810 Vol. 39, No. 3 / March 2022 / Journal of the Optical Society of America B

Research Article

Journal of the Optical Society of America B

OPTICAL PHYSICS

Optimal input signal distribution for a nonlinear optical fiber channel with small Kerr nonlinearity

A. V. Reznichenko,[1,*] A. I. Chernykh,[2,3] E. V. Sedov,[2,4] and I. S. Terekhov[1]

[1]Budker Institute of Nuclear Physics of Siberian Branch Russian Academy of Sciences, 11 Acad. Lavrentieva Pr., Novosibirsk 630090, Russia
[2]Novosibirsk State University, 2 Pirogova St., Novosibirsk 630090, Russia
[3]Institute of Automation and Electrometry of the Siberian Branch of the Russian Academy of Sciences, 1 Ac. Koptyuga Avenue, Novosibirsk 630090, Russia
[4]Aston Institute of Photonics Technologies, Aston University, Aston Triangle, Birmingham, B4 7ET, UK
*Corresponding author: A.V.Reznichenko@inp.nsk.su

Received 12 October 2021; accepted 23 December 2021; posted 21 January 2022; published 17 February 2022

We consider the information channel described by the Schrödinger equation with additive Gaussian noise. We introduce the model of the input signal and the model of the output signal receiver. For this channel, using perturbation theory for the small nonlinearity parameter, we calculate the first three terms of the expansion of the conditional probability density function in the nonlinearity parameter. At a large signal-to-noise power ratio, we calculate the conditional entropy, output signal entropy, and mutual information in the leading and next-to-leading order in the nonlinearity parameter and in the leading order in the parameter 1/SNR. Using the mutual information, we find the optimal input signal distribution and channel capacity in the leading and next-to-leading order in the nonlinearity parameter. Finally, we present the method of the construction of the input signal with the optimal statistics for the given shape of the signal.

https://doi.org/10.1364/JOSAB.445376

1. INTRODUCTION

Nowadays, fiber-optic communication channels are actively developed. That is why it is important to know their maximum information transmission rate, i.e., channel capacity. For small powers of the outcoming signal, these channels are well described by linear models. For the powers in question, the capacity of the noisy channel was found analytically in Shannon's famous papers [1,2]:

$$C \propto \log(1 + \mathrm{SNR}), \tag{1}$$

where $\mathrm{SNR} = P/N$ is the signal-to-noise power ratio, P is the average input signal power, and N is the noise power. As one can see from this relation, to increase the channel capacity with the noise power being fixed, it is necessary to increase the signal power. However, as the signal power increases, the nonlinear effects become more important. In this case, a simple expression for the capacity of this noisy nonlinear channel is unknown. The reason is that the expression should take into account all details of the mathematical model of the nonlinear channel. This model implies the following components: the input signal model, i.e., the method for encoding incoming information, the physical signal propagation model across the fiber wire, the receiver model (i.e., signal detection features, frequency filtering), and the procedure of signal post-processing. Thus, for different channel models, the expressions for capacity will also be different even for some matching components. It is very difficult to find an explicit expression for capacity even for a specific (and often highly simplified) model of a nonlinear communication channel. A more realistic problem for such channels is to find the upper or lower bounds for the capacity. For example, in the case of commonly used models involving the signal propagation governed by the nonlinear Schrödinger equation (NLSE) with additive white Gaussian noise (see Refs. [3–6] and references therein), an expression for capacity has not yet been obtained. However, in the case when a small parameter is present in the model, often it is possible to invoke the perturbation theory for this parameter if the zero-parameter model turns out to be solvable. For instance, in the case of low noise power in the channel (i.e., for a large SNR parameter), one can develop an analog of the semiclassical approach in quantum mechanics [7]. Further, in the case when the coefficient of the second dispersion in the channel is small, it is also possible to construct the perturbation theory based on this parameter [8–10], since the zero-parameter model is well developed [4,11–18]. Finally, in the case of moderate input power, it is also possible to develop the perturbation theory for the Kerr nonlinearity parameter of the fiber [19]. In different approaches, the capacity for an optical fiber channel with nonzero dispersion and Kerr nonlinearity has been studied both analytically and numerically; see Refs. [20–27] and references therein.

In the present paper, we focus on the study of the nonlinear channel described by NLSE with additive Gaussian noise using

the perturbation theory in the small parameter of Kerr nonlinearity and large SNR. To this end, we consistently build the model of the input signal $X(t)$, we study the impact of the spectral noise width on the output signal $Y(t)$ (i.e., the raw signal in the receiving point), and we investigate the influence of the signal detection procedure in the receiver and the post-processing, i.e., the procedure of input data extraction from the received signal $Y(t)$. We carry out all our calculations in the leading and next-to-leading orders in the Kerr nonlinearity parameter.

To study the mutual information, we use the representation for the conditional probability density function (PDF) $P[Y(t)|X(t)]$, i.e., the probability density to get the output signal $Y(t)$, if the input signal is $X(t)$, through the path-integral [7]. This representation for $P[Y(t)|X(t)]$ is especially convenient to use the perturbation theory. Generally speaking, the function spaces of the input signal $X(t)$ and output signal $Y(t)$ are infinite-dimensional. However, information is transmitted using some finite set pulses of a certain shape, spread in either time or frequency space. For example, the input signal $X(t)$ can be constructed as follows:

$$X(t) = \sum_{k=-M}^{M} C_k s\,(t - k T_0). \tag{2}$$

Here $s\,(t)$ is a fixed envelope function of time, C_k are complex variables that carry information about the input signal, and T_0 is a time interval between two successive pulses. The problem of information transfer is reduced to recover coefficients $\{C_{-M}, \ldots, C_M\}$ from the signal $Y(t)$ received at the output. To find the informational characteristics of the communication channel, we need to reduce density functional $P[Y(t)|X(t)]$ to functional $P[\{\tilde{C}\}|\{C\}]$, i.e., the conditional probability density to get a set of coefficients $\{\tilde{C}_k\}$, if the input signal is encrypted by coefficients $\{C_k\}$. Functional $P[\{\tilde{C}\}|\{C\}]$ depends on both the physical laws of signal propagation along the communication channel and the detection procedure with post-processing of the signal. Generally, functional $P[\{\tilde{C}\}|\{C\}]$ cannot be reduced to a factorized form,

$$P[\{\tilde{C}\}|\{C\}] = \prod_{k=-M}^{M} P^{(k)}[\tilde{C}_k|C_k], \tag{3}$$

due to dispersion effects in the first place. This means that we deal with a communication channel with memory (commonly, an infinite one). Fiber optical channels with memory were previously considered in many papers; see, for example, Ref. [26].

In our work, we calculate density $P[\{\tilde{C}\}|\{C\}]$ for a nonlinear fiber optic communication channel, in which the signal propagation is governed by the NLSE with additive Gaussian noise of finite spectral width. Our model also includes a receiver model and post-processing procedure of the extraction of coefficients $\{\tilde{C}_k\}$ from the detected signal $Y(t)$. Density functional $P[Y(t)|X(t)]$ as well as density $P[\{\tilde{C}\}|\{C\}]$ were found with the use of two different methods. The first method is based on direct calculation of the path-integral representing $P[Y(t)|X(t)]$ via the effective two-dimensional action [7] in the leading and next-to-leading orders in parameter 1/SNR and in the parameter of Kerr nonlinearity, correspondingly. The second method is based on independent calculation of the correlators of the solution of the NLSE with additive Gaussian noise for a fixed input signal $X(t)$.

Using the found density $P[\{\tilde{C}\}|\{C\}]$, we calculated the entropy of the output signal and conditional entropy. It allowed us to find mutual information in leading and next-to-leading orders in parameter 1/SNR and in the parameter of Kerr nonlinearity. Then we found the extremum of the mutual information and calculated the probability density of the input signal $P_{\mathrm{opt}}[\{C\}]$ that delivers this extremum. We demonstrated that in the first non-vanishing order in Kerr nonlinearity, probability density $P_{\mathrm{opt}}[\{C\}]$ is not factorized, i.e., already in the leading order in the nonlinearity parameter, the fiber optic channel is the channel with memory. The optimal distribution $P_{\mathrm{opt}}[\{C\}]$ allowed us to construct conditional probabilities $P_{\mathrm{opt}}[C_k|C_{-M}, \ldots, C_{k-1}, C_{k+1}, \ldots, C_M]$, which, in turn, are needed to construct the input signal with the given statistics $P_{\mathrm{opt}}[\{C\}]$. Using the explicit form of distribution $P_{\mathrm{opt}}[\{C\}]$, we demonstrated that the difference between the mutual information found using the optimal statistics and the mutual information calculated using Gaussian distribution occurs only in the fourth order in the small parameter of Kerr nonlinearity. To demonstrate our analytical results, we performed numerical calculations of mutual information, optimal distribution function, and correlators of the output signal for various parameters of the second dispersion, as well as for pulse sequences of different lengths.

The paper is organized as follows. The next section is dedicated to the channel model description: we describe the structure of the input signal, then introduce the procedures of receiving and post-processing. We introduce the conditional PDF in the case of small Kerr nonlinearity in the third section. In this section, we propose two approaches to the perturbative calculation of the conditional PDF. The details of this calculation are presented in Supplement 1. The fourth section is devoted to the derivation of mutual information. The resulting expression for mutual information uses tensor notations for the coefficients calculated in detail in Supplement 1. These universal coefficients allow us to present the optimal input signal distribution in the fifth section. In this section, we present the theoretical and numerical results. We present the statistical method of the construction of the optimal input signal in the sixth section: we describe the correlations of input symbols resulting in optimal distribution. The conclusion finalizes our consideration of the optimal input signal distribution for the nonlinear channel with small Kerr nonlinearity.

2. MODEL OF THE CHANNEL

Let us start the consideration from the input signal representation.

A. Input Signal Model

In our model the input signal $X(t)$ has the following form:

$$X(t) = \sum_{k=-M}^{M} C_k s\,(t - k T_0). \tag{4}$$

Thus, the signal is the sequence of $2M+1$ pulses of the shape $s(t)$ spaced by time T_0. The complex coefficients C_k carry the transmitted information. We chose the pulse envelope $s(t)$ possessing two properties. The first property is the normalization condition:

$$\int_{-\infty}^{\infty} \frac{dt}{T_0} s^2(t) = 1. \tag{5}$$

The second property is the orthogonality condition,

$$\int_{-\infty}^{\infty} \frac{dt}{T_0} s(t - kT_0) s(t - mT_0) = \delta_{km}, \tag{6}$$

where δ_{km} is the Kronecker δ-symbol. Below, we will consider two different types of function $s(t)$. The first one is the sinc-type function,

$$s(t) = \mathrm{sinc}\,[Wt/2] = 2\frac{\sin(Wt/2)}{Wt}, \tag{7}$$

where $W = 2\pi/T_0$ is the input signal bandwidth. Note that these envelopes are overlapping; however, properties (6) and (7) are fulfilled. We focus our attention in the following calculations primarily on the sinc-type of envelope.

The second type is the Gaussian function,

$$s(t) = \sqrt{\frac{T_0}{\tau\sqrt{\pi}}} \exp\left(-\frac{t^2}{2\tau^2}\right), \tag{8}$$

where τ is characteristic signal duration. Below, we imply that $\tau \ll T_0$. It is parameter τ that determines the frequency bandwidth of the input signal. So the orthogonality condition (6) can be satisfied only approximately with any specified precision by choosing the value of time τ.

Complex coefficients C_k are distributed with PDF $P_X[\{C\}]$. Below, we refer function $P_X[\{C\}]$ as the input signal PDF, where $\{C\} = \{C_{-M}, C_{-M+1}, \ldots, C_M\}$ is the ordered set of coefficients C_k. In our model, we consider the continuous PDF $P_X[\{C\}]$ normalized by the condition

$$\int \left(\prod_{k=-M}^{M} \mathrm{d}^2 C_k\right) P_X[\{C\}] = 1, \tag{9}$$

where $d^2C_k = d\mathrm{Re}C_k d\mathrm{Im}C_k$. We also restrict our consideration by the input signal $X(t)$ with the fixed averaged power P:

$$P = \int \left(\prod_{k=-M}^{M} \mathrm{d}^2 C_k\right) P_X[\{C\}] \frac{1}{2M+1} \int_{-\infty}^{\infty} \frac{dt}{T_0} |X(t)|^2. \tag{10}$$

B. Signal Propagation Model

In our model, the propagation of the signal $\psi(z, t)$ is described by the NLSE with additive Gaussian noise (see Refs. [3–6]):

$$\partial_z \psi(z, t) + i\beta\partial_t^2 \psi(z, t) - i\gamma|\psi(z, t)|^2 \psi(z, t) = \eta(z, t), \tag{11}$$

with the input condition $\psi(0, t) = X(t)$. In Eq. (11), β is the second dispersion coefficient, γ is the Kerr nonlinearity coefficient, and $\eta(z, t)$ is an additive complex noise with zero mean:

$$\langle \eta(z, t) \rangle_\eta = 0. \tag{12}$$

Here $\langle \ldots \rangle_\eta$ is the averaging over the realization of noise $\eta(z, t)$. We also imply that the correlation function $\langle \eta(z, t)\bar{\eta}(z', t') \rangle_\eta$ has the following form:

$$\langle \eta(z, t)\bar{\eta}(z', t') \rangle_\eta = Q\frac{\tilde{W}}{2\pi} \mathrm{sinc}\left(\frac{\tilde{W}(t - t')}{2}\right) \delta(z - z'). \tag{13}$$

Here and below the bar means complex conjugation. Parameter Q in Eq. (13) is a power of noise $\eta(z, t)$ per unit length and per unit frequency. Parameter $\tilde{W}$ is the bandwidth of the noise. Note that $\lim_{\tilde{W}\to\infty} \frac{\tilde{W}}{2\pi} \mathrm{sinc}(\frac{\tilde{W}(t-t')}{2}) = \delta(t - t')$.

Below, we imply that noise bandwidth $\tilde{W}$ is much greater than bandwidth W of input signal $X(t)$ and much greater than bandwidth W' of the solution $\Phi(z = L, t)$ of Eq. (11) with zero noise. Here L is the signal propagation distance. So in our consideration, we set

$$\tilde{W} \gg W' > W. \tag{14}$$

The solution $\Phi(z, t)$ of Eq. (11) with zero noise and with the input boundary condition $\Phi(z = 0, t) = X(t)$ will play an important role in our further consideration. Details of the perturbative calculation of the solution $\Phi(z, t)$ and its properties are presented in Subsection 2 of Supplement 1.

C. Receiver Model and Post-Processing

To recover the transmitted information, we perform the procedure of output signal detection at $z = L$. Our receiver detects noisy signal $\psi(L, t)$ [solution of Eq. (11) with noise], then it filters the detected signal in the frequency domain. After that, we remove the phase incursion $e^{i\beta\omega^2 L}$ related with the second dispersion coefficient and obtain signal $Y_d(t)$. So in the frequency domain, we finally obtain detected signal $Y_d(\omega)$:

$$Y_d(\omega) = e^{-i\beta\omega^2 L}\theta(W_d/2 - |\omega|) \int_{-\infty}^{\infty} \mathrm{d}t e^{i\omega t} \psi(L, t), \tag{15}$$

where W_d is the frequency bandwidth of our receiver. In our model, bandwidth W_d is much less than $\tilde{W}$ as well as in Eq. (14). Also, it is reasonable to consider the receiver with bandwidth $W_d \geq W$, as it is our case.

To obtain the information, we should recover coefficients $\{C\}$ from signal $Y_d(\omega)$. To this aim, we project signal $Y_d(t)$ on shape functions $s(t - kT_0)$:

$$\begin{aligned} \tilde{C}_k &= \frac{1}{T_0} \int_{-\infty}^{\infty} \mathrm{d}t s(t - kT_0) Y_d(t) \\ &= \frac{1}{2\pi T_0} \int_W \mathrm{d}\omega \overline{s^{(k)}}(\omega) Y_d(\omega), \end{aligned} \tag{16}$$

where $s^{(k)}(\omega)$ is the Fourier transform of function $s(t - kT_0)$:

$$s^{(k)}(\omega) = \int_{-\infty}^{\infty} \mathrm{d}t s(t - kT_0) e^{i\omega t} = e^{i\omega k T_0} s(\omega). \tag{17}$$

Due to the noise and nonlinearity of Eq. (11), the recovered coefficient $\tilde{C}_k$ does not coincide with coefficient C_k. However, in the case of zero nonlinearity and zero noise, our detection procedure allows us to recover all coefficients $\{C\}$.

The informational characteristics of the channel are described by the conditional PDF $P[\{\tilde{C}\}|\{C\}]$ to receive the sequence $\{\tilde{C}\}$ for input sequence $\{C\}$. So we have to find the conditional PDF $P[\{\tilde{C}\}|\{C\}]$.

3. CONDITIONAL PDF $P[\{\tilde{C}\}|\{C\}]$

In this section, we find the conditional PDF $P[\{\tilde{C}\}|\{C\}]$ using two approaches. The first one is based on the result of Ref. [7] where the conditional PDF $P[Y(\omega)|X(\omega)]$ to receive the output signal $Y(\omega)$ for the input signal $X(\omega)$ is represented in the form of path-integral. The second approach is based on calculation of the output signal correlators in the leading and next-to-leading orders in parameter Q. Let us briefly discuss the first and second approaches.

The base of the path-integral approach is the representation for the conditional PDF $P[Y(\omega)|X(\omega)]$ in the frequency domain (see Ref. [7]):

$$P[Y(\omega)|X(\omega)]=\int_{\psi(0,\omega)=X(\omega)}^{\psi(L,\omega)=Y(\omega)}\mathcal{D}\psi(z,\omega)\exp\left[-\frac{S[\psi]}{Q}\right], \tag{18}$$

where the effective action $S[\psi]$ reads

$$\begin{aligned}S[\psi]=\int_0^L \mathrm{d}z\int_{\tilde{W}}\frac{d\omega}{2\pi}\Big|&\partial_z\psi(z,\omega)-i\beta\omega^2\psi(z,\omega)\\&-i\gamma\int_{\tilde{W}}\frac{d\omega_1 d\omega_2 d\omega_3}{(2\pi)^2}\delta(\omega_1+\omega_2-\omega_3-\omega)\\&\times\psi(z,\omega_1)\psi(z,\omega_2)\bar{\psi}(z,\omega_3)\Big|^2,\end{aligned} \tag{19}$$

and the integration measure $\mathcal{D}\psi(z,\omega)$ is defined in such a way to obey the normalization condition $\int\mathcal{D}Y(\omega)P[Y(\omega)|X(\omega)]=1$; for details, see Ref. [7]. As mentioned above, function $P[Y(\omega)|X(\omega)]$ contains redundant degrees of freedom, since the receiver does not detect all frequencies of the output signal $Y(\omega)$. That is why we have to introduce the conditional PDF $P_d[Y_d(\omega)|X(\omega)]$, which is the result of the integration of function $P[Y(\omega)|X(\omega)]$ over redundant degrees of freedom $Y(\omega)$, $|\omega|>W_d/2$:

$$P_d[Y_d(\omega)|X(\omega)]=\int[\mathcal{D}Y(\omega)]_{|\omega|>W_d/2}P[Y(\omega)|X(\omega)]. \tag{20}$$

So function $P_d[Y_d(\omega)|X(\omega)]$ contains only detectable degrees of freedom $Y_d(\omega)$, $|\omega|<W_d/2$; see Eq. (15). If one knows function $P_d[Y_d(\omega)|X(\omega)]$, it is easy to calculate an arbitrary correlator $\langle\tilde{C}_{k_1}\dots\bar{\tilde{C}}_{k_N}\rangle$, where

$$\langle\tilde{C}_{k_1}\dots\bar{\tilde{C}}_{k_N}\rangle=\int\mathcal{D}Y_d(\omega)P_d[Y_d(\omega)|X(\omega)]\bar{\tilde{C}}_{k_1}\dots\bar{\tilde{C}}_{k_N}, \tag{21}$$

where $\tilde{C}_k$ is defined in Eq. (16). For our purposes, we should know correlators in the leading order in noise parameter Q, and up to second order in nonlinearity parameter γ. Knowledge of these correlators allows us to construct the conditional PDF $P[\{\tilde{C}\}|\{C\}]$ that reproduces all correlators with necessary accuracy. Details of this calculation are presented in Supplement 1.

Note that the parameter γ has the dimension, however, all terms of the expansion in the parameter γ of any physical quantity have the same dimension. Therefore, when we write "expansion in the parameter γ" we imply asymptotical expansion in the dimensionless parameter γLP, where P is the averaged input signal power.

The second approach allows us to calculate the same correlators (21) by solving Eq. (11) up to second order in parameter γ and up to first order in noise parameter Q (i.e., the second order in function $\eta(z,t)$). We substitute the solution $\psi(L,t)$ of Eq. (11) into Eq. (15); then the result of Eq. (15) is substituted into Eq. (16), and we arrive at the expression for the measured coefficient $\tilde{C}_k$. Note that, since the solution $\psi(L,t)$ depends on the noise, coefficient $\tilde{C}_k$ depends on noise function $\eta(z,t)$ as well. To calculate any correlator $\langle\tilde{C}_{k_1}\dots\bar{\tilde{C}}_{k_N}\rangle$, we should average the product $\tilde{C}_{k_1}\dots\bar{\tilde{C}}_{k_N}$ over the noise realizations using Eqs. (12) and (13). As it should be, the results for correlators $\langle\tilde{C}_{k_1}\dots\bar{\tilde{C}}_{k_N}\rangle$ are the same for both approaches. The results for the correlators are presented in Subsection 4 of Supplement 1.

Using the obtained correlators, we build the conditional PDF $P[\{\tilde{C}\}|\{C\}]$:

$$\begin{aligned}P[\{\tilde{C}\}|\{C\}]=\Lambda_c\exp\Bigg\{-\frac{T_0}{QL}\sum_{k,k'=-M}^{M}\Big[&\delta\tilde{C}_{k'}F^{k',k}\overline{\delta\tilde{C}_k}\\&+\delta\tilde{C}_{k'}G^{k',k}\delta\tilde{C}_k+\overline{\delta\tilde{C}_{k'}}H^{k',k}\overline{\delta\tilde{C}_k}\Big]\Bigg\},\end{aligned} \tag{22}$$

where $F^{k',k}=\bar{F}^{k,k'}=\delta^{k',k}+F_2^{k',k}$, $H^{k',k}=H^{k,k'}=H_1^{k',k}+H_2^{k',k}$, and $G^{k',k}=\bar{H}^{k',k}=G_1^{k',k}+G_2^{k',k}$ are dimensionless coefficients with $k,k'=-M,\dots,M$. Indices 1 and 2 indicate terms proportional to γ and γ^2, respectively. Quantity $\delta\tilde{C}_k$ is defined as follows:

$$\delta\tilde{C}_k=\tilde{C}_k-\langle\tilde{C}_k\rangle. \tag{23}$$

Here the correlator $\langle\tilde{C}_k\rangle$ is a known function of $\{C\}$; see Eq. (S33). Note that quantity $\langle\tilde{C}_k\rangle$ contains the bandwidth of noise $\tilde{W}$.

The dimensionless coefficients can be presented through pair correlators as

$$\begin{aligned}F^{k',k}&=\bar{F}^{k,k'}=\delta_{k',k}+F_2^{k',k},\\F_2^{k',k}&=4\sum_{m=-M}^{M}\bar{H}_1^{k',m}H_1^{m,k}-\gamma^2\frac{T_0}{2QL}\frac{\partial^2}{\partial\gamma^2}\langle\delta\tilde{C}_{k'}\delta\tilde{C}_k\rangle|_{\gamma=0},\end{aligned} \tag{24}$$

where

$$\begin{aligned}H_1^{m,k}&=-\gamma\frac{T_0}{2QL}\frac{\partial}{\partial\gamma}\langle\delta\tilde{C}_k\delta\tilde{C}_m\rangle|_{\gamma=0},\\H_2^{m,k}&=-\gamma^2\frac{T_0}{4QL}\frac{\partial^2}{\partial\gamma^2}\langle\delta\tilde{C}_k\delta\tilde{C}_m\rangle|_{\gamma=0}.\end{aligned} \tag{25}$$

All needed correlators $\langle \tilde{C}_k \rangle$, $\langle \delta\tilde{C}_k \delta\tilde{C}_m \rangle$, $\langle \delta\tilde{C}_{k'} \delta\tilde{C}_k \rangle$ are presented explicitly in Supplement 1, Eqs. (S33), (S34), and (D35). The normalization factor Λ_c reads up to γ^2 order as

$$\Lambda_c = \left(\frac{T_0}{\pi QL}\right)^{2M+1} \left[1 + \left(\sum_{k=-M}^{M} F_2^{k,k} - 2 \sum_{k,k'=-M}^{M} G_1^{k,k'} H_1^{k',k}\right)\right]. \tag{26}$$

At first sight, the found PDF (22) has a Gaussian form, and it might be suggested that we have reduced the channel to the linear one. But it is not the case, since the dimensionless coefficients depend nonlinearly on the input signal coefficients $\{C\}$. The Gaussian structure is the consequence of the consideration of the problem in the leading order in parameter Q.

Now we turn to the consideration of channel entropies $H[\tilde{C}]$ and $H[\{\tilde{C}\}|\{C\}]$, which are necessary for mutual information calculation.

4. MUTUAL INFORMATION

The conditional entropy reads

$$H[\{\tilde{C}\}|\{C\}] = -\int \mathrm{d}C\mathrm{d}\tilde{C}\, P_X[\{C\}] P[\{\tilde{C}\}|\{C\}] \times \log P[\{\tilde{C}\}|\{C\}], \tag{27}$$

where

$$dC = \prod_{k=-M}^{M} d\mathrm{Re}C_k d\mathrm{Im}C_k, \quad d\tilde{C} = \prod_{k=-M}^{M} d\mathrm{Re}\tilde{C}_k d\mathrm{Im}\tilde{C}_k. \tag{28}$$

To calculate the conditional entropy $H[\{\tilde{C}\}|\{C\}]$, we substitute the conditional PDF (22) into expression (27), then perform the integration over $\{\tilde{C}\}$ and obtain

$$H[\{\tilde{C}\}|\{C\}] = -\int \mathrm{d}C\, P_X[\{C\}] \left(\log \Lambda_c - (2M+1)\right). \tag{29}$$

To perform the integration in Eq. (29), we expand $\log \Lambda_c$ up to γ^2 terms then integrate over $\{C\}$ and arrive at

$$H[\{\tilde{C}\}|\{C\}] = (2M+1)\log(\pi e\, QL/T_0) - \gamma^2 L^2 J_\Lambda^{s_1,s_2;s_3,s_4} \int \mathrm{d}C\, P_X[\{C\}] C_{s_1} C_{s_2} \bar{C}_{s_3} \bar{C}_{s_4}. \tag{30}$$

To obtain Eq. (30), we have used the normalization condition

$$\int \mathrm{d}C\, P_X[\{C\}] = 1. \tag{31}$$

In Eq. (30) and below, unless otherwise stated, we imply that there is summation over the repeated indices. The explicit expression for coefficients $J_\Lambda^{s_1,s_2;s_3,s_4}$ is cumbersome, and therefore we present it in Supplement 1, Eq. (S43).

Now we proceed to calculation of the output signal entropy:

$$H[\{\tilde{C}\}] = -\int \mathrm{d}\tilde{C}\, P_{\mathrm{out}}[\{\tilde{C}\}] \log P_{\mathrm{out}}[\{\tilde{C}\}], \tag{32}$$

where the output signal distribution reads

$$P_{\mathrm{out}}[\{\tilde{C}\}] = \int \mathrm{d}C\, P[\{\tilde{C}\}|\{C\}] P_X[\{C\}]. \tag{33}$$

To calculate the PDF of the output signal, we change the integration variables in Eq. (33) from C_k to $\delta\tilde{C}_k = \tilde{C}_k - \langle \tilde{C}_k \rangle$.

Since in our model the average noise power is much less than the average input signal power ($QL/T_0 \ll P$), we calculate the integral (33) using the Laplace method [28,29] and obtain the following result in the leading order in parameter $1/\mathrm{SNR} = QL/(T_0 P)$:

$$P_{\mathrm{out}}[\{\tilde{C}\}] \approx \left|\frac{\partial(C, \bar{C})}{\partial(\tilde{C}^{(0)}, \bar{\tilde{C}}^{(0)})}\right| P_X[\{F\}], \tag{34}$$

where $\tilde{C}_k^{(0)}$ is the known function of C_k [see Eq. (S8) in Supplement 1]:

$$\tilde{C}_k^{(0)}[\{C\}] = C_k + i\gamma L C_{k_1} C_{k_2} \bar{C}_{k_3} a_1^{k_1,k_2;k_3,k} - \gamma^2 L^2 C_{m_1} C_{m_2} C_{m_3} \bar{C}_{m_4} \bar{C}_{m_5} a_2^{m_1,m_2,m_3;m_4,m_5,k}, \tag{35}$$

and $F[\{\tilde{C}\}]$ is the solution of the equation

$$\tilde{C}_k = \tilde{C}_k^{(0)}[\{F\}]. \tag{36}$$

Solution F of this equation can be found using perturbation theory in parameter γ. One can see that the distribution of the output signal coincides with the input signal distribution $P_X[\{F\}]$ up to the Jacobian determinant $|\frac{\partial(C,\bar{C})}{\partial(\tilde{C}^{(0)},\bar{\tilde{C}}^{(0)})}|$. It allows us to calculate the output signal entropy in the leading order in parameter Q, or $1/\mathrm{SNR} = QL/(T_0 P)$ in dimensionless quantities.

Substituting the result (34) into the expression for the output signal entropy (32), we perform the integration over $\tilde{C}$ and arrive at

$$H[\{\tilde{C}\}] = H[\{C\}] + \int dC\, P_X[\{C\}] \log \left|\frac{\partial(\tilde{C}^{(0)}, \bar{\tilde{C}}^{(0)})}{\partial(C, \bar{C})}\right|, \tag{37}$$

where $H[\{C\}]$ is the input signal entropy:

$$H[\{C\}] = -\int \mathrm{d}C\, P_X[\{C\}] \log P_X[\{C\}]. \tag{38}$$

Therefore, to find the output entropy, we should calculate the logarithm of the Jacobian determinant in Eq. (37). The straightforward calculation in the first non-vanishing order in parameter γ leads to the following result:

$$\log \left|\frac{\partial(\tilde{C}^{(0)}, \bar{\tilde{C}}^{(0)})}{\partial(C, \bar{C})}\right| = \gamma^2 L^2 C_{s_1} C_{s_2} \bar{C}_{s_3} \bar{C}_{s_4} J^{s_1,s_2;s_3,s_4}, \tag{39}$$

where dimensionless coefficients $J^{s_1,s_2;s_3,s_4}$ are given by Eq. (S16).

To calculate the mutual information, we subtract the conditional entropy (30) from the output signal entropy (37):

$$I_{P_X} = H[\{\tilde{C}\}] - H[\{\tilde{C}\}|\{C\}]$$
$$= -(2M+1)\log[\pi e QL/T_0] - \int dC\, P_X[\{C\}]\log P_X[\{C\}]$$
$$+ \gamma^2 L^2 J_I^{s_1,s_2;s_3,s_4} \int \mathrm{d}C\, P_X[\{C\}] C_{s_1} C_{s_2} \bar{C}_{s_3} \bar{C}_{s_4}, \tag{40}$$

where coefficients

$$J_I^{s_1,s_2;s_3,s_4} = J^{s_1,s_2;s_3,s_4} + J_\Lambda^{s_1,s_2;s_3,s_4}. \tag{41}$$

These two contributions to coefficients $J_I^{s_1,s_2;s_3,s_4}$ are presented explicitly in Eqs. (S16) and (S42). The method of the numerical calculation of these coefficients is presented in Supplement 1. The first two terms in the mutual information (40) coincide with that for linear channel ($\gamma = 0$). The third term describes the contribution of the Kerr nonlinearity effects. One can see that the first nonlinear correction to the mutual information is of the order of γ^2. Since coefficients $J_I^{s_1,s_2;s_3,s_4}$ depend on the envelope function $s(t)$, the mutual information also depends on this envelope function. It is worth noting that mutual information depends on the bandwidth of the input signal W via coefficients $J_I^{s_1,s_2;s_3,s_4}$ and does not depend on detector bandwidth W_d. The reason for this is that bandwidth W_d of the receiver is greater than or equal to the bandwidth of the input signal W: $W_d \geq W$. Therefore, all integrals over frequency in the interval $[-W_d/2,\ W_d/2]$ with envelope $s^{(k)}(\omega)$ are reduced to integrals over the frequency interval $[-W/2,\ W/2]$ determined by function $s(\omega)$.

5. OPTIMAL INPUT SIGNAL DISTRIBUTION

Now we can calculate the optimal input signal distribution $P_{\mathrm{opt}}[\{C\}]$ that maximizes mutual information (40). The optimal distribution $P_{\mathrm{opt}}[\{C\}]$ obeys the normalization condition,

$$\int \mathrm{d}C\, P_{\mathrm{opt}}[\{C\}] = 1, \tag{42}$$

and the condition of the fixed average power,

$$\int \mathrm{d}C\, P_{\mathrm{opt}}[\{C\}] \frac{1}{2M+1} \sum_{k=-M}^{M} |C_k|^2 = P. \tag{43}$$

To find $P_{\mathrm{opt}}[\{C\}]$, we solve the variational problem $\delta\mathcal{K}[P_X] = 0$ for the functional

$$\mathcal{K}[P_X] = I_{P_X} - \lambda_1 \left(\int \mathcal{D}C\, P_X[\{C\}] - 1 \right)$$
$$- \lambda_2 \left(\int \mathcal{D}C\, P_X[\{C\}] \frac{1}{2M+1} \sum_{k=-M}^{M} |C_k|^2 - P \right), \tag{44}$$

where $\lambda_{1,2}$ are the Lagrange multipliers, related with restrictions (42) and (43). The solution of the variational problem in the first and second orders in parameter γ and in the leading order in parameter 1/SNR has the form

$$P_{\mathrm{opt}}[\{C\}] = P^{(0)}[\{C\}] \left\{ 1 + \gamma^2 L^2 J_I^{s_1,s_2;s_3,s_4} C_{s_1} C_{s_2} \bar{C}_{s_3} \bar{C}_{s_4} \right.$$
$$+ (\gamma L P)^2 (J_I^{r,s;r,s} + J_I^{r,s;s,r})$$
$$\left. \times \left(1 - \frac{2}{P(2M+1)} \sum_{k=-M}^{M} |C_k|^2 \right) \right\}, \tag{45}$$

where $P^{(0)}[\{C\}]$ is the optimal input signal distribution for the channel with zero nonlinearity parameter γ:

$$P^{(0)}[\{C\}] = \left(\frac{1}{\pi P} \right)^{2M+1} \exp\left[-\frac{1}{P} \sum_{k=-M}^{M} |C_k|^2 \right]. \tag{46}$$

One can see that $P^{(0)}[\{C\}]$ is the Gaussian distribution, whereas distribution (45) is not Gaussian due to the nonlinear corrections. Thus, $P_{\mathrm{opt}}[\{C\}]$ leads to nonzero correlations between coefficients C_k with different k.

Note that the found distribution $P_{\mathrm{opt}}[\{C\}]$ is not the exact optimal input signal distribution for the channel described by the NLSE, since $P_{\mathrm{opt}}[\{C\}]$ is calculated for the given envelope $s(t)$ and only up to γ^2 terms. Nevertheless, $P_{\mathrm{opt}}[\{C\}]$ takes into account the first nonzero nonlinear corrections that lead to nontrivial correlations of input coefficients C_k.

To find the maximal value of mutual information (40), we substitute $P_{\mathrm{opt}}[\{C\}]$ (45) into expression (40), perform the integration over C, and obtain

$$I_{P_{\mathrm{opt}}} = (2M+1)\left(\log\left[\frac{PT_0}{QL} \right] + (\gamma L P)^2 J_\Sigma \right), \tag{47}$$

where

$$J_\Sigma = \frac{J_I^{r,s;r,s} + J_I^{r,s;s,r}}{2M+1}. \tag{48}$$

One can see that the mutual information is proportional to the number of the coefficients C_k, i.e., $2M+1$. The first term in the second brackets in Eq. (47) coincides with Shannon's result for the linear channel at large SNR. The second term is the first nonzero nonlinear correction. Below, we will demonstrate numerically that quantity J_Σ depends weakly on $2M+1$.

We emphasize that calculation of mutual information (40) using Gaussian distribution (46) leads to the result $I_{P^{(0)}}$, which coincides with the result (47). It means that in this order in parameter γ, both distributions give the same result for the mutual information. Thus, one might think that it does not matter what distribution, the optimal, see Eq. (45), or the Gaussian, see Eq. (46), is used for the calculation of mutual information. However, the optimal input signal distribution (45) leads to mutual information $I_{P_{\mathrm{opt}}}$, which is greater than $I_{P^{(0)}}$ in the higher orders in nonlinearity parameter γ. To demonstrate that we have calculated the difference between $I_{P_{\mathrm{opt}}}$ and $I_{P^{(0)}}$ in the leading nonzero order in the nonlinearity parameter γ, we obtain

$$I_{P_{\mathrm{opt}}} - I_{P^{(0)}} = \frac{(\gamma L P)^4}{2} (\langle A^2 \rangle_{P^{(0)}} - \langle A \rangle_{P^{(0)}}^2 - \frac{4}{2M+1} \langle A \rangle_{P^{(0)}}^2), \tag{49}$$

where

$$A=\frac{J_I^{s_1,s_2;s_3,s_4}}{P^2}C_{s_1}C_{s_2}\bar{C}_{s_3}\bar{C}_{s_4}, \tag{50}$$

and here we introduce the averaging over the zero-order distribution [see Eq. (46)]:

$$\langle(\ldots)\rangle_{P^{(0)}}=\int \mathrm{d}C P^{(0)}[\{C\}](\ldots). \tag{51}$$

Performing the averaging in Eq. (49), we arrive at the following result:

$$I_{P_{\mathrm{opt}}}-I_{P^{(0)}}=2(\gamma LP)^4\Big(4\tilde{J}_I^{a,b;b,d}\tilde{J}_I^{d,k;a,k}+\tilde{J}_I^{a,b;c,d}\tilde{J}_I^{c,d;a,b}-\frac{4}{2M+1}\tilde{J}_I^{a,b;b,a}\tilde{J}_I^{c,d;d,c}\Big), \tag{52}$$

where

$$\tilde{J}_I^{a,b;c,d}=\left(J_I^{a,b;c,d}+J_I^{b,a;c,d}+J_I^{a,b;d,c}+J_I^{b,a;d,c}\right)/4. \tag{53}$$

We have checked numerically that the right-hand side of Eq. (52) is positive for all values of the parameter β which are considered in the present paper.

Below, we present results for the mutual information for different envelopes $s(t)$ and for different values of dispersion.

A. Zero Dispersion Case

Direct calculation of mutual information (47) in the case of zero dispersion and non-overlapping envelopes $s(t)$, obeying condition (6) [see, for instance, envelope (8)], gives the result

$$\left.\frac{I_{P_{\mathrm{opt}}}}{2M+1}\right|_{\beta=0}=\log\left[\frac{PT_0}{QL}\right]-(\gamma LP)^2\frac{22N_6-21N_4^2}{3}, \tag{54}$$

where N_λ is the integral

$$N_\lambda=\frac{1}{T_0}\int_{-\infty}^{\infty}\mathrm{d}t s^\lambda(t). \tag{55}$$

We note that $22N_6-21N_4^2>0$ due to Cauchy–Schwarz–Bunyakovsky inequality. For the case of a rectangular pulse $s(t)=\theta(T_0/2-|t|)$ (which corresponds to the case of the per-sample model; see [15]), Eq. (54) passes to

$$\left.\frac{I_{P_{\mathrm{opt}}}}{2M+1}\right|_{\beta=0}=\log\left[\frac{PT_0}{QL}\right]-\frac{(\gamma LP)^2}{3}. \tag{56}$$

This result coincides with Eq. (53) in Ref. [15].

The difference (49) for the case of $\beta=0$ and non-overlapping envelopes $s(t)$ [see Eq. (8)] has the form

$$I_{P_{\mathrm{opt}}}-I_{P^{(0)}}\big|_{\beta=0}=(2M+1)(\gamma LP)^4\frac{(22N_6-21N_4^2)^2}{18}. \tag{57}$$

One can see that the difference is positive and in agreement with the general results of Ref. [15], which is valid for arbitrary nonlinearity.

For the case of the envelope of the sinc form,

$$s(t)=\mathrm{sinc}\,(Wt/2)\,, \tag{58}$$

we obtain the following result for the maximum value of mutual information (47) (see details in Supplement 1, Subsection 3):

$$\frac{I_{P_{\mathrm{opt}}}}{(2M+1)}=\log\left[\frac{PT_0}{QL}\right]+\frac{(\gamma LP)^2}{2M+1}\left\{-\frac{22}{3}\int_{-\infty}^{+\infty}\mathrm{d}\tau S^6(\tau,\tau)+\int_{-\infty}^{+\infty}\mathrm{d}\tau_1\int_{-\infty}^{+\infty}d\tau_2(3S^8(\tau_1,\tau_2)+4S^4(\tau_1,\tau_2)S^2(\tau_1,\tau_1)S^2(\tau_2,\tau_2))\right\}, \tag{59}$$

where

$$S^2(\tau_1,\tau_2)=\sum_{r=-M}^{M}\mathrm{sinc}(\pi(\tau_1+r))\mathrm{sinc}(\pi(\tau_2+r)),$$

$$S^2(\tau,\tau)=\sum_{r=-M}^{M}\mathrm{sinc}^2(\pi(\tau+r)). \tag{60}$$

The numerical result for quantity (59) has the form

$$\frac{I_{P_{\mathrm{opt}}}}{(2M+1)}=\log\left[\frac{PT_0}{QL}\right]-1.26(\gamma LP)^2, \tag{61}$$

where the coefficient at nonlinearity factor $(\gamma LP)^2$ weakly depends on parameter M.

B. Nonzero Dispersion Case

Here we present results for the nonzero dispersion parameter. For this consideration, we choose the following parameters of the channel: $\beta=2\times10^{-23}\mathrm{sec}^2/\mathrm{km}$, propagation length is equal to $L=800$ km, with different values of input signal bandwidth W. The dispersion effects can be described by dimensionless parameter $\tilde{\beta}$:

$$\tilde{\beta}=\beta LW^2/2. \tag{62}$$

Below, we present numerical results for the mutual information for various values of $\tilde{\beta}$. Figure 1 presents the dependence of quantity J_Σ [see Eq. (48)] on different values of parameter $\tilde{\beta}$ for $M=5$; see Eq. (47). We checked that quantity J_Σ weakly depends on M for $M>5$. The points in Fig. 1 were obtained by two different numerical approaches see Supplement 1. Both approaches lead to the same results, and it is the guarantee of correctness of the numerical calculations.

One can see that J_Σ has the minimum at $\tilde{\beta}=0$. It means that nonlinear correction to the mutual information [see Eq. (47)] has the maximum absolute value at $\tilde{\beta}=0$. At small $\tilde{\beta}$, quantity J_Σ demonstrates quadratic dependence on the dispersion parameter; see the inset in Fig. 1. At large $\tilde{\beta}$, quantity J_Σ goes to zero. Our estimations result in the dependence $J_\Sigma\sim\frac{\log\tilde{\beta}}{\tilde{\beta}}$ for large $\tilde{\beta}$; see Ref. [30]. Therefore, nonlinear correction decreases

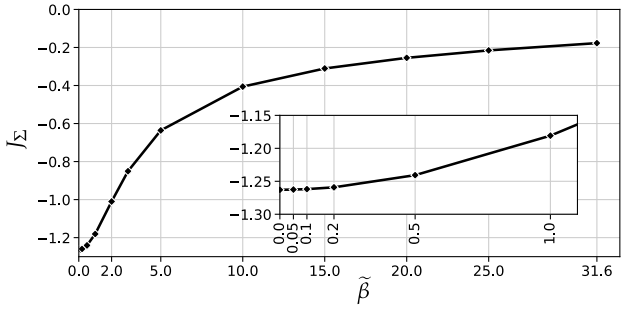

Fig. 1. Dependence of J_Σ on dispersion parameter $\tilde{\beta}$ for $M = 5$. In the inset, we plot J_Σ for the region of small $\tilde{\beta}$.

with increasing $\tilde{\beta}$, as $(\gamma L P)^2/\tilde{\beta}$. It means that the effective nonlinear parameter at large $\tilde{\beta}$ is not $\gamma L P$, but $\gamma L P/\sqrt{\tilde{\beta}}$.

6. CONSTRUCTION OF THE INPUT SIGNALS

To increase the mutual information, we should be able to create signals that obey the optimal input signal distribution (45). To create the input sequence that has statistics determined by the PDF $P_{\rm opt}[\{C\}]$, we represent this PDF in the form [31]

$$P_{\rm opt}[\{C\}] = P_{\rm opt}[C_{i_1}]P_{\rm opt}[C_{i_2}|C_{i_1}]P_{\rm opt}[C_{i_3}|C_{i_2}, C_{i_1}] \times \ldots \times P_{\rm opt}[C_{i_{2M+1}}|C_{i_{2M}}, \ldots C_{i_2}, C_{i_1}], \quad (63)$$

where

$$P_{\rm opt}[C_{i_1}] = \int dC_{i_2} \ldots dC_{i_{2M+1}} P_{\rm opt}[\{C\}], \quad (64)$$

$$P_{\rm opt}[C_{i_2}|C_{i_1}] = \frac{\int dC_{i_3} \ldots dC_{i_{2M+1}} P_{\rm opt}[\{C\}]}{P_{\rm opt}[C_{i_1}]}, \quad (65)$$

$$P_{\rm opt}[C_{i_3}|C_{i_2}, C_{i_1}] = \frac{\int dC_{i_4} \ldots dC_{i_{2M+1}} P_{\rm opt}[\{C\}]}{P_{\rm opt}[C_{i_1}] \times P_{\rm opt}[C_{i_2}|C_{i_1}]}, \quad (66)$$

$$P_{\rm opt}[C_{i_{2M}}|C_{i_{2M-1}}, \ldots C_{i_2}, C_{i_1}] = \frac{\int dC_{i_{2M+1}} P_{\rm opt}[\{C\}]}{P_{\rm opt}[C_{i_1}] \times \ldots \times P_{\rm opt}[C_{i_{2M-1}}|C_{i_{2M-2}}, \ldots, C_{i_1}]}, \quad (67)$$

$$P_{\rm opt}[C_{i_{2M+1}}|C_{i_{2M}}, \ldots C_{i_2}, C_{i_1}] = \frac{P_{\rm opt}[\{C\}]}{P_{\rm opt}[C_{i_1}] \times \ldots \times P_{\rm opt}[C_{i_{2M}}|C_{i_{2M-1}}, \ldots, C_{i_1}]}. \quad (68)$$

Using Eqs. (64)–(68), we can build sequences that have the necessary statistics in the following way. First, we choose the first element C_1 of the sequence distributed with PDF (64). The statistics of the second element C_2 depends on the value of C_1, and should be distributed with PDF (65), etc. In our approximation (γ^2 order of the calculation), the optimal PDF $P_{\rm opt}[\{C\}]$ contains the fourth order polynomial in coefficients C_k. So we have two nontrivial correlators, $\langle C_k \bar{C}_m \rangle_{P_{\rm opt}}$ and $\langle C_k C_q \bar{C}_m \bar{C}_p \rangle_{P_{\rm opt}}$, which determine all higher order correlators.

For a very long sequence, the correlation between the first and last coefficients should be neglectable. To find the characteristic length $|k-m|$ of the correlation between elements C_k and C_m of the input sequence, we calculate correlator $\langle C_k \bar{C}_m \rangle_{P_{\rm opt}}$. After straightforward calculation, we obtain

$$\langle C_k \bar{C}_m \rangle_{P_{\rm opt}} = P\delta_{km}(1-(\gamma L P)^2 \frac{2}{2M+1} \times \sum_{r,s=-M}^{M} (J_I^{r,s;r,s} + J_I^{r,s;s,r})) + P(\gamma L P)^2 \times \sum_{r=-M}^{M} \left[J_I^{r,m;r,k} + J_I^{r,m;k,r} + J_I^{m,r;r,k} + J_I^{m,r;k,r}\right]. \quad (69)$$

The first term on the right-hand side of this equation contains the Kronecker delta symbol, i.e., it is zero for $k \neq m$. The second term describes the correlation between different elements of the input sequence. To find the correlation length, we should investigate the dependence of this term on parameter $m-k$. The correlation length depends on parameter $\tilde{\beta}$. For a small $\tilde{\beta}$, only the nearest neighbors are correlated since the spreading of the input signal due to dispersion is small. When increasing dispersion parameter $\tilde{\beta}$, the correlation length is increasing. The numerical values of coefficients $J_I^{i,j;k,l}$ are presented in Ref. [32]. So one can calculate any necessary correlators.

As an example, we consider the sequence where only the nearest elements are correlated. To build the sequence, we should know only two distributions: $P_{\rm opt}[C_i]$ and $P_{\rm opt}[C_i|C_j] = P_{\rm opt}[C_i, C_j]/P_{\rm opt}[C_j]$. We have performed the calculation of these distributions and obtain

$$P_{\rm opt}[C_q] = P^{(0)}[C_q]\left(1+(\gamma L P)^2 D_1^{(q)}\left(|C_q|/\sqrt{P}\right)\right), \quad (70)$$

where $P^{(0)}[C_q] = \frac{1}{\pi P}\exp\{-\frac{|C_q|^2}{P}\}$ is the Gaussian distribution, and $D_1^{(q)}(x)$ is the following polynomial function:

$$D_1^{(q)}(x) = \left[(1-x^2)\left(2\sum_{r,s=-M}^{M}\frac{J_I^{r,s;r,s}+J_I^{r,s;s,r}}{2M+1} - \sum_{r=-M}^{M}\left[J_I^{r,q;r,q} + J_I^{r,q;q,r} + J_I^{q,r;r,q} + J_I^{q,r;q,r}\right]\right) + J_I^{q,q;q,q}(x^4-4x^2+2)\right], \quad (71)$$

$$P_{\rm opt}[C_i, C_j] = P_{\rm opt}[C_j, C_i] = \int dC_{i_3} \ldots dC_{i_{2M+1}} P_{\rm opt}[\{C\}] = P_{\rm opt}[C_i]P_{\rm opt}[C_j] \times \left\{1+(\gamma P L)^2 D^{i,j}\left(\frac{C_i}{\sqrt{P}}, \frac{C_j}{\sqrt{P}}\right)\right\}, \quad (72)$$

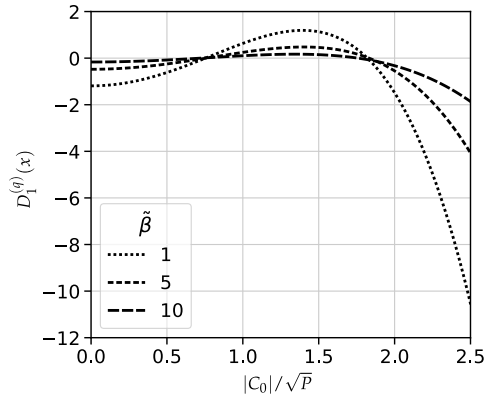

Fig. 2. Dependence of function $D_1^{(q)}(x)$ on $x = |C_0|/\sqrt{P}$ for $q = 0$ and different values of dispersion parameter $\tilde{\beta}$. The dotted, dashed, and long-dashed lines are plotted for dispersion parameter $\tilde{\beta}$ equal to $\tilde{\beta} = 1$, $\tilde{\beta} = 5$, and $\tilde{\beta} = 10$, correspondingly.

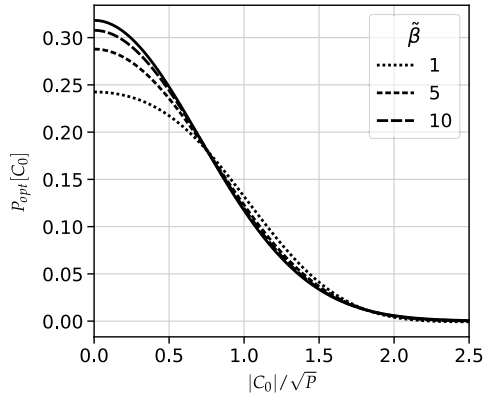

Fig. 3. Function $P_{\text{opt}}[C_0]$ for different values of dispersion parameter $\tilde{\beta}$. The solid line corresponds to the Gaussian distribution $P^{(0)}[C_0]$ for the power parameter $P = 1$ in conventional units. The dotted, dashed, and long-dashed lines are plotted for dispersion parameter $\tilde{\beta}$ equal to $\tilde{\beta} = 1$, $\tilde{\beta} = 5$, and $\tilde{\beta} = 10$, correspondingly.

$$D^{i,j}(x,y) = J_I^{i,i;j,j}x^2\bar{y}^2 + J_I^{j,j;i,i}\bar{x}^2y^2 + (|x|^2-1)(|y|^2-1)\left(J_I^{i,j;i,j} + J_I^{i,j;j,i} + J_I^{j,i;i,j} + J_I^{j,i;j,i}\right) + x\bar{y}\{(J_I^{i,i;i,j} + J_I^{i,i;j,i})$$
$$\times(|x|^2-2) + (J_I^{i,j;j,j} + J_I^{j,i;j,j})(|y|^2-2)\} + y\bar{x}\{(J_I^{i,j;i,i} + J_I^{j,i;i,i})(|x|^2-2) + (J_I^{j,j;i,j} + J_I^{j,j;j,i})(|y|^2-2)\}$$
$$+ \sum_{m=-M}^{M}\left(x\bar{y}\left[J_I^{i,m;j,m} + J_I^{i,m;m,j} + J_I^{m,i;j,m} + J_I^{m,i;m,j}\right] + \bar{x}y\left[J_I^{j,m;i,m} + J_I^{j,m;m,i} + J_I^{m,j;i,m} + J_I^{m,j;m,i}\right]\right). \quad (73)$$

One can see that the corrections to these PDFs are the fourth order polynomials in parameter $C_q/\sqrt{P}$. Let us consider these polynomials. In Fig. 2, we plot function $D_1^{(q)}(x)$ [see Eq. (71)] for different values of $\tilde{\beta}$. Function $D_1^{(q)}(x)$ for different $\tilde{\beta}$ has the maximum in the vicinity of value $x \approx 1.5$. For $x > 1.5$, this function decreases for all values of $\tilde{\beta}$. For smaller $\tilde{\beta}$, the absolute value of function $D_1^{(q)}(x)$ is larger for $x > 2$. It means the applicability region determined by the relation $(\gamma L P)^2 D_1^{(q)}(|C_q|/P) \ll 1$ is wider for larger $\tilde{\beta}$. The reason is the decreasing character of coefficients $J_I^{i,j;k,l}$ for increasing $\tilde{\beta}$; see, e.g., Fig. 2.

To demonstrate the behavior of function $P_{\text{opt}}[C_q]$, we plot it for the different dispersion parameters $\tilde{\beta}$; see Fig. 3. In Fig. 3, we choose nonlinear parameter $(\gamma L P)^2 = 0.2$, which corresponds to the averaged signal power $P \sim 4.5 \times 10^{-4}$W for the realistic channel parameters $\gamma = 1.25\ \text{W}^{-1}\text{km}^{-1}$, $L = 800$ km. One can see that the function $P_{\text{opt}}[C_0]$ decreases slowly for smaller $\tilde{\beta}$. It means that nonlinear correction decreases with increasing $\tilde{\beta}$. The difference $|P^{(0)}[C_0] - P_{\text{opt}}[C_0]|$ gets smaller with increasing $\tilde{\beta}$. The reason is that for the larger dispersion parameter, the signal spreading is larger. It results in the decreasing of the effective nonlinearity parameter, i.e., decreasing coefficients $J_I^{i,j;k,l}$.

The expression in the big curly brackets in Eq. (72) is symmetric in coefficients C_i and C_j. Since we know function $P_{\text{opt}}[C_j, C_i]$, probability $P_{\text{opt}}[C_j|C_i]$ can be easily obtained using Eq. (65):

$$P_{\text{opt}}[C_i|C_j] = P_{\text{opt}}[C_i]\left\{1 + (\gamma P L)^2 D^{i,j}\left(\frac{C_i}{\sqrt{P}}, \frac{C_j}{\sqrt{P}}\right)\right\}. \quad (74)$$

In Fig. 4, we plot function $P_{\text{opt}}[C_0|C_{-1}]$ for different values of coefficient $C_{-1}/\sqrt{P}$ (real and imaginary), nonlinearity parameter $(\gamma L P)^2 = 0.2$, and dispersion parameters $\tilde{\beta}$ equal to one and five. We plot the dependence of $P_{\text{opt}}[C_0|C_{-1}]$ on dimensionless variable $C_0/\sqrt{P}$, where P is chosen to be equal to unity. One can see that function $P_{\text{opt}}[C_0|C_{-1}]$ differs from $P_{\text{opt}}[C_0]$ and depends on value C_{-1} essentially. Also, the PDFs deviate more from Gaussian distribution for larger absolute values of C_{-1}. One can see in Figs. 4(a) and 4(c) that the PDF reaches the negative value at the vicinity of $|C_0| \sim 2$. The reason is the large chosen nonlinearity parameter $(\gamma L P)^2 = 0.2$. The negative value of the PDF makes no sense, but it demonstrates the region of applicability of our approximation. For smaller parameters $\gamma L P$ or for larger parameter $\tilde{\beta}$, our perturbative result (72) is valid wherever function $P_{\text{opt}}[C_j, C_i]$ is not small. For large values of $\tilde{\beta}$, it is necessary to consider not only PDF $P_{\text{opt}}[C_j, C_i]$, but other PDFs presented in Eqs. (64)–(68), since the spreading effects become significant. If necessary, PDFs $P_{\text{opt}}[C_i|C_{j_1}, C_{j_2}], \ldots, P_{\text{opt}}[C_i|C_{j_1}, C_{j_2}, \ldots, C_{j_{2M}}]$ can be derived from Eq. (45) analogously to Eq. (74). We do not present these PDFs here because of their cumbersomeness. Files with coefficients $J_I^{i,j;k,l}$ for different M and $\tilde{\beta}$ that are needed to calculate these PDFs can be found in Ref. [32].

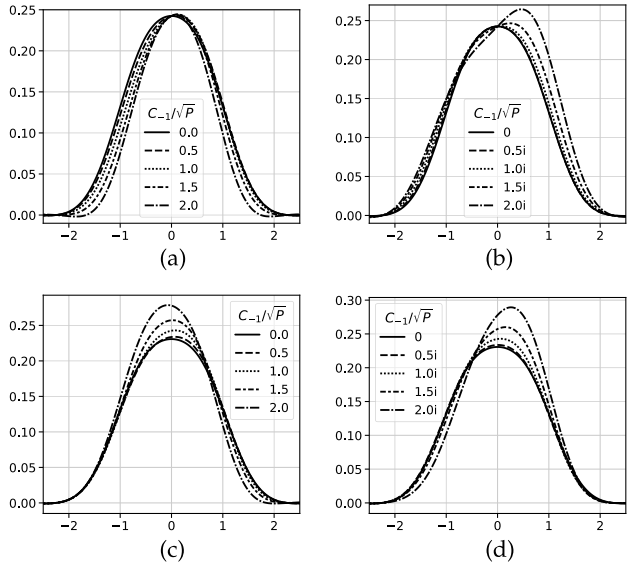

Fig. 4. PDF $P_{\mathrm{opt}}[C_0|C_{-1}]$ as a function of real C_0 for different values of $C_{-1}/\sqrt{P}$ and $\tilde{\beta}$. (a), (b) Cases of $\tilde{\beta}=1$ for real and imaginary values of $C_{-1}/\sqrt{P}$, respectively. (c), (d) Cases of $\tilde{\beta}=5$ for real and imaginary values of $C_{-1}/\sqrt{P}$, respectively.

7. CONCLUSION

In the present paper, we develop a method of calculation of the conditional PDF $P[\{\tilde{C}\}|\{C\}]$ for the channel described by the NLSE with additive noise and nonzero second dispersion coefficient β. To illustrate our method, we calculate PDF $P[\{\tilde{C}\}|\{C\}]$ in the leading and next-to-leading order in the Kerr nonlinearity parameter γ and in the leading order in parameter $1/\mathrm{SNR}$. To obtain $P[\{\tilde{C}\}|\{C\}]$, we calculated $P[Y(\omega)|X(\omega)]$ using two different approaches. The first approach is based on direct calculation of the path-integral; see Eq. (18). In the second approach, we calculate the output signal correlators for the fixed input signal $X(t)$ and then construct the conditional PDF. Both approaches give the same result; see Eq. (S22). To take into account the envelope of the input signal and the detection procedure of the receiver, we integrate PDF $P[Y(\omega)|X(\omega)]$ over the redundant degrees of freedom and obtain the conditional PDF $P[\{\tilde{C}\}|\{C\}]$. Using PDF $P[\{\tilde{C}\}|\{C\}]$ we calculate the mutual information, solve the variational problem, and find the optimal input signal distribution $P_{\mathrm{opt}}[\{C\}]$ in the leading order in parameter $1/\mathrm{SNR}$ and in the second order in the parameter of Kerr nonlinearity γ. We demonstrate that $P_{\mathrm{opt}}[\{C\}]$ differs from Gaussian distribution. Using the distribution $P_{\mathrm{opt}}[\{C\}]$, we calculated the maximal value of mutual information in the leading order in parameter $1/\mathrm{SNR}$ and in the second order in parameter γ for the given pulse envelope, average power, and detection procedure. We demonstrate that the γ^2-correction to the mutual information is negative. Its absolute value is maximal for zero dispersion, and it decreases for increasing dispersion parameter $\tilde{\beta}$; see Fig. 1. We also prove that the mutual information calculated using Gaussian distribution and that calculated with the optimal one coincide in the γ^2 order. The difference appears only in the γ^4 order. It means that the Gaussian distribution of the input signal is a good approximation of the optimal distribution for a small nonlinearity parameter. However, for a not extremely small nonlinearity parameter, it is necessary to take into account the exact PDF $P_{\mathrm{opt}}[\{C\}]$. So, we are able to construct the sequences $\{C\}$ obeying the statistics $P_{\mathrm{opt}}[\{C\}]$. In Section 6, we propose the method of this construction using conditional PDFs (64)–(68). For channels with the small correlation lengths, we calculated explicitly $P_{\mathrm{opt}}[C_i]$ and $P_{\mathrm{opt}}[C_i|C_j]$, and demonstrate the dependence of the probability of the subsequent coefficient C_i on the previous one C_j; see Fig. 4.

Funding. Russian Science Foundation (17-72-30006).

Acknowledgment. The work of A.V. Reznichenko was supported by the Ministry of Education and Science of the Russian Federation. The work of I.S. Terekhov and E.V. Sedov was supported by the RSF.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are available in Ref. [32].

Supplemental document. See Supplement 1 for supporting content.

REFERENCES

1. C. Shannon, "A mathematical theory of communication," Bell Syst. Tech. J. **27**, 379–423 (1948).
2. C. E. Shannon, "Communication in the presence of noise," Proc. Inst. Radio Eng. **37**, 10–21 (1949).
3. H. A. Haus, "Quantum noise in a solitonlike repeater system," J. Opt. Soc. Am. B **8**, 1122–1126 (1991).
4. A. Mecozzi, "Limits to long-haul coherent transmission set by the Kerr nonlinearity and noise of the in-line amplifiers," J. Lightwave Technol. **12**, 1993–2000 (1994).
5. E. Iannoe, F. Matera, A. Mecozzi, and M. Settembre, *Nonlinear Optical Communication Networks* (Wiley, 1998).
6. S. K. Turitsyn, E. G. Turitsyna, S. B. Medvedev, and M. P. Fedoruk, "Averaged model and integrable limits in nonlinear double-periodic Hamiltonian systems," Phys. Rev. E **61**, 3127–3132 (2000).
7. I. S. Terekhov, S. S. Vergeles, and S. K. Turitsyn, "Conditional probability calculations for the nonlinear Schrödinger equation with additive noise," Phys. Rev. Lett. **113**, 230602 (2014).
8. A. V. Reznichenko and I. S. Terekhov, "Channel capacity and simple correlators for nonlinear communication channel at large SNR and small dispersion," in *IEEE Xplore, IEEE Information Theory Workshop* (IEEE, 2017), pp. 186–190.
9. A. V. Reznichenko and I. S. Terekhov, "Calculation at large SNR and small dispersion within path-integral approach," J. Phys. Conf. Series **999**, 012016 (2018).
10. A. V. Reznichenko and I. S. Terekhov, "Investigation of nonlinear communication channel with small dispersion via stochastic correlator approach," J. Phys. Conf. Series **1206**, 012013 (2019).
11. A. Mecozzi and M. Shtaif, "On the capacity of intensity modulated systems using optical amplifiers," IEEE Photon. Technol. Lett. **13**, 1029–1031 (2001).
12. J. Tang, "The Shannon channel capacity of dispersion-free nonlinear optical fiber transmission," J. Lightwave Technol. **19**, 1104–1109 (2001).
13. K. S. Turitsyn, S. A. Derevyanko, I. V. Yurkevich, and S. K. Turitsyn, "Information capacity of optical fiber channels with zero average dispersion," Phys. Rev. Lett. **91**, 203901 (2003).
14. M. I. Yousefi and F. R. Kschischang, "On the per-sample capacity of nondispersive optical fibers," IEEE Trans. Inf. Theory **57**, 7522–7541 (2011).
15. I. S. Terekhov, A. V. Reznichenko, Ya. A. Kharkov, and S. K. Turitsyn, "The log log growth of channel capacity for nondispersive nonlinear optical fiber channel in intermediate power range," Phys. Rev. E **95**, 062133 (2017).

16. A. A. Panarin, A. V. Reznichenko, and I. S. Terekhov, "Next-to-leading order corrections to capacity for nondispersive nonlinear optical fiber channel in intermediate power region," Phys. Rev. E **95**, 012127 (2017).
17. G. Kramer, "Autocorrelation function for dispersion-free fiber channels with distributed amplification," IEEE Trans. Inf. Theory **64**, 5131–5155 (2018).
18. A. V. Reznichenko, A. I. Chernykh, S. V. Smirnov, and I. S. Terekhov, "Log-log growth of channel capacity for nondispersive nonlinear optical fiber channel in intermediate power range: extension of the model," Phys. Rev. E **99**, 012133 (2019).
19. F. J. Garsía-Gómez and G. Kramer, "Mismatched models to lower bound the capacity of optical fiber channels," J. Lightwave Technol. **38**, 6779–6787 (2020).
20. P. P. Mitra and J. B. Stark, "Nonlinear limits to the information capacity of optical fibre communications," Nature **411**, 1027–1030 (2001).
21. E. E. Narimanov and P. Mitra, "The channel capacity of a fiber optics communication system: perturbation theory," J. Lightwave Technol. **20**, 530–537 (2002).
22. J. M. Kahn and K.-P. Ho, "Spectral efficiency limits and modulation detection techniques for DWDM systems," IEEE. J. Sel. Top. Quantum Electron. **10**, 259–272 (2004).
23. R.-J. Essiambre, G. J. Foschini, G. Kramer, and P. J. Winzer, "Capacity limits of information transport in fiber-optic networks," Phys. Rev. Lett. **101**, 163901 (2008).
24. R.-J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity limits of optical fiber networks," J. Lightwave Technol. **28**, 662–701 (2010).
25. R. Killey and C. Behrens, "Shannon's theory in nonlinear systems," J. Mod. Opt. **58**, 1–10 (2011).
26. E. Agrell, A. Alvarado, G. Durisi, and M. Karlsson, "Capacity of a nonlinear optical channel with finite memory," J. Lightwave Technol. **32**, 2862–2876 (2014).
27. M. A. Sorokina and S. K. Turitsyn, "Regeneration limit of classical Shannon capacity," Nat. Commun. **5**, 3861–3866 (2014).
28. M. A. Lavrentiev and B. V. Shabat, *Method of Complex Function Theory* (1987).
29. M. Lavrentiev and B. Chabot, *Methodes de la Theorie des Fonctions d'une Variable Complexe* (1977).
30. I. S. Terekhov, A. V. Reznichenko, and S. K. Turitsyn, "Calculation of mutual information for nonlinear communication channel at large signal-to-noise ratio," Phys. Rev. E **94**, 042203 (2016).
31. A. V. Voytishek, *Foundations of the Monte Carlo Methods in Algorithms and Problems. Parts 1-5. (in Russian)* (Novosibirsk State University, 1997–1999).
32. E. Sedov, "Optimal signal data," GitHub, 2021, https://github.com/esf0/optimal_signal_data.