

# Calculation of mutual information for nonlinear optical fiber communication channel at large SNR within path-integral formalism

A V Reznichenko<sup>1,2</sup>, I S Terekhov<sup>1,2</sup>, S K Turitsyn<sup>2,3</sup>

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

<sup>2</sup> Novosibirsk State University, Novosibirsk 630090, Russia

<sup>3</sup> Aston Institute of Photonics Technologies, Aston University, Aston Triangle, Birmingham B4 7ET, UK

E-mail: a.v.reznichenko@inp.nsk.su

**Abstract.** Using the path-integral technique we calculate the mutual information for the fiber optical channel modelled by the nonlinear Schrödinger equation with additive Gaussian noise. At large signal-to-noise ratio (SNR) we present the mutual information through the path-integral which is convenient for the perturbative expansion both in nonlinearity and dispersion. In the leading order in  $1/\text{SNR}$  we demonstrate that the mutual information is determined through the averaged logarithm of the normalization factor  $\Lambda$  of the conditional probability density function  $P[Y|X]$ . In the limit of small noise and small nonlinearity we derive analytically the first nonzero nonlinear correction to the mutual information for the channel. For the arbitrary nonlinearity we restrict the mutual information by the low bound obtained from the Jensen's inequality and analyze the bound for the case of large dispersion.

## 1. Introduction

We consider the fiber optical channel described by the nonlinear Schrödinger equation (NLSE) with additive white Gaussian noise when the signal-to-noise ratio (SNR) is large:

$$\partial_z \psi + i\beta \partial_t^2 \psi - i\gamma |\psi|^2 \psi = \eta(z, t), \quad (1)$$

where  $\psi(z, t)$  is the outgoing signal,  $\gamma$  is the Kerr nonlinearity,  $\beta = \beta_2/2$  is the dispersion parameter,  $\eta(z, t)$  is the white Gaussian noise:  $\langle \eta(z, t) \bar{\eta}(z', t') \rangle_\eta = Q \delta(z - z') \delta(t - t')$ , where  $Q$  is a power of the white Gaussian noise (per unit length and frequency). The bar here and hereafter means complex conjugation.

Our goal is to calculate the mutual information [1] of the channel for given  $P[X]$  — probability density function (PDF) of the input signal  $X(t)$ :  $\psi(z = 0, t) = X(t)$ . In our approach we use the representation of the mutual information in the form of path-integral

$$I_{P[X]} = \int \mathcal{D}X \mathcal{D}Y P[X] P[Y|X] \log \left[ \frac{P[Y|X]}{P_{out}[Y]} \right] = H[Y] - H[Y|X], \quad (2)$$

where  $P[X]$  is the PDF of the input signal  $X$  with the fixed finite average power  $P_{ave}$ . The function  $P[Y|X]$  here is the conditional probability density function, that is the



probability density of receiving output signal  $\psi(z = L, t) = Y(t)$  when the input signal is  $\psi(z = 0, t) = X(t)$ . The output signal PDF reads  $P_{out}[Y] = \int \mathcal{D}X P[X] P[Y|X]$ . The entropy  $H[Y] = - \int \mathcal{D}Y P_{out}[Y] \log [P_{out}[Y]]$  is the output signal entropy (responsible for the signal transmission), and the entropy  $H[Y|X] = - \int \mathcal{D}X \mathcal{D}Y P[X] P[Y|X] \log [P[Y|X]]$  is the conditional entropy responsible for the noise impact.

In our approach we use the formulation for the conditional PDF  $P[Y|X]$  through the path-integral (Martin-Siggia-Rose formalism [2]). In Ref.[3] the following (“quasiclassical”) representation for the conditional probability density function was obtained:

$$P[Y|X] = \Lambda \exp \left\{ -\frac{S[\Psi]}{Q} \right\}, \quad \Lambda = \int_{\phi(z=0)=0}^{\phi(z=L)=0} \mathcal{D}\phi \exp \left\{ -\frac{1}{Q} (S[\Psi + \phi] - S[\Psi]) \right\}, \quad (3)$$

where the action  $S[\psi]$  in (3) reads in the time domain

$$S[\psi] = \int_0^L dz \int_T dt |\mathcal{L}[\psi(z, t)]|^2, \quad \mathcal{L}[\psi] = \partial_z \psi(z, t) + i\beta \partial_t^2 \psi(z, t) - i\gamma \psi(z, t) |\psi(z, t)|^2. \quad (4)$$

Here  $T$  is a time interval containing both signals  $X(t)$  and  $Y(t)$ . In what follows we use the discretization scheme in the time and in the frequency domain and the discrete Fourier transform. There are some relations between  $T$  and discretization intervals in the time domain ( $\delta_t$  for the dense time grid with  $M'$  intervals and  $\tilde{\delta}_t$  for the coarse time sub-grid with  $M$  intervals) and in the frequency domain ( $\delta_\omega$ ):  $T = 1/\delta_\omega = M' \delta_t = M \tilde{\delta}_t = 2\pi M/W = 2\pi M'/W'$ . Here  $W$  is the bandwidth of the input signal, and  $W'$  is auxiliary bandwidth:  $W' \supset W$ . In the frequency discretization scheme we should perform the following substitutions in our equations:  $\partial_t^2 \Phi(z, t) \rightarrow -\Omega_k^2 \Phi_{\omega_k}(z)$ ,  $\delta(\omega_{i_1} + \omega_{i_2} + \dots) \rightarrow \frac{\Delta_{(M')}(i_1+i_2+\dots)}{2\pi\delta_\omega}$ ,  $\int \frac{d\omega}{2\pi} \dots \rightarrow \delta_\omega \sum_{j=0}^{M'-1}$ , where  $\omega_{n'} = -W'/2 + 2\pi\delta_\omega n'$ ,  $\Omega_{n'} = 2\sin[\pi n'/M'] M'/T = 2M'\delta_\omega \sin[\pi n'/M']$ ,  $n' = 0, 1, \dots, M' - 1$ , and we use the discrete analog of the Dirac delta-function  $\Delta_{(M)}(k) = \frac{1}{M} \sum_{n=0}^{M-1} \exp \{ -2\pi i \frac{n}{M} k \} = \sum_{m=-\infty}^{\infty} \delta_{k, mM}$ .

The function  $\Psi$  in Eq. (3) is the solution (referred to as “the classical solution”) of the Euler-Lagrange equation  $\delta S[\Psi] = 0$  with the boundary conditions:  $\Psi(0) = X$ ,  $\Psi(L) = Y$ . In the time domain this equation for  $\Psi(z, t)$  has a notably simple form [4]

$$\begin{aligned} (\partial_z + i\beta \partial_t^2 - 2i\gamma |\Psi(z, t)|^2) \mathcal{L}[\Psi(z, t)] + i\gamma \Psi^2(z, t) \overline{\mathcal{L}[\Psi(z, t)]} &= 0, \\ \mathcal{L}[\Psi(z, t)] &= (\partial_z + i\beta \partial_t^2 - i\gamma |\Psi(z, t)|^2) \Psi(z, t), \end{aligned} \quad (5)$$

and the function  $\Psi(z, t)$  obeys the boundary conditions:  $\Psi(0, t) = X(t)$ ,  $\Psi(L, t) = Y(t)$ .

It is convenient to introduce the function  $\Phi(z, t)$  which is the solution of the nonlinear Schrödinger equation (NLSE) with zero noise, i.e.,  $\mathcal{L}[\Phi(z, t)] = 0$ , and with the boundary condition  $\Phi(0, t) = X(t)$ . It is obvious that the function  $\Phi(z, t)$  obeys the equation Eq. (5) and the boundary condition at  $z = 0$ , but it does not obey the boundary condition at  $z = L$ . It globally minimizes the action as well:  $S[\Phi(z, t)] = 0$ . Since we imply that the noise power is much less than the signal power we can present the solution of Eq. (5) in the form

$$\Psi(z, t) = \Phi(z, t) + \varkappa(z, t), \quad (6)$$

where the function  $\varkappa(z, t)$  is of order of  $\sqrt{Q}$  for unsuppressed configurations  $\Psi(z, t)$ , since in the leading order in  $1/\text{SNR}$  the action is quadratic functional in  $\varkappa(z, t)$ , see details in [4]. Therefore

we substitute the function  $\Psi$  in the form (6) to the Eq. (5), then linearizing Eq. (5) in  $\varkappa(z, t)$  we obtain the following linear problem on  $\varkappa(z, t)$ :

$$\begin{aligned} (\partial_z + i\beta\partial_t^2 - 2i\gamma|\Phi(z, t)|^2) l[\varkappa(z, t)] + i\gamma\Phi^2(z, t)\overline{l[\varkappa(z, t)]} &= 0, \\ l[\varkappa(z, t)] &= (\partial_z + i\beta\partial_t^2) \varkappa(z, t) - i\gamma \left( 2\varkappa(z, t)|\Phi(z, t)|^2 + \overline{\varkappa(z, t)}\Phi^2(z, t) \right), \end{aligned} \quad (7)$$

with the boundary conditions  $\varkappa(z = 0, t) = 0$ ,  $\varkappa(z = L, t) = Y(t) - \Phi(L, t) \equiv \delta Y(t)$ .

In some sense, the introduction of the function  $\varkappa(z, t) \sim \sqrt{Q}$  is the crucial idea of our consideration in the case of large SNR, since it allows us to reduce the difficult nonlinear problem of  $P[Y|X]$  calculation to the linear one.

## 2. Factorization of pre-exponential factor $\Lambda$

Let us consider the separation of different scales in the conditional PDF  $P[Y|X]$  when PDF  $P[Y|X]$  is considered under an integral over  $X$  together with the input signal PDF  $P[X]$  which has the following form in the frequency domain:  $P[\vec{X}] = P_X^{(M)}[\vec{X}_1]\delta(\vec{X}_2)$ , where  $\delta(\vec{X}_2)$  means  $2(M' - M)$ -dimensional delta-function corresponding  $M' - M$  complex remnant channels. The vector notations ( $\vec{X} = \vec{X}_1 \oplus \vec{X}_2$ ) relate to the frequency domain:  $\vec{X}_1$  is  $2M$ -dimensional vector corresponding to  $M$  meaning complex channels in the frequency domain  $W$ , whereas  $\vec{X}_2$  is  $2(M' - M)$ -dimensional vector corresponding to remnant  $M' - M$  complex channels in the domain  $W' \setminus W$ . In the leading order in  $1/\text{SNR}$  the conditional PDF  $P[Y|X]$  reads

$$P[Y|X] \approx \Lambda[X] e^{-S_2[\varkappa]/Q}, \quad (8)$$

where the action  $S_2[\varkappa] = \delta_t \sum_{k=0}^{M'-1} \Delta \sum_{n=1}^{N-1} \mathcal{L}_{eff}[\varkappa(z_n, t_k)]$  is quadratic functional in  $\phi$ , where  $\delta_t = T/M'$  is the discretization parameter in the time domain:  $t_k = k\delta_t$ ,  $k = 0, 1, \dots, M' - 1$ . Here  $\Delta = L/N$  is the distance discretization parameter:  $z_n = n\Delta$ ,  $n = 1, 2, \dots, N - 1$ . We have introduced the ‘‘Lagrangian’’:

$$\mathcal{L}_{eff}[\varkappa] = \left| \partial_z \varkappa(z_n, t_k) + i\beta\partial_t^2 \varkappa(z_n, t_k) - i\gamma \left( 2\varkappa(z_n, t_k)|\Phi(z_n, t_k)|^2 + \bar{\varkappa}(z_n, t_k)\Phi^2(z_n, t_k) \right) \right|^2. \quad (9)$$

Here derivatives should be regarded as difference derivatives in our discretization scheme. The function  $\varkappa(z, t)$  in the exponent (8) is the solution of the Euler-Lagrange equation  $-\delta\mathcal{L}_{eff}[\varkappa]/\delta\bar{\varkappa} = 0$  with the boundary conditions  $\varkappa(z = 0, t) = 0$ ,  $\varkappa(z = L, t) = Y(t) - \Phi(L, t)$ . The normalization factor  $\Lambda[X]$  has the form

$$\Lambda[X] = \int_{\phi(0,t)=0}^{\phi(L,t)=0} \mathcal{D}\phi \exp \left\{ -\frac{\delta_t}{Q} \sum_{k=0}^{M'-1} \Delta \sum_{n=1}^{N-1} \mathcal{L}_{eff}[\phi(z_n, t_k)] \right\}. \quad (10)$$

Note that the sum in expression for  $S_2[\varkappa]$  is performed over the dense time grid. To demonstrate the factorization we have to separate the scales in the action into the coarse and dense parts. In other words, we have to separate the summation over  $M$  meaning channels and  $M' - M$  remnant channels. The scale separation procedure in some sense is similar to Wilson’s renormalization procedure for the Lagrangian  $\mathcal{L}_{eff}[\varkappa]$ , see [5]. But in our approximation the Lagrangian (9) is quadratic functional in  $\varkappa$  that is why there are no corrections to the effective action when we perform integration over remnant  $2(M' - M)$  degrees of freedom  $\varkappa(z, t_k)$  where  $t_k$  runs through values only on the dense grid without the coarse sub-grid. Let us demonstrate this fact.

First we perform the separation of variables:

$$\begin{aligned}\varkappa(z, t_k) &= \varkappa^{(c)}(z, t_k) + \varkappa^{(d)}(z, t_k), \\ Y(t_k) &= Y^{(c)}(t_k) + Y^{(d)}(t_k),\end{aligned}\quad (11)$$

where  $\varkappa^{(c)}(z, t_k)$ , or  $Y^{(c)}(t_k)$ , is completely defined only by the values  $\varkappa^{(c)}(z, T_{\tilde{i}})$ , or  $Y^{(c)}(T_{\tilde{i}})$ , on the coarse time grid  $T_{\tilde{i}} = \tilde{i}\tilde{\delta}_t$ ,  $\tilde{i} = 0, 1, \dots, M-1$ . Here and below superscript “(c)” means coarse variable. In other words, the function  $\varkappa^{(c)}(z, t_k)$  evaluated at all grid points is the interpolation of some order (i.e., the interpolating polynomial degree)  $N_0 > 2$  calculated on the base of values  $\varkappa^{(c)}(z, T_{\tilde{i}})$  of the coarse time grid. The function  $\varkappa^{(c)}(z, t_k)$  coincides with  $\varkappa(z, T_{\tilde{i}})$  when  $t_k$  falls on the coarse time grid  $T_{\tilde{i}}$  (i.e.,  $k = [\tilde{i}M'/M]$ ,  $\tilde{i} = 0, 1, \dots, M-1$ ), i.e.,  $\varkappa^{(d)}(z, t_k) = 0$  on the coarse grid. In other grid points of the dense grid the function  $\varkappa^{(c)}(z, t_k)$  smoothly interpolates the values of  $\varkappa(z, t_k)$  with interpolation order  $N_0 > 2$ :  $\varkappa^{(d)} = \mathcal{O}(\tilde{\delta}_t^{N_0})$  and  $\partial_t^2 \varkappa^{(d)}(z, t_k) = \mathcal{O}(\tilde{\delta}_t^{N_0-2})$ , where we have used that  $\partial_t^2 \varkappa(z, t_k) = \partial_t^2 \varkappa^{(c)}(z, t_k) + \partial_t^2 \varkappa^{(d)}(z, t_k)$ , here the derivatives are assumed as the difference derivatives on the dense grid. The boundary conditions are as follows:

$$\begin{aligned}\varkappa^{(c)}(0, t_k) &= 0, & \varkappa^{(c)}(L, t_k) &= Y^{(c)}(t_k) - \Phi^{(c)}(L, t_k), \\ \varkappa^{(d)}(0, t_k) &= 0, & \varkappa^{(d)}(L, t_k) &= Y^{(d)}(t_k) - \Phi^{(d)}(L, t_k),\end{aligned}\quad (12)$$

where we have used that  $\Phi(L, t_k)$  has the coarse and dense parts as well:

$$\Phi(z, t_k) = \Phi^{(c)}(z, t_k) + \Phi^{(d)}(z, t_k), \quad \Phi^{(d)}(z, t_k) = \mathcal{O}(\tilde{\delta}_t^{N_0}). \quad (13)$$

Note that if we consider (8) under the integral over  $\mathcal{D}X$  with the input signal PDF, then the function  $\Phi(z, t_k)$  is the (nonlinear) function of the input signal only on the coarse time grid  $X(T_{\tilde{i}})$ . This means that the dense part  $\Phi^{(d)}(z, t_k) = \mathcal{O}(\tilde{\delta}_t^{N_0})$  is always small for sufficiently large  $M$ . Now we insert the representation (11) in the action  $S_2$ . The action fractionizes into three parts

$$\begin{aligned}S_2[\varkappa] &= \delta_t \sum_{k=0}^{M'-1} \Delta \sum_{n=1}^{N-1} \mathcal{L}_{eff}[\varkappa(z_n, t_k)] = \delta_t \sum_{k=0}^{M'-1} \Delta \sum_{n=1}^{N-1} \mathcal{L}_{eff}[\varkappa^{(c)}(z_n, t_k)] + \\ &\delta_t \sum_{k=0}^{M'-1} \Delta \sum_{n=1}^{N-1} \mathcal{L}_{eff}[\varkappa^{(d)}(z_n, t_k)] + \delta_t \sum_{k=0}^{M'-1} \Delta \sum_{n=1}^{N-1} \mathcal{L}_{int}[\varkappa^{(c)}(z_n, t_k), \varkappa^{(d)}(z_n, t_k)],\end{aligned}\quad (14)$$

where the third part with interaction of coarse ( $\varkappa^{(c)}$ ) and dense ( $\varkappa^{(d)}$ ) degrees of freedom contains Lagrangian

$$\begin{aligned}\mathcal{L}_{int}[\varkappa^{(c)}(z, t_k), \varkappa^{(d)}(z, t_k)] &= \\ &\left( \partial_z \varkappa^{(c)}(z, t_k) + i\beta \partial_t^2 \varkappa^{(c)}(z, t_k) - i\gamma \left( 2\varkappa^{(c)}(z, t_k) |\Phi(z, t_k)|^2 + \bar{\varkappa}^{(c)}(z, t_k) \Phi^2(z, t_k) \right) \right) \times \\ &\left( \partial_z \bar{\varkappa}^{(d)}(z, t_k) - i\beta \partial_t^2 \bar{\varkappa}^{(d)}(z, t_k) + i\gamma \left( 2\bar{\varkappa}^{(d)}(z, t_k) |\Phi(z, t_k)|^2 + \varkappa^{(d)}(z, t_k) \bar{\Phi}^2(z, t_k) \right) \right) + c.c.\end{aligned}\quad (15)$$

Here “c.c.” means the same complex conjugated term.

The first part in the r.h.s. of Eq. (14) can be simplified as follows:

$$\delta_t \sum_{k=0}^{M'-1} \Delta \sum_{n=1}^{N-1} \mathcal{L}_{eff}[\varkappa^{(c)}(z_n, t_k)] = \tilde{\delta}_t \sum_{\tilde{i}=0}^{M-1} \Delta \sum_{n=1}^{N-1} \mathcal{L}_{eff}[\phi(z_n, T_{\tilde{i}})] \left( 1 + \mathcal{O}(\tilde{\delta}_t^{N_0-2}) \right), \quad (16)$$

where  $\tilde{\delta}_t = \frac{2\pi}{W} = \delta_t \left[ M'/M \right]$  is the grid spacing of the coarse time grid. Here we have replaced every term under the sum over the dense grid with its average value on the coarse grid. The accuracy in Eq. (16) is governed by the interpolation order of the second derivative in time.

The third part in the r.h.s. of Eq. (14) can be integrated (summed) over  $z$  by part resulting in the following expression:

$$\delta_t \sum_{k=0}^{M'-1} \Delta \sum_{n=1}^{N-1} \mathcal{L}_{int}[\boldsymbol{\varkappa}^{(c)}(z_n, t_k), \boldsymbol{\varkappa}^{(d)}(z_n, t_k)] = \delta_t \sum_{k=0}^{M'-1} \Delta \sum_{n=1}^{N-1} \left( \bar{\boldsymbol{\varkappa}}^{(d)}(z_n, t_k) \frac{\delta \mathcal{L}_{eff}[\boldsymbol{\varkappa}^{(c)}]}{\delta \boldsymbol{\varkappa}} + c.c. \right) + S_{surf}, \quad (17)$$

where the variation  $-\delta \mathcal{L}_{eff}[\boldsymbol{\varkappa}^{(c)}]/\delta \boldsymbol{\varkappa}$  is linear in  $\boldsymbol{\varkappa}^{(c)}$ , and it represents the l.h.s. of the Euler-Lagrange equation, i.e., for the function  $\boldsymbol{\varkappa}^{(c)}$  we obtain  $\delta \mathcal{L}_{eff}[\boldsymbol{\varkappa}^{(c)}]/\delta \boldsymbol{\varkappa} = \mathcal{O}(\tilde{\delta}_t^{N_0-2})$ , i.e., it is always small. The term  $S_{surf}$  results from the surface term in integration by part over  $z$  in Eq. (17), and taking into account the boundary conditions (12) it reads

$$S_{surf} = \delta_t \sum_{k=0}^{M'-1} \left[ \bar{Y}^{(d)}(t_k) - \bar{\Phi}^{(d)}(L, t_k) \right] \left( \partial_z \boldsymbol{\varkappa}^{(c)}(L, t_k) + i\beta \partial_t^2 \boldsymbol{\varkappa}^{(c)}(L, t_k) - i\gamma \left( 2\boldsymbol{\varkappa}^{(c)}(L, t_k) |\Phi(L, t_k)|^2 + \bar{\boldsymbol{\varkappa}}^{(c)}(L, t_k) \Phi^2(L, t_k) \right) \right) + c.c. \quad (18)$$

We can omit the surface term (18), since it is linear both in the coarse and dense variables, but they are orthogonal when integrating over  $t$  (they have not intersecting supports in the frequency domain). It is obvious for the first two terms in the parentheses in Eq. (18). The last terms containing  $\Phi(L, t_k)$  are coarse variables as well: we can replace  $\Phi(L, t_k)$  with  $\Phi^{(c)}(L, t_k)$  with the interpolation accuracy  $\mathcal{O}(\tilde{\delta}_t^{N_0})$  and then replace  $\Phi^{(c)}(L, t_k)$  with  $Y^{(c)}(t_k)$  with the accuracy  $\mathcal{O}(\sqrt{Q})$  (we remind that  $\boldsymbol{\varkappa}$  in Eq. (12) is of order of  $\sqrt{Q}$ ). Then we can replace  $Y^{(c)}(t_k)$  with the constants inside the whole interval of an coarse space with the interpolation accuracy  $\mathcal{O}(\tilde{\delta}_t)$  and now use the orthogonality of the coarse and dense variables.

To summarize, with the accuracy of our interpolation  $\mathcal{O}(\tilde{\delta}_t) = \mathcal{O}(1/M)$  we can omit the interaction term (17), and our action fractionizes into coarse and dense parts:  $S_2[\boldsymbol{\varkappa}] \approx S_2[\boldsymbol{\varkappa}^{(c)}] + S_2[\boldsymbol{\varkappa}^{(d)}]$ . Both actions are expressed through the same Lagrangian (9) and are quadratic forms. The coefficients of these quadratic forms depend on input signal  $X$  only. The factorization of  $\Lambda$  can be shown using the normalization condition:  $1 = \int \mathcal{D}Y P[Y|X] = \Lambda \int \mathcal{D}Y \exp \{-S_2[\boldsymbol{\varkappa}]/Q\}$ , where we have used that  $\Lambda$  does not depend on  $Y$  in the leading order in  $1/\text{SNR}$ . Finally, one has

$$\Lambda^{-1} = \int \mathcal{D}Y e^{-\frac{S_2[\boldsymbol{\varkappa}^{(c)}]}{Q} - \frac{S_2[\boldsymbol{\varkappa}^{(d)}]}{Q}} = \int \mathcal{D}Y^{(c)} e^{-\frac{S_2[\boldsymbol{\varkappa}^{(c)}]}{Q}} \mathcal{D}Y^{(d)} e^{-\frac{S_2[\boldsymbol{\varkappa}^{(d)}]}{Q}} = \Lambda_1^{-1} \Lambda_2^{-1}, \quad (19)$$

or  $\Lambda = \Lambda_1 \times \Lambda_2$ . Here the normalization factor  $\Lambda_1$  depends on the input signal  $X$  on the coarse grid only and it reads

$$\Lambda_1 = \int_{\phi(0,t)=0}^{\phi(L,t)=0} \left[ \mathcal{D}\phi(z, t) \right]_M \exp \left\{ -\frac{\tilde{\delta}_t}{Q} \sum_{\tilde{i}=0}^{M-1} \Delta \sum_{n=1}^{N-1} \mathcal{L}_{eff}[\phi(z_n, T_{\tilde{i}})] \right\} \quad (20)$$

with the Lagrangian  $\mathcal{L}_{eff}[\phi(z_n, T_{\tilde{i}})]$ , see (9), considered on the coarse time grid  $T_{\tilde{i}}$  only.

The measure  $\left[ \mathcal{D}\phi(z, t) \right]_M$  on the coarse time grid is defined as

$$\left[ \mathcal{D}\phi(z, t) \right]_M = \lim_{\tilde{\delta}_t \rightarrow 0} \lim_{\Delta \rightarrow 0} \left( \frac{\tilde{\delta}_t}{\Delta \pi Q} \right)^M \prod_{\tilde{j}=0}^{M-1} \prod_{i=1}^{N-1} \left\{ \frac{\tilde{\delta}_t}{\Delta \pi Q} d\text{Re}\phi(z_i, T_{\tilde{j}}) d\text{Im}\phi(z_i, T_{\tilde{j}}) \right\}. \quad (21)$$

The normalization factor  $\Lambda_1$  corresponds to  $M$  meaning complex channels, and  $\Lambda_2$  corresponds to  $M' - M$  complex remnant channels. In this demonstration we have used that the quantity  $P[Y|X]$  is considered under the integral over  $\mathcal{D}X$  with the input signal PDF  $P[X]$ . The accuracy of our factorization is at least  $\mathcal{O}(\delta_t) = \mathcal{O}(1/M)$ .

### 3. Perturbative expansion vs Jensen's inequality for mutual information

The factorization of  $\Lambda$  guarantees, see details in [4], that the mutual information has the following form in the leading order in  $1/\text{SNR}$  and for arbitrary nonlinearity:

$$I_{P[X]} = H[X] - M + \int d\vec{X}_1 P_X^{(M)}[\vec{X}_1] \log \Lambda_1[\vec{X}_1], \quad (22)$$

where  $\vec{X}_1$  is  $2M$  dimensional vector corresponding  $M$  complex meaning channels. We can calculate  $\Lambda_1[\vec{X}_1]$  within the perturbation theory in dimensionless nonlinearity parameter  $\tilde{\gamma} = \gamma LPW/(2\pi)$  and perform the averaging over Gaussian input signal PDF  $P_X^{(M)}$ :  $P_X^{(M)}[\vec{X}_1] = P_G[\vec{X}_1] = \Lambda_P e^{-|\vec{X}_1|^2 \delta_\omega/P}$ , where  $\Lambda_P = (\delta_\omega/(\pi P))^M$ . We will denote the last term in (22) as  $\int d\vec{X}_1 P_X^{(M)}[\vec{X}_1] \log \Lambda_1[\vec{X}_1] \equiv \langle \log \Lambda_1 \rangle_X$ . For the averaging over Gaussian input signal PDF we use the Wick's theorem [5] and the following correlator

$$\langle X(\omega_k) \bar{X}(\omega_{k'}) \rangle_X = P \delta_{k, k'} / \delta_\omega, \quad k, k' = 0, 1, \dots, M-1. \quad (23)$$

From the representation (20) we obtain the expression for  $\Lambda_1[\vec{X}_1]$  in the frequency domain

$$\Lambda_1[\vec{X}_1] = \int_{\phi(0, \omega)=0}^{\phi(L, \omega)=0} [\mathcal{D}\phi(z, \omega)]_M \exp \left\{ -\frac{\delta_\omega}{Q} \sum_{k=0}^{M-1} \Delta \sum_{n=1}^{N-1} \mathcal{L}_{eff}[\phi(z_n, \omega_k)] \right\}, \quad (24)$$

where the measure reads

$$[\mathcal{D}\phi(z, \omega)]_M = \lim_{\delta_\omega \rightarrow 0} \lim_{\Delta \rightarrow 0} \left( \frac{\delta_\omega}{\Delta \pi Q} \right)^M \prod_{n=1}^{N-1} \prod_{k=0}^{M-1} \left\{ \frac{\delta_\omega}{\Delta \pi Q} d\text{Re}\phi(z_n, \omega_k) d\text{Im}\phi(z_n, \omega_k) \right\}. \quad (25)$$

Now we present the ‘‘Lagrangian’’  $\mathcal{L}_{eff}$  as a sum:  $\mathcal{L}_{eff} = \mathcal{L}_{eff}^{(0)} + \mathcal{L}_{eff}^{(1)} + \mathcal{L}_{eff}^{(2)}$ , where the first term reads  $\mathcal{L}_{eff}^{(0)}[\phi(z_n, \omega_k)] = |(\partial_z - i\beta \bar{\Omega}_k^2) \phi(z_n, \omega_k)|^2$ . Here we have introduced  $\bar{\Omega}_k = 2 \sin[\pi k/M] M/T' = 2M \delta_\omega \sin[\pi k/M] = W \sin[\pi k/M]/\pi$ . In the continuous limit  $M \rightarrow \infty$  we can assume that  $\bar{\Omega}_k \approx 2\pi \delta_\omega k = \omega_k - \omega_0$ . The second and the third terms are the terms of  $\mathcal{L}_{eff}$ , see Eq. (9) in the frequency representation, explicitly proportional to  $\tilde{\gamma}$  and  $\tilde{\gamma}^2$ , correspondingly.

*Perturbative calculation.* We present the perturbation expansion in  $\tilde{\gamma}$  of the normalization factor  $\Lambda_1$  in the form:  $\Lambda_1 = \gamma^0 \Lambda_1^{(0)} + \gamma \Lambda_1^{(1)} + \gamma^2 \Lambda_1^{(2)} + \mathcal{O}(\tilde{\gamma}^3)$ . Thus the last term in the expression (22) for the mutual information has the following expansion in  $\gamma$ :  $\langle \log \Lambda_1 \rangle_X = \langle \log \Lambda_1^{(0)} \rangle_X + \gamma \left\langle \frac{\Lambda_1^{(1)}}{\Lambda_1^{(0)}} \right\rangle_X + \gamma^2 \left( \left\langle \frac{\Lambda_1^{(2)}}{\Lambda_1^{(0)}} \right\rangle_X - \frac{1}{2} \left\langle \left[ \frac{\Lambda_1^{(1)}}{\Lambda_1^{(0)}} \right]^2 \right\rangle_X \right) + \mathcal{O}(\tilde{\gamma}^3)$ . Retaining only the first term  $\mathcal{L}_{eff}^{(0)}$  in Eq. (24) we have:

$$\Lambda_1^{(0)} = \int_{\phi(0, \omega)=0}^{\phi(L, \omega)=0} [\mathcal{D}\phi(z, \omega)]_M \exp \left\{ -\frac{\delta_\omega}{Q} \sum_{k=0}^{M-1} \Delta \sum_{n=1}^{N-1} |(\partial_z - i\beta \bar{\Omega}_k^2) \phi(z_n, \omega_k)|^2 \right\} = \left( \frac{\delta_\omega}{\pi Q L} \right)^M. \quad (26)$$

This expression for  $\log \Lambda_1^{(0)}$  results in the main (Shannon's) contribution  $M \log [P/(QL)]$  to the mutual information (22) by taking into account that  $H[X] = M - M \log [\delta_\omega/(\pi P)]$ .

We introduce the averaging  $\langle \dots \rangle_\phi$  over fields  $\phi(z, \omega)$  defined as

$$\langle \dots \rangle_\phi = \frac{1}{\Lambda_1^{(0)}} \int_{\phi(0, \omega)=0}^{\phi(L, \omega)=0} [\mathcal{D}\phi(z, \omega)]_M \dots \exp \left\{ -\frac{\delta_\omega}{Q} \sum_{k=0}^{M-1} \Delta \sum_{n=1}^{N-1} |(\partial_z - i\beta \bar{\Omega}_k^2) \phi(z_n, \omega_k)|^2 \right\}. \quad (27)$$

The paired correlator has the form:

$$\langle \phi(z, \omega_k) \bar{\phi}(z', \omega_{k'}) \rangle_\phi = -\frac{Q}{\delta_\omega} \delta_{k, k'} G(z, z') \exp [i\beta \bar{\Omega}_k^2 (z - z')], \quad (28)$$

where the Green function reads  $G(z, z') = z \frac{z'-L}{L} \theta(z' - z) + z' \frac{z-L}{L} \theta(z - z')$ . In what follows for brevity sake we will write the sum over  $z$  as the integral. Let us stress that in Eq. (27) and hereinafter the derivative with respect to  $z$  is assumed in the "causative" manner,  $\partial_z \phi(z_n, \omega_k) = (\phi(z_{n+1}, \omega_k) - \phi(z_n, \omega_k)) / \Delta$ , as provided by our approach [3].

In the notations (27) we can present the mutual information (22) in the form

$$I_{P[X]} = M \log \left[ \frac{P}{QL} \right] + \left\langle \log \langle e^{-S_{nl}[\phi(z, \omega)]/Q} \rangle_\phi \right\rangle_X, \quad S_{nl} = \int_0^L dz \delta_\omega \sum_{k=0}^{M-1} (\mathcal{L}_{eff}^{(1)} + \mathcal{L}_{eff}^{(2)}). \quad (29)$$

It is easy to see by the direct calculation that there are no corrections to the mutual information (22) of order of  $\tilde{\gamma}$ :  $\gamma \left\langle \frac{\Lambda_1^{(1)}}{\Lambda_1^{(0)}} \right\rangle_X = 0$ . The second order calculation is cumbersome but straightforward. One can find the details of this perturbative calculations in Ref. [4]. Finally we obtain the following expression for the mutual information

$$I_{P_G[X]} = M \log \text{SNR} - M \frac{\tilde{\gamma}^2}{3} g(\tilde{\beta}) + \mathcal{O}(\tilde{\gamma}^3), \quad (30)$$

where  $M$  is the number of complex meaning channels. This number is implied to be large  $M \gg 1$ . The parameter  $\tilde{\beta} = \beta L W^2$  is the dimensionless dispersion parameter. The function  $g(\tilde{\beta})$  in discretization scheme can be presented as the triple sum:

$$g(\tilde{\beta}) = \frac{1}{M^3} \sum_{k_1, k_2, k_3=0}^{M-1} F \left( \frac{\tilde{\beta}}{2} [\bar{\Omega}_{k_1}^2 + \bar{\Omega}_{k_2}^2 - \bar{\Omega}_{k_3}^2 - \bar{\Omega}_{k_1+k_2-k_3}^2] \right). \quad (31)$$

The elementary function  $F(\mu)$  in Eq. (31) is the result of integration of the derivatives of the dimensionless Green function,  $G_0(\zeta_1, \zeta_2) = \zeta_1(\zeta_2 - 1)\theta(\zeta_2 - \zeta_1) + \zeta_2(\zeta_1 - 1)\theta(\zeta_1 - \zeta_2)$ :

$$F(\mu) = -12 \int_0^1 d\zeta_1 \int_0^1 d\zeta_2 \frac{\partial G_0(\zeta_1, \zeta_2)}{\partial \zeta_1} \frac{\partial G_0(\zeta_1, \zeta_2)}{\partial \zeta_2} e^{-2i\mu(\zeta_1 - \zeta_2)} = 3 \frac{\mu^2 - \sin^2(\mu)}{\mu^4}, \quad (32)$$

and for convenience it is normalized as  $F(0) = 1$ . In the continuous limit of sufficiently large  $M$

we can present  $g(\tilde{\beta})$  as the integral, as the series, and as the explicit hypergeometric expression:

$$\begin{aligned}
 g(\tilde{\beta}) &= \int_0^1 dx_1 \int_0^1 dx_2 \int_0^1 dx_3 F\left(\frac{\tilde{\beta}}{4}(x_1 - x_3)(x_2 - x_3)\right) = \\
 &4! \sum_{n=0}^{\infty} \frac{(-1)^n \tilde{\beta}^{2n} [(4n+2)! + (1+2n)!^2]}{2^{2n-1} (2n+4)! (4n+3)! (1+2n)^2} = \\
 &\frac{1}{33075 \tilde{\beta}^4} \left\{ -3528000 \tilde{\beta}^2 {}_2F_3\left(1, 1; \frac{1}{4}, \frac{3}{4}, 2; -\frac{\tilde{\beta}^2}{256}\right) + 140 \tilde{\beta}^6 {}_2F_3\left(\frac{3}{2}, 2; \frac{9}{4}, \frac{11}{4}, \frac{7}{2}; -\frac{\tilde{\beta}^2}{256}\right) - \right. \\
 &576 \tilde{\beta}^6 {}_2F_3\left(\frac{7}{4}, 2; \frac{9}{4}, \frac{11}{4}, \frac{11}{4}; -\frac{\tilde{\beta}^2}{256}\right) + 210 \tilde{\beta}^6 {}_2F_3\left(2, \frac{5}{2}; \frac{9}{4}, \frac{11}{4}, \frac{7}{2}; -\frac{\tilde{\beta}^2}{256}\right) - \\
 &14 \tilde{\beta}^6 {}_3F_4\left(\frac{3}{2}, \frac{3}{2}, 2; \frac{9}{4}, \frac{5}{2}, \frac{11}{4}, \frac{7}{2}; -\frac{\tilde{\beta}^2}{256}\right) - 588000 \tilde{\beta}^4 {}_1F_2\left(\frac{1}{2}; \frac{3}{2}, \frac{5}{2}; -\frac{\tilde{\beta}^2}{16}\right) + \\
 &564480 \tilde{\beta}^4 {}_1F_2\left(\frac{3}{4}; \frac{7}{4}, \frac{5}{2}; -\frac{\tilde{\beta}^2}{16}\right) - 270480 \tilde{\beta}^4 {}_1F_2\left(1; \frac{5}{4}, \frac{7}{4}; -\frac{\tilde{\beta}^2}{256}\right) + \\
 &176400 \tilde{\beta}^4 {}_2F_3\left(\frac{1}{2}, \frac{1}{2}; \frac{3}{2}, \frac{3}{2}, \frac{5}{2}; -\frac{\tilde{\beta}^2}{16}\right) - 294000 \tilde{\beta}^4 {}_2F_3\left(\frac{1}{2}, 1; \frac{5}{4}, \frac{7}{4}, \frac{5}{2}; -\frac{\tilde{\beta}^2}{256}\right) + \\
 &564480 \tilde{\beta}^4 {}_2F_3\left(\frac{3}{4}, 1; \frac{5}{4}, \frac{7}{4}, \frac{7}{4}; -\frac{\tilde{\beta}^2}{256}\right) + 33075 \tilde{\beta}^4 {}_2F_3\left(1, 1; \frac{5}{4}, \frac{7}{4}, 3; -\frac{\tilde{\beta}^2}{256}\right) - \\
 &147000 \tilde{\beta}^4 {}_2F_3\left(1, \frac{3}{2}; \frac{5}{4}, \frac{7}{4}, \frac{5}{2}; -\frac{\tilde{\beta}^2}{256}\right) + 88200 \tilde{\beta}^4 {}_3F_4\left(\frac{1}{2}, \frac{1}{2}, 1; \frac{5}{4}, \frac{3}{2}, \frac{7}{4}, \frac{5}{2}; -\frac{\tilde{\beta}^2}{256}\right) + \\
 &\left. 352800 \tilde{\beta}^2 + 6209280 \tilde{\beta} \sin\left(\frac{\tilde{\beta}}{2}\right) - 564480 \cos\left(\frac{\tilde{\beta}}{2}\right) + 564480 \right\}, \tag{33}
 \end{aligned}$$

where  ${}_pF_q$  is generalized hypergeometric function.

For the large  $\tilde{\beta}$  we can deal with the asymptotics of the function  $g(\tilde{\beta})$  at large  $\tilde{\beta}$ :

$$g(\tilde{\beta}) = \frac{16\pi}{\tilde{\beta}} \left( \log[\tilde{\beta}] - \log[2] + \gamma_E - \frac{23}{6} \right) + \mathcal{O}\left(\frac{1}{\tilde{\beta}^{3/2}}\right), \quad \tilde{\beta} \rightarrow \infty. \tag{34}$$

The function  $g(\tilde{\beta})$ , see Eq. (33), and the asymptotics (34) for large  $\tilde{\beta}$  are presented in Fig. 1.

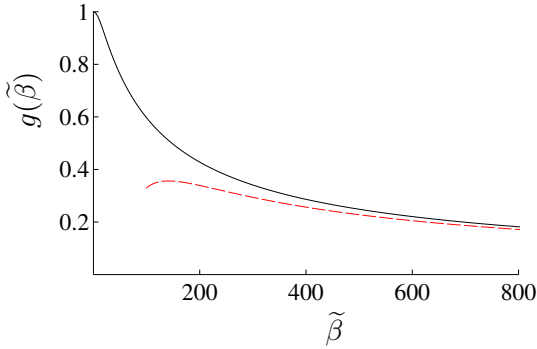
For  $\tilde{\beta} = 0$  we have  $g(\tilde{\beta} = 0) = 1$  and we arrive at the result [6] for the nondispersive channel in expansion in  $\tilde{\gamma}$ :

$$I_{P[X]}^{(\beta=0)} = M \log \text{SNR} - \frac{M}{2} \int_0^{\infty} d\tau e^{-\tau} \log\left(1 + \frac{\tau^2 \tilde{\gamma}^2}{3}\right) = M \log [\text{SNR}] - M \frac{\tilde{\gamma}^2}{3} + \mathcal{O}(\tilde{\gamma}^4). \tag{35}$$

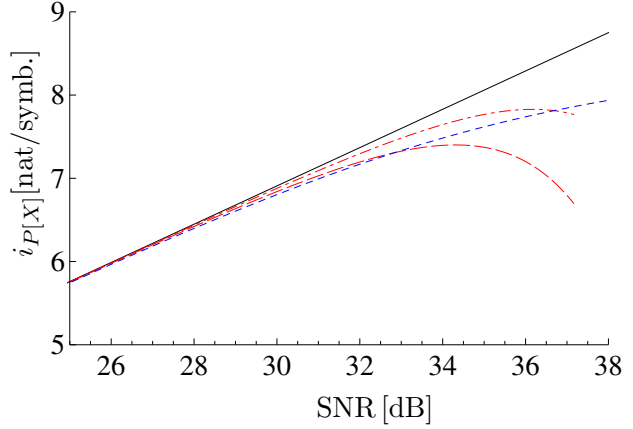
Let us estimate the mutual information  $I_{P[X]}$  for typical fiber optical links [7]:  $\beta = 20 \text{ ps}^2/\text{km}$ ,  $L = 1000 \text{ km}$ ,  $\gamma = 1.31(\text{Wkm})^{-1}$ ,  $W = 100 \text{ GHz}$ ,  $P_{\text{noise}} = QLW/(2\pi) = 5.3 \times 10^{-4} \text{ mW}$ . For these parameters one has  $\tilde{\beta} = \beta LW^2 \approx 200$ , and  $g(\tilde{\beta}) \approx 0.42$ . From Eq. (30) we obtain

$$I_{P[X]} \approx M \{ \log [\text{SNR}] - 7 \times 10^{-8} \times \text{SNR}^2 \}. \tag{36}$$





**Figure 1.** The function  $g(\tilde{\beta})$  (the solid black line), see Eq. (33), and the asymptotics of  $g(\tilde{\beta})$  for large  $\tilde{\beta}$  (red dashed line), see Eq. (34).



**Figure 2.** The spectral efficiency  $i_{P[X]} = I_{P[X]}/M$  corresponding to Eq. (30) for different  $\tilde{\beta}$ : Shannon's result (the solid black line),  $\tilde{\beta} = 200$  (red long-dashed line),  $\tilde{\beta} = 0$  (blue dashed line),  $\tilde{\beta} = 800$  (red chain line).

For these parameters we present the spectral efficiency  $i_{P[X]} = I_{P[X]}/M$  for different dispersion parameters  $\tilde{\beta}$  in Fig. 2. Increasing parameter  $\tilde{\beta}$  the first nonlinear correction, see Eq. (30), and the asymptotics, see Eq. (34), goes to zero as  $\tilde{\gamma}^2 \log(\tilde{\beta})/\tilde{\beta}$ . Therefore for larger  $\tilde{\beta}$  the result (30) is closer to Shannon's result than the result (35) in wider region in SNR.

*Low bound from the Jensen's inequality.* From the direct calculation using Eq. (28) and Wick's theorem [5] we have the following average for the action part with nonlinearity terms

$$-\left\langle \frac{\langle S_{nl}[\phi(z, \omega)] \rangle_\phi}{Q} \right\rangle_X = 5\tilde{\gamma}^2 \int_T dt \int_0^L dz G(z, z) \left\langle |\Phi(z, t)|^4 \right\rangle_X. \quad (37)$$

In the first nonvanishing order for small  $\tilde{\gamma}$  we have  $-\left\langle \frac{\langle S_{nl}[\phi(z, \omega)] \rangle_\phi}{Q} \right\rangle_X = -\frac{5}{3}M\tilde{\gamma}^2 + \mathcal{O}(\tilde{\gamma}^4)$ .

The relation (37) plays significant role since it delivers the low bound for the capacity. Indeed, from the Jensen's inequality we have

$$\left\langle \exp \left[ -\frac{S_{nl}[\phi(z, \omega)]}{Q} \right] \right\rangle_\phi \geq \exp \left[ -\frac{\langle S_{nl}[\phi(z, \omega)] \rangle_\phi}{Q} \right], \quad (38)$$

and it means that we can find the low bound for the mutual information:

$$I_{P[X]} \geq M \log [\text{SNR}] - \left\langle \frac{\langle S_{nl}[\phi(z, \omega)] \rangle_\phi}{Q} \right\rangle_X. \quad (39)$$

The representation (39) is convenient for the nonperturbative analysis. For instance, it is useful for the arbitrary nonlinearity at large or at small  $\tilde{\beta}$ , i.e., when the explicit expression for the NLSE solution  $\Phi(z, t)$  is known. Let us consider large dispersion parameter  $\tilde{\beta}$ . From the asymptotic representation of the NLSE solution [8] we obtained that for large  $z$ ,  $z \gg z_0 = L/\tilde{\beta}$ ,

$$\Phi(z, t) \sim \frac{1}{\sqrt{4\pi\beta z}} X(\omega = \frac{t}{2\beta z}) \exp \left\{ -i \frac{t^2}{4\beta z} - i \frac{\pi}{4} + i \frac{\gamma}{4\pi\beta} \left| X(\omega = \frac{t}{2\beta z}) \right|^2 \log \left[ \frac{z}{z_0} \right] \right\}. \quad (40)$$

This equation demonstrates that for the case of large  $\tilde{\beta} \gg 1$  the low bound, i.e., r.h.s. of Eq. (39), tends to the Shannon limit. The same result is obvious in the perturbative expansion (30) from the asymptotics (34): see Fig. 2. Physically it means that in the case of large  $\tilde{\beta}$  the dispersion leads to signal spreading in time domain. It results in the amplitude decreasing and thereby effectively decreasing of the nonlinear term in the equation (1).

#### 4. Concluding remarks

We have derived the analytical expression for the mutual information  $I_{P[X]}$  of the channel modelled by the nonlinear Schrödinger equation with the additive Gaussian noise at large SNR. We used the path-integral approach to the calculation of the conditional probability density function based on Martin-Siggia-Rose formalism. In the leading order in  $1/\text{SNR}$  we obtained the general representation (in the case of arbitrary nonlinearity and dispersion) for the mutual information for the Gaussian input signal PDF. The correction to Shannon's result has the form of the averaged logarithm of the normalization factor  $\Lambda$  of the conditional PDF  $P[Y|X]$  expressed in terms of the path-integral. In the case of small nonlinearity (in expansion in  $\tilde{\gamma} = \gamma LP_{ave} \lesssim 1$ ) we obtained the first nonvanishing correction to Shannon's result. It is of order of  $\tilde{\gamma}^2$  and decreases for increasing dispersion parameter. Using Jensen's inequality we obtained the low bound for the mutual information in the case of arbitrary nonlinearity and dispersion.

#### Acknowledgments

The general expressions for the entropies and the mutual information have been obtained with the support of the Russian Science Foundation (RSF) (grant No. 16-11-10133). Part of the work (perturbative calculations, mutual information estimations) was supported by the Russian Foundation for Basic Research (RFBR), Grant No. 16-31-60031/15. A. V. Reznichenko thanks the President program (SP-2415.2015.2) for support.

#### References

- [1] Shannon C E 1949 *Communication in the presence of noise. Proc. Institute of Radio Engineers* vol 37 p 1
- [2] Martin P C, Siggia E D, and Rose H A 1973 *Phys. Rev. A* **8** 423
- [3] Terekhov I S, Vergeles S S, and Turitsyn S K 2014 *Phys. Rev. Lett.* **113** 230602
- [4] Terekhov I S, Reznichenko A V, and Turitsyn S K 2016 *Phys. Rev. E* **94** 042203
- [5] Itzykson C, Zuber J B 1980 *Quantum Field Theory* (New York: McGraw-Hill Book Company)
- [6] Terekhov I S, Reznichenko A V, Kharkov Ya A, and Turitsyn S K, *Preprint* arXiv: 1508.05774v2
- [7] Essiambre R -J *et al.* 2010 *J. of Lightwave Technol.* **28** 0733
- [8] Novikov S, Manakov S V, Pitaevskii L P, and Zakharov V E 1984 *Theory of Solitons: The Inverse Scattering Method (Monographs in Contemporary Mathematics)* (Rumford: Springer)