiber channel in intermediate power region. *Phys. Rev. E* **2016**, *95*, 012127. [\[CrossRef\]](#)
21. Reznichenko, A.V.; Smirnov, S.V.; Chernykh, A.I.; Terekhov, I.S. The 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. [\[CrossRef\]](#)
22. Keykhosravi, K.; Durisi, G.; Agrell, E. Accuracy Assessment of Nondispersive Optical Perturbative Models Through Capacity Analysis. *Entropy* **2019**, *21*, 760. [\[CrossRef\]](#)
23. Mitra, P.; Stark, J.B. Nonlinear limits to the information capacity of optical fibre communications. *Nature* **2001**, *411*, 1027–1030. [\[CrossRef\]](#)
24. Hochberg, D.; Molina-Paris, C.; Perez-Mercader, J.; Visser, M. Effective action for stochastic partial differential equations. *Phys. Rev. E* **1999**, *60*, 6343–6360. [\[CrossRef\]](#)
25. Gardiner, C.W. *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*; Springer: New York, NY, USA, 1985.
26. Feynman, R.P.; Hibbs, A.R. *Quantum Mechanics and Path Integrals*; McGraw-Hill: New York, NY, USA, 1965.
27. Lavrentiev, M.A.; Shabat, B.V. *Method of Complex Function Theory*; Nauka: Moscow, Russia, 1987.
28. Gradshteyn, I.S.; Ryzik, I.M. *Table of Integrals and Series, and Products*; Academic Press: Orlando, FL, USA, 2014.
29. Terekhov, I.S.; Reznichenko, A.V.; Turitsyn, S.K. Calculation of mutual information for nonlinear communication channel at large SNR. *Phys. Rev. E* **2016**, *94*, 042203. [\[CrossRef\]](#)
30. Reznichenko, A.V.; Terekhov, I.S. Channel Capacity and Simple Correlators for Nonlinear Communication Channel at Large SNR and Small Dispersion. *IEEE Int. Symp. Inf. Theory Proc.* **2018**, *2018*, 186–190.
31. Reznichenko, A.V.; Terekhov, I.S.; Turitsyn, S.K. Calculation of mutual information for nonlinear optical-fiber communication channel at large SNR within path-integral formalism. *J. Phys. Conf. Ser.* **2017**, *826*, 012026. [\[CrossRef\]](#)
32. Reznichenko, A.V.; Terekhov, I.S. Channel Capacity Calculation at Large SNR and Small Dispersion within Path-Integral Approach. *J. Phys. Conf. Ser.* **2018**, *999*, 012016. [\[CrossRef\]](#)
33. Reznichenko, A.V.; Terekhov, I.S. Investigation of Nonlinear Communication Channel with Small Dispersion via Stochastic Correlator Approach. *J. Phys. Conf. Ser.* **2019**, *1206*, 012013. [\[CrossRef\]](#)

