Supplementary Materials for the article: Optimal input signal distribution and per-sample mutual information for nondispersive nonlinear optical fiber channel in large SNR limit

I. S. Terekhov*†‡, A. V. Reznichenko*†, Ya. A. Kharkov†, S. K. Turitsyn‡
*Budker Institute of Nuclear Physics, Russian Academy of Sciences, Novosibirsk, 630090, Russia
†Novosibirsk State University, Novosibirsk, 630090, Russia
‡School of Physics, University of New South Wales, Sydney 2052, Australia
§Aston Institute of Photonics Technologies, Aston University, Aston Triangle, Birmingham, B4 7ET, UK

I. CONDITIONAL PROBABILITY DENSITY FUNCTION IN NON-DISPERSIVE CASE

In Ref.[1] we have shown that in the case \( \text{SNR} = \frac{P}{Q} \gg 1 \) the conditional probability can be written in the form:

\[
P[Y|X] \approx \exp \left\{ -\frac{S[\Psi_{cl}(z) \bigl(0\bigr)]}{Q} \right\} \int D\tilde{\psi} \exp \left\{ -\frac{S[\Psi_{cl}(z) + \tilde{\psi}(z) - S[\Psi_{cl}(z)]]}{Q} \right\},
\]

in non-dispersive model one has the effective action

\[
S[\psi] = \int_0^L dz \left| \partial_z \psi - i \gamma |\psi|^2 \psi \right|^2
\]

is associated with l.h.s. of the nonlinear Shrödinger equation

\[
\partial_z \psi(z) - i \gamma |\psi(z)|^2 \psi(z) = \eta(z),
\]

and Gaussian nature of the noise \( \eta(z) \):

\[
\langle \eta(z) \rangle_\eta = 0, \quad \langle \eta(z) \eta(z') \rangle_\eta = Q \delta(z - z').
\]

The measure \( D\tilde{\psi} \) in Eq. (1) is defined as

\[
D\tilde{\psi} = \lim_{\Delta \to 0} \left( \frac{1}{\Delta \pi Q} \right)^N \prod_{i=1}^{N-1} dRe \tilde{\psi}_i dIm \tilde{\psi}_i,
\]

with \( \tilde{\psi}_i = \tilde{\psi}(z_i) \) and \( \Delta = \frac{L}{N} \) being the grid space.

In Eq. (1) the function \( \Psi_{cl}(z) \) is the solution of the equation \( \delta S[\Psi_{cl}] = 0 \) which has the form

\[
\frac{d^2 \Psi_{cl}}{dz^2} - 4i \gamma |\Psi_{cl}|^2 \frac{d\Psi_{cl}}{dz} - 3 \gamma^2 |\Psi_{cl}|^4 \Psi_{cl} = 0,
\]

with the boundary condition \( \Psi_{cl}(0) = X = |X| \exp[i \phi(X)], \quad \Psi_{cl}(L) = Y = |Y| \exp[i \phi(Y)] \). It is easy to find the solution of Eq. (6) in the polar coordinate system: \( \Psi_{cl}(z) = \rho(\zeta) e^{i \theta(\zeta)}, \quad \zeta = z/L \). The solution depends on four real integration constants. We denote them as \( E, \tilde{\mu}, \zeta_0, \) and \( \theta_0 \). There are two different regimes of the solution: in the trigonometric regime one has \( E = \frac{k^2}{2} \geq 0 \), and in the hyperbolic regime \( E = -\frac{k^2}{2} \leq 0 \). For both cases instead of \( E \) we introduce the non-negative parameter \( k = \sqrt{2E} \).

- In the trigonometric case \( E = \frac{k^2}{2} \geq 0 \) we have the solution for \( \tilde{\mu} \geq k \geq 0 \):

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

\[
\theta(\zeta) = \frac{\tilde{\mu}}{2} (\zeta - \zeta_0) + \sqrt{\tilde{\mu}^2 - k^2} \frac{\sin[2k(\zeta - \zeta_0)]}{4k} + \arctan \left[ (\tilde{\mu} - \sqrt{\tilde{\mu}^2 - k^2}) \frac{\tan[k(\zeta - \zeta_0)]}{k} \right] + \theta_0.
\]
where integration constants $\hat{\mu}$, $k$ and $\zeta_0$ must be found from the boundary conditions: $|X|^2 = \rho^2(0)$, $|Y|^2 = \rho^2(1)$, and $\phi(Y) - \phi(X) = \theta(1) - \theta(0)$, and after that the last parameter $\theta_0 = \phi(X) - \theta(0)$. Then one can find the action

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

(8)

- In the hyperbolic case ($E = -\frac{k^2}{2} \leq 0$) we have the solution for $k \geq 0$ and arbitrary $\hat{\mu}$ in the following form

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

$$\theta(\zeta) = -\frac{\hat{\mu}}{2}(\zeta - \zeta_0) + \frac{\sqrt{\hat{\mu}^2 + k^2 \sinh[2k(\zeta - \zeta_0)]}}{4k} - \arctan\left[(\hat{\mu} + \sqrt{\hat{\mu}^2 + k^2})\tanh[k(\zeta - \zeta_0)]\right] + \theta_0,$$

where $\hat{\mu}$, $k$, $\zeta_0$, and $\theta_0$ are derived from the same procedure as in the trigonometric regime. The action reads

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

(10)

Note, there are two solutions of Eq. (6) with the constant $\rho(z) = \rho(0) \equiv \rho$ and obeying only the input boundary condition $\Psi_0(0) = X$. The first one

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

(11)

where $\mu = \gamma L \rho^2 = \gamma L |X|^2$, corresponds to the solution representation (4) with $k = 0$ and $\hat{\mu} = \mu$ or to the solution representation (9) with $k = 0$ and $\hat{\mu} = -\mu$. The function $\Psi_0(z)$ is the solution of the Eq. (1) with zero noise and with the input boundary condition. Furthermore, $\Psi_0(z)$ delivers the absolute minimum of the action $S[\Psi_0(z)] = 0$.

The second solution of Eq. (6) with the constant $\rho(z)$ is the trigonometric regime (7) case with $\hat{\mu} = k = 2\mu$:

$$\Psi_{\rho=const}(z) = \rho \exp\left\{3i\mu \frac{z}{L} + i\phi(X)\right\}, \quad \mu = \gamma L \rho^2 = \gamma L |X|^2.$$

(12)

To find the solution of Eq. (6) one should express the integration constant through the boundary condition. Instead, we exploit the fact that the noise power $QL$ much less than input signal power $P = |X|^2 \equiv \rho^2$. In other words, we will find the solution of Eq. (6) that is close to the $\Psi_0(z)$: it is the solution of Eq. (3) with zero noise which provides the absolute minimum of the action $S[\Psi_0(z)] = 0$. In that fashion we perform the substitution in Eq. (6):

$$\Psi_{cl}(z) = (\rho + \varkappa(z)) \exp\left\{i\mu \frac{z}{L} + i\phi(X)\right\},$$

(13)

where the function $\varkappa(z)$ is assumed to be small: $\varkappa(z) \ll \rho$ for all configurations of $\Psi_{cl}(z)$ providing $S[\Psi_{cl}(z)]/Q = O(1)$ when $QL$ tends to zero.

We present $\varkappa(z)$ in the perturbation theory decomposition in $1/\sqrt{\text{SNR}}$ parameter: $\varkappa(z) = \varkappa_1(z) + \varkappa_2(z) + \ldots$, where $\varkappa_1(z)$ is of $1/\sqrt{\text{SNR}}$ order and provides the leading order contribution, $\varkappa_2(z)$ is of $1/\text{SNR}$ order, and so on.

- The linearized equation for the function $\varkappa_1(z) = x_1(z) + iy_1(z)$ has the form

$$\frac{d^2 x_1}{dz^2} - 2i \frac{\mu}{L} \frac{dx_1}{dz} - 4 \frac{\mu^2}{L^2} \Re[\varkappa_1] = 0.$$

(14)

The boundary conditions $\Psi_{cl}(0) = X$ and $\Psi_{cl}(1) = Y \equiv \rho e^{i\phi(Y)}$ leads to

$$\varkappa_1(0) = 0,$$

$$\varkappa_1(1) = x_0 + iy_0 = \rho e^{i\phi(Y) - \phi(X) - \mu} - \rho.$$

(15)

The solution $\varkappa_1(z) = x_1(z) + iy_1(z)$ of the linearized boundary problem (14), (15) is polynomial

$$x_1(z) = \left(-\mu a_1(X,Y) \frac{z}{L} + a_2(X,Y)\right) \frac{z}{L},$$

$$y_1(z) = \left(-\frac{2}{3} \mu^2 a_1(X,Y) \frac{z^2}{L^2} + \mu a_2(X,Y) \frac{z}{L} + a_1(X,Y)\right) \frac{z}{L},$$

(16)

where coefficients $a_1(X,Y)$ and $a_2(X,Y)$ can be found from the boundary conditions (15) and have the form:

$$a_1(X,Y) = \frac{-\mu x_0 + y_0}{1 + \mu^2/3},$$

$$a_2(X,Y) = \frac{1 - 2\mu^2/3}{1 + \mu^2/3} x_0 + \mu y_0.$$
with \( x_0 = x_0(X, Y) \) and \( y_0 = y_0(X, Y) \) being determined from Eq. (15). In the leading order the action reads

\[
\frac{1}{Q} S\left[\Psi_0(z) + \chi_1(z)e^{i\Phi_0 + i\phi(x)}\right] = \frac{1}{Q} \int_0^L dz \left[ \left( \frac{\partial^2 \chi_1}{\partial x^2} - 2i\frac{\mu}{L} Re[\chi_1] \right)^2 + O\left( \frac{\chi_1^3}{\rho^3} \right) \right] = 
\]

\[
= (1 + 4\mu^2/3)a_0^2 - 2\mu a_1 a_2 + a_2^2 + O\left( \frac{1}{\sqrt{\text{SNR}}} \right) = \frac{(1 + 4\mu^2/3)x_0^2 - 2\mu x_0 y_0 + y_0^2 + O\left( \frac{1}{\sqrt{\text{SNR}}} \right)}{QL(1 + \mu^2/3)}. \tag{18}
\]

Let us proceed to the next-to-leading order correction to the \( P[Y|X] \). We should calculate the next approximation \( \chi_2(z) \) to the solution (13). Taking into account Eq. (14) we present the equation for \( \chi_2(z) \) in the form

\[
\frac{d^2 \chi_2}{dz^2} - 2i\frac{\mu}{L} \frac{d \chi_2}{dz} - 4\frac{\mu}{L} Re[\chi_1] = \frac{2\mu}{\rho L} \left[ 2Re[\chi_1]\left( 3\frac{\mu}{L} Re[\chi_1] + \frac{\mu}{L} \chi_2 + 2i\frac{d\chi_1}{dz} \right) + \frac{\mu}{L} |\chi_1|^2 \right]. \tag{19}
\]

where the boundary conditions for \( \chi_2(z) \) read \( \chi_2(0) = \chi_2(L) = 0 \). The solution \( \chi_2(z) = x_2(z) + iy_2(z) \) of the Eq. (19) is polynomial in \( z \) and quadratic in \( x_0 \) and \( y_0 \):

\[
x_2(z) = -\frac{\mu/\rho}{270(1 + \mu^2/3)^3} \left( 1 - \frac{z}{L} \right) \frac{z}{L} \left\{ \begin{array}{c}
\mu (2\mu^4 - 15\mu^2 + 585) x_0^2 + 2 (13\mu^2 (\mu^2 + 15) - 180) x_0 y_0 + \mu (2\mu^2 + 15) (5\mu^2 - 9) y_0^2 - \\
5 (\mu^2 + 3) \frac{z}{L} \left( \mu (\mu^2 - 15) x_0^2 - 4 (\mu^2 - 6) x_0 y_0 + \mu (\mu^2 + 9) y_0^2 \right) + \\
5\mu (\mu^2 + 3) \frac{z^2}{L^2} \left( 3 (\mu^2 - 3) x_0^2 + 36\mu x_0 y_0 - (\mu^2 - 15) y_0^2 \right) + \\
20\mu^2 (\mu^2 + 3) \frac{z^3}{L^3} \left( y_0 - \mu x_0 \right) \left( 2\mu y_0 - (\mu^2 - 3) x_0 \right) - \\
20\mu^2 (\mu^2 + 3) \frac{z^4}{L^4} (y_0 - \mu x_0)^2 \end{array} \right\}. \tag{20}
\]

\[
y_2(z) = -\frac{\mu/\rho}{270(1 + \mu^2/3)^3} \left( 1 - \frac{z}{L} \right) \frac{z}{L} \left\{ \begin{array}{c}
(7\mu^4 - 75\mu^2 + 360) x_0^2 + 6 \mu (\mu^2 + 75) x_0 y_0 + 3\mu^2 (5\mu^2 + 39) y_0^2 + \\
2\frac{z}{L} \left( \mu^6 - 4\mu^4 + 255\mu^2 + 180 \right) x_0^2 + \mu (\mu^2 + 15) (13\mu^2 + 3) x_0 y_0 + \mu^2 (5\mu^4 + 36\mu^2 - 9) y_0^2 - \\
14\mu (\mu^2 + 3) \frac{z^2}{L^2} (y_0 - \mu x_0) \left( (15 - 4\mu^2) x_0 + 9\mu y_0 \right) + \\
84\mu^2 (\mu^2 + 3) \frac{z^3}{L^3} (y_0 - \mu x_0)^2 \end{array} \right\}. \tag{21}
\]

In the leading (see Eq. (18)) and next-to-leading order the action reads

\[
\frac{1}{Q} S[\Psi_{cl}(z)] = \frac{(1 + 4\mu^2/3)x_0^2 - 2\mu x_0 y_0 + y_0^2}{QL(1 + \mu^2/3)} + \frac{\mu/\rho}{135QL(1 + \mu^2/3)} \left\{ \mu (4\mu^4 + 15\mu^2 + 225) x_0^4 + \\
(23\mu^4 + 255\mu^2 - 90) x_0 y_0^2 + \mu (20\mu^4 + 117\mu^2 - 45) x_0 y_0^2 - 3 (5\mu^4 + 33\mu^2 + 30) y_0^3 \right\} + O\left( \frac{1}{\sqrt{\text{SNR}}} \right). \tag{22}
\]

To calculate the conditional probability density \( P[Y|X] \) one should find the pre-exponent path-integral (so-called quantum corrections over the classical solution \( \Psi_{cl}(z) \)) in (11) in the leading 1/\( \text{SNR} \) order:

\[
I_{QC}[\Psi_{cl}(z)] = \int_{\Psi(0)=0}^{\Psi(L)=0} \mathcal{D}\Psi \exp \left\{ -\frac{S[\Psi_{cl}(z)] + \bar{\Psi}(z) - S[\Psi_{cl}(z)]}{Q} \right\} \approx \int_{u(L)=0}^{u(0)=0} \mathcal{D}u \exp \left\{ -\frac{1}{Q} \int_0^L dz \left[ \partial_z u - 2i\frac{\mu}{L} (u + \bar{u}) \right]^2 \right\} = 
\]

\[
\lim_{N \to \infty} \left( \frac{N}{\pi QL} \right) \prod_{i=1}^{N-1} dX_i dY_i \exp \left\{ -\frac{N}{QL} \sum_{i=0}^{N-1} \left\{ (X_{i+1} - X_i)^2 + (Y_{i+1} - Y_i - 2\frac{\mu}{N} X_i)^2 \right\} \right\}, \tag{23}
\]
where use the measure and discretisation scheme \( u(z_i) = X_i + Y_i, \ z_i = \Delta \frac{z_i}{N}, \ \Delta = \frac{1}{N} \) and \( X_0 = X_{N+1} = Y_0 = Y_{N+1} = 0 \).

The sequential integration over \( Y_{N-1}, Y_{N-2}, \ldots, Y_1 \) is trivial:

\[
\int dY \exp \left\{ \frac{- (A - Y)^2}{2\tau_1} - \frac{(Y - B)^2}{2\tau_2} \right\} = \left( \frac{2\pi}{\tau_1 \tau_2} \right)^{1/2} \exp \left\{ \frac{- (A - B)^2}{2(\tau_1 + \tau_2)} \right\}.
\]

It leads to the remainder integral (over \( X_i, i = 1, \ldots, N - 1 \)) of the form

\[
\lim_{N \to \infty} \left( \frac{N}{\pi QL} \right)^N \frac{(\pi QL/N)^{N-1}}{\sqrt{N}} \int_{-\infty}^{\infty} \prod_{i=1}^{N-1} dX_i \exp \left\{ - N \frac{Q}{QL} \sum_{i,j=1}^{N-1} X_i M_{i,j}(\alpha) X_j \right\},
\]

where we denote \( \alpha = \frac{4}{N} \left( \frac{\mu}{N} \right)^2 \), and \( (N - 1) \) by \( (N - 1) \) matrix \( M(\alpha) \) has the following elements: \( M_{i,i} = 2 + \alpha, \ M_{i,i\pm 1} = -1 + \alpha, \ i = 1, \ldots, N - 1 \), and with others elements are \( M_{i,j} = \alpha, j \neq i, j \neq i \pm 1 \). It is straightforward to calculate the determinant of \( M(\alpha) \) and to perform the Gaussian integration over \( X_i \)

\[
\det[M(\alpha)] = N + \alpha \frac{N^2(N^2 - 1)}{12},
\]

\[
I_{QC}[\Psi_0(z)] = \frac{1}{\pi QL \sqrt{1 + \mu^2/3}}.
\]

Finally, we have

\[
P[Y|X] = \exp \left\{ \frac{- (1 + 4\mu^2/3)x_0^2 - 2\mu x_0 y_0 + y_0^2}{QL(1 + \mu^2/3)} \right\} \left( 1 - \frac{\mu/\rho}{135QL (1 + \mu^2/3)^3} \left( \mu (4\mu^4 + 15\mu^2 + 225) x_0^3 + (23\mu^4 + 255\mu^2 - 90) x_0^2 y_0 + \mu (20\mu^4 + 117\mu^2 - 45) x_0 y_0^2 - 3 (5\mu^4 + 33\mu^2 + 30) y_0^3 \right) + \mathcal{O} \left( \frac{1}{\text{SNR}} \right) \right) .
\]

Note that the correction (the second term in big parentheses in Eq. (28)) does not change the leading order contribution \( (\pi QL \sqrt{1 + \mu^2/3})^{-1} \) to the path integral expression.

Now it is easy to see, that the normalization condition

\[
\int DY P[Y|X] = 1
\]

is fulfilled: indeed, \( DY \equiv dReYdImY = dx_0dy_0 \) when \( \rho \) is fixed, and the normalization integral becomes Gaussian.

REFERENCES

[1] I. S. Terekhov, S. S. Vergeles, and S. K. Turitsyn, Phys. Rev. Lett. 113, 230602 (2014).
Optimal input signal distribution and per-sample mutual information for nondispersive nonlinear optical fiber channel in large SNR limit

I. S. Terekhov*†‡, A. V. Reznichenko*†, Ya. A. Kharkov†, S. K. Turitsyn§†‡

*Budker Institute of Nuclear Physics, Russian Academy of Sciences, Novosibirsk, 630090, Russia
†Novosibirsk State University, Novosibirsk, 630090, Russia
‡School of Physics, University of New South Wales, Sydney 2052, Australia
§Aston Institute of Photonics Technologies, Aston University, Aston Triangle, Birmingham, B4 7ET, UK

Abstract—For a model nondispersive nonlinear optical fiber channel the analytical expressions for the conditional probability, conditional entropy, output signal entropy, and per-sample mutual information in the large SNR (signal-to-noise ratio) limit were derived. The conditional probability was calculated using Feynman path-integral technique in the leading and next-to-leading order in 1/SNR. In the leading order in 1/SNR the input signal distribution maximizing the mutual information in the intermediate power region was found. In this region the mutual information for optimal distribution was calculated in large SNR limit.

I. INTRODUCTION.

The channel capacity $C$ introduced by Shannon in his seminal work [1] is related to the maximum amount of information that can be reliably transmitted over noisy communication channel. Therefore, this quantity is of great practical importance. Shannon calculated the capacity of the linear channel with the additive white Gaussian noise (AWGN) and found its famous logarithmic dependence on the signal power:

$$ C \propto \log_2 (1 + \text{SNR}) , $$

where $\text{SNR} = P/N$ is the signal-to-noise power ratio, $P$ is the signal power, and $N$ is the noise power. This, in particular, means that when noise power $N$ is fixed to increase the capacity one has to increase the signal power $P$. Recently, it was a growing interest in studying properties of nonlinear communication channels (see e.g. [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17] and numerous references therein) - channels, where transmission is affected and changed by the signal power, especially, in case of large SNR [2]. Analysis of the information capacity of nonlinear channels is technically challenging and new techniques and methods are highly desirable to advance these studies. In this work we present new method for computation of the mutual information (and, thus, a lower bound on the capacity) in nonlinear channels and new analytical results for the nondispersive nonlinear fiber channel [10], [11], [18] in the limit of large SNR.

The channel capacity $C$ is the maximum of the mutual information $I_{P_X[X]}$ with respect to probability density function $P_X[X]$ of an input signal $X$:

$$ C = \max_{P_X[X]} I_{P_X[X]} , $$

where maximum value of $I_{P_X[X]}$ should be found at the given average power: $P = \int |X|^2 P_X[X]$. The mutual information of a memoryless channel is defined in terms of the output signal entropy $H[Y]$ and conditional entropy $H[Y|X]$:

$$ I_{P_X[X]} = H[Y] - H[Y|X] , $$

where

$$ H[Y|X] = - \int D XP_X[X] P[Y|X] \log P[Y|X] , $$

$$ H[Y] = - \int D Y P_{out}[Y] \log P_{out}[Y] , $$

$$ P_{out}[Y] = \int D XP_X[X] P[Y|X] , $$

where $P[Y|X]$ is conditional probability density function to have output signal $Y$ when the input signal is $X$, the measure $DY$ is defined as $\int D Y P[Y|X] = 1$, and $DX$ is defined as $\int D XP_X[X] = 1$. The capacity $C$, as defined by (2)-(6), is measured in units $(\log 2)^{-1}$ bit per symbol (also known as hartley per symbol). The input and output signals are the functions of time in certain bandwidth. In general, a sampling of the temporal signal should be introduced to define discrete-time memoryless channel, however, here we consider only per-sample quantities.

For the nonlinear fiber channels with zero dispersion and additive white Gaussian noise the analytical expression for the conditional probability density function (PDF) was obtained in the form of infinite series [10], [18] within Martin-Siggia-Rose formalism based on the quantum field theory methods [19]. Furthermore, the lower bound at large SNR for nondispersive optical fiber channel was analytically derived in Ref. [10]:

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

where $\gamma_E \approx 0.5772$ is the Euler constant. The pre-logarithmic factor $1/2$ in Eq. (7) arises as a result of the fact that in the high power regime, when the signal power $P \gtrsim (N^2 L^2)^{-1}$, the signal-dependent phase noise takes over the entire phase...
interval and, as a result, the phase does not transfer information, see Ref. [11]. Here $\gamma$ is the Kerr nonlinearity coefficient, $L$ is a fiber link length, see below. In [11] the capacity estimates were presented in the intermediate power interval $N \ll P \ll 6\gamma^2 / (N\gamma^2 L^2)^{-1}$ as well. For that range of power $P$ the following estimate on the lower bound was derived [11]:

$$C \geq - \log(\gamma NL) + \frac{\gamma E - 1 + \log(3\pi)}{2} + O\left(\frac{1}{\sqrt{\text{SNR}}}\right),$$

(8)

where instead of $O\left(\frac{1}{\sqrt{\text{SNR}}}\right)$ the authors presented the explicit functions of the parameter SNR which decrease at large SNR; see eq. (40) in [11]. However, the authors of [11] did not take into account the $\sqrt{\text{SNR}}$ corrections in the output signal entropy $H(Y)$, therefore, keeping these functions in the capacity inequality is beyond the calculation accuracy. To obtain the estimate in the half-Gaussian distribution $P_X^{(1)}(X)$ of the input signal was exploited as the capacity-approaching distribution:

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

(9)

where $P$ is the average power of the input signal:

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

(10)

In the present paper we show that the optimal input signal distribution function for nondispersive nonlinear optical fiber channel differs from half-Gaussian one [3].

In [11] method to describe the nonlinear optical fiber channel with nonzero dispersion in the large SNR limit was introduced. Here we illustrate this general approach in application to more simple nondispersive nonlinear optical fiber channel considered in [10], [11], [18]. Since the channel is dispersionless, the temporal signal waveform is not changing during propagation (note, though, that the signal bandwidth can grow due to fiber nonlinearity and signal modulation). Therefore, instead of consideration of the evolution of $\psi(z,t)$ we can consider a set of parallel independent scalar channels (per-sample channels [11]) governed by the following model:

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

(11)

where $\psi(z)$ is the signal function that is assumed to obey the boundary conditions $\psi(0) = X$, $\psi(L) = Y$. The white noise $\eta(z)$ has the zero mean and the correlation function $\langle \eta(z)\eta(z') \rangle_z = Q\delta(z-z')$, so that $\text{SNR} = P/QL$, where $P$ and $QL$ are the per-sample signal and noise power, respectively. The connection between the differential model [11] and conventional information-theoretic presentation in the form of an explicit input-output probabilistic model and appropriate sampling has been discussed in details in [10], [11], [18]. Therefore, here we will deal with the per-sample analysis and refer for all the specific information-theoretic aspects to a very detailed paper [11].

For this per-sample channel we calculate conditional probability (in order to illustrate how our method works), conditional entropy [4], output signal entropy [5], and the mutual information [3]. By varying the mutual information we find the optimal input signal distribution $P_X[X]$ maximizing the mutual information in the leading order in $1/\text{SNR}$. The paper is organized in the following way. In Section 2 we develop the quasi-classical method for the calculation of the conditional PDF $P[Y|X]$ for arbitrary nonlinearity in the leading and next-to-leading order in $1/\text{SNR}$. We find the simple representation for $P[Y|X]$ in this limit. This allows us to calculate output signal distribution $P_{out}[Y]$. The optimal signal distribution $P_{out}^{(2)}[X]$ is found in Section 3. Section 4 is focused on the calculation and the comparison of the mutual informations for the various input signal distributions. We demonstrate that there is a range of power $P$ where the mutual information $I_{P_{out}^{(2)}|X}$ calculated for Gaussian distribution $P_{out}^{(2)}[X]$ is closer to $I_{P_{out}^{(2)}|X}$, whereas at large enough power $P$ the mutual information $I_{P_{out}^{(2)}|X}$ is closer to $I_{P_{out}^{(2)}|X}$ than the mutual information $I_{P_{out}^{(2)}|X}$. We draw our conclusions in Section 5.

II. CALCULATION OF THE CONDITIONAL PDF $P[Y|X]$ AND OUTPUT SIGNAL PDF $P_{out}[Y]$ IN THE LARGE SNR LIMIT

A. "Quasiclassical" method for the conditional PDF $P[Y|X]$ calculation

The conditional probability written via the path-integral form [10] in retarded discretization scheme [19], [20]

$$P[Y|X] = \int \mathcal{D}\psi \exp\left\{ - \frac{S[\psi]}{Q} \right\},$$

(12)

can be reduced to the quasi-classical form, see Ref. [17]:

$$P[Y|X] = e^{-\frac{S[\Psi_{cl}(z)]}{Q}} \int \mathcal{D}\psi e^{-\frac{S[\Psi_{cl}(z)+\psi(z)]-S[\Psi_{cl}(z)]}{Q}},$$

(13)

where the effective action $S[\psi] = \int_0^L dz [\partial_z\psi - i\gamma|\psi|^2\psi]^2$, and the function $\Psi_{cl}(z)$ is the "classical" solution of the equation $\delta S[\Psi_{cl}] = 0$ which has the form

$$\frac{d^2\Psi_{cl}}{dz^2} - 4i\gamma |\Psi_{cl}|^2 \frac{d\Psi_{cl}}{dz} - 3\gamma^2|\Psi_{cl}|^4 \Psi_{cl} = 0,$$

(14)

with the boundary condition $\Psi_{cl}(0) = X$, $\Psi_{cl}(L) = Y$. The exact solution $\Psi_{cl}(z)$ can be calculated analytically, see Ref. [21]. In order to find $P[Y|X]$ one should calculate the exponent $e^{-\frac{S[\Psi_{cl}(z)]}{Q}}$ and the path integral in Eq. (13).

First, we evaluate the exponent in the leading order in $1/\text{SNR}$. Although the general solution $\Psi_{cl}(z)$ of (14) can be found analytically (see (21)), there is no closed form for the solution of the boundary problem. Instead, we find such solution in the leading $1/\text{SNR}$ order, linearizing Eq. (14) in the vicinity of the solution $\Psi_0(z)$ of the channel equation [11] with zero noise. The function $\Psi_0(z)$ reads

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

(15)

where $\mu = \gamma L\rho^2 = \gamma L|X|^2$. Note that this solution satisfies only the input boundary condition $\Psi_0(0) = X = \rho e^{i\phi(X)}$, and
it is the solution of Eq. (14) as well. Therefore, we look for the solution of Eq. (14) in the following form

$$\Psi_{cl}(z) = (\rho + \omega(z)) \exp \left\{ i\mu \frac{z}{L} + i\phi(X) \right\}, \quad (16)$$

where the function $\omega(z)$ is assumed to be small: $|\omega(z)| \ll \rho$. In general case, the ratio $|\omega(z)|/\rho$ is not necessarily small and it depends on the output boundary condition $\omega(L)$. However, the configurations of $\omega(z)$ at which $\Psi_{cl}(z)$ significantly deviates from $\Psi_0(z)$ is $\omega(L)$ are statistically irrelevant. Indeed, the exponent $e^{-\frac{\omega(z)}{\rho}}$ and, as a result, the conditional probability $P[Y | X]$ vanish if the type of $\omega(z)$ is greater than $\sqrt{QL}$. Note, that the expansion $S[\Psi_0(z) + \delta\Psi(z)] \propto \omega^2(z)$ starts from the quadratic term at small $\omega(z)$, since the action achieves extremum (the absolute minimum $S[\Psi_0(z)] = 0$) on the solution $\Psi_0(z)$.

The linearized in $|\omega(z)/\rho| \sim 1/\sqrt{SNR}$ and exact in non-linearity parameter $\mu$ equation for the function $\omega(z)$ reads

$$\frac{d^2\omega}{dz^2} - 2\frac{\mu}{L} \frac{d\omega}{dz} - 4\frac{\mu^2}{L^2} \text{Re}[\omega] = 0. \quad (17)$$

The boundary conditions for the function $\omega(z)$ read

$$\omega(0) = 0, \quad \omega(L) = Ye^{-i\phi(X)}-i\mu - \rho \equiv x_0 + iy_0. \quad (18)$$

The solution of the linearized boundary problem (17), (18) is polynomial (see details in [21]). After substitution of the solution $\omega(z)$ of the Eq. (17) in the action we get

$$\frac{1}{Q} \int_0^L dz \left| \frac{\partial}{\partial \omega} \omega - 2\frac{\mu}{L} \text{Re}[\omega] \right|^2 = \frac{(1 + 4\mu^2/3)x_0^2 - 2\mu x_0 y_0 + y_0^2}{QL(1 + \mu^2/3)}. \quad (19)$$

Note that here we retain only the main, quadratic in $\omega$ terms. However, it is straightforward to calculate the next correction to the action [19] in $1/\sqrt{SNR}$ parameter, see detail in [21].

To calculate the path-integral in Eq. (13) in the leading $1/\text{SNR}$ order we retain only the quadratic in $\psi$ expression, because higher order in $\psi$ terms give corrections to the path-integral suppressed in the parameter $QL$ and can be neglected in the large SNR limit. Moreover, as soon as we calculate the path-integral in the leading order in $Q$, we can substitute $\Psi_0(z)$ for $\Psi_{cl}(z)$ in the action difference $S[\Psi_{cl}(z) + \psi(z)] - S[\Psi_{cl}(z)]$, see details in [21]. The result for $P[Y | X]$ calculation in the leading and next-to-leading in $1/\sqrt{SNR}$ order is presented in [21], here we show $P[Y | X]$ only in the leading order:

$$P[Y | X] = \exp \left\{ \frac{-(1 + 4\mu^2/3)x_0^2 - 2\mu x_0 y_0 + y_0^2}{QL(1 + \mu^2/3)} \right\}. \quad (20)$$

where $x_0$ and $y_0$ are the functions of $X$ and $Y$ defined in (13). One can see that the normalization condition $\int D\psi P[Y | X] = 1$ is fulfilled. Also one can check that the distribution (20) obeys the following important property

$$\lim_{Q \to 0} P[Y | X] = \delta \left( Y - \Psi_0(L) \right). \quad (21)$$

The expression (21) is nothing else, but the deterministic limit of $P[Y | X]$ for the vanishing noise.

B. Output signal PDF $P_{out}[Y]$

Now we consider the distribution of the output signal $P_{out}[Y]$. In the leading order in SNR the distribution $P_{out}[Y]$ can be easily calculated using property (21). The substitution of Eq. (21) to Eq. (6) leads to following expression:

$$P_{out}[Y] = P_X \left[ Ye^{-i\phi Y/\sqrt{L^2}} \right], \quad Q \to 0. \quad (22)$$

For the case of distribution $P_X[X]$ which depends only on $|X|$ we have $P_{out}[Y] = P_X[|Y|]$. For such type of distribution $P_X[X]$ we can calculate corrections to (22) in the parameter $QL$ in any order in $QL$. A reader, not interested in the details of the calculation of the corrections can proceed to Section III.

Let us restrict our consideration below in the Section by the distributions $P_X[X]$ depending only on $|X|$. We can use the $P[Y | X]$ found in Ref. [10], see Eqs. (11)–(13) therein. In this case $P_{out}[Y]$ is a function of $|Y| = \rho' = \rho e^{-i\phi}$

$$P_{out}[\rho'] = \frac{2}{QL} \int_0^\infty dp \rho'^2 \left( \frac{2p}{QL} \right) P_X[\rho], \quad (23)$$

where $I_0(z)$ is the modified Bessel function of the first kind. Using the representation we can obtain the simple relation for $P_{out}[\rho']$ calculation in perturbation theory in $QL$. To this end we perform the zero order Hankel transformation:

$$\hat{P}_{out}[k] = \int_0^\infty dp J_0(k \rho) P_X[\rho]. \quad (24)$$

of both sides of Eq. (23), then we use the standard integral [22] with Bessel and modified Bessel functions

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

and arrive at a simple relation on the Hankel images

$$\hat{P}_{out}[k] = e^{-k^2 \frac{QL}{16}} \hat{P}[k]. \quad (25)$$

Performing the inverse Hankel transformation

$$P_X[\rho] = \int_0^\infty dk k J_0(k \rho) \hat{P}[k], \quad (26)$$

we obtain

$$P_{out}[\rho] = e^{\frac{QL}{4} \Delta_\rho} P_X[\rho], \quad (27)$$

where $\Delta_\rho = \frac{\partial^2}{\partial \rho^2} + \frac{1}{\rho} \frac{\partial}{\partial \rho}$ is the two-dimensional radial Laplace operator. From the relation (27) the problem of finding $(QL)^n$ corrections to $P_{out}[\rho]$ reduces to exponent expansion and straightforward calculations of the action of the differential operator $\Delta_\rho^n$ on $P_X[\rho]$. 

Let us consider the widely used example of the modified Gaussian distribution
\[ P_X^{(\beta)}[\rho] = \frac{\exp\left\{ -\beta \rho^2 / (2P) \right\}}{\pi^{(\beta/2)} (2P/\beta)^{1/2}}. \] (28)
For \( \beta > 0 \) the distribution \( P_X^{(\beta)}[\rho] \) is normalized to unity \((2\pi \int_0^\infty d\rho \rho P_X^{(\beta)}[\rho] = 1)\) and has the average power \( P \)
\((2\pi \int_0^\infty d\rho \rho^2 P_X^{(\beta)}[\rho] = P)\). The distribution \( P_X^{(\beta)}[X] \) generalizes the half-Gaussian distribution for \( \beta = 1 \) and the
Gaussian one for \( \beta = 2 \):
\[ P_X^{(2)}[X] = \frac{1}{\pi P} e^{-[X]^2/P}. \] (29)
The output PDF is computed by the standard integration for the Gaussian case and to \( e^z \) for the Gaussian case and to \( e^z/2I_0(z/2) \) for the half-Gaussian case:
\[ P_{out}^{(2)}[Y] = \frac{1}{\pi(P + QL)} \exp \left\{ -\frac{[Y]^2}{P + QL} \right\}, \] (31)
\[ P_{out}^{(1)}[Y] = \frac{\exp[-[Y]^2/(2P + QL)]}{\pi [Y]/\sqrt{\pi(2P + QL)}} \times \frac{\pi[Y]^2}{QL} e^{-\frac{[Y]^2}{2QL(2P + QL)}} I_0 \left( \frac{[Y]^2}{QL(2P + QL)} \right). \] (32)
Note that the result for \( P_{out}^{(1)}[Y] \) in Ref. \( \text{[11]} \) (see Eq. \( \text{[38]} \) therein) for half-Gaussian distribution is incorrect. At small \( QL \) the second line of our Eq. \( \text{[32]} \) tends to unity approaching our asymptotic expression Eq. \( \text{[22]} \) and the asymptotic given in Ref. \( \text{[11]} \).

We note again that in the limit \( Q \to 0 \) in our example \( \text{[30]} \) \( P_{out}[Y] = P_X[H[Y]] \), this is nothing more, but the demonstration of the evident general property of \( P[Y|X] \) for arbitrary PDF \( P_X[X] \) resulting from \( \text{[21]} \), see also Eq. \( \text{[22]} \).

III. OPTIMAL INPUT DISTRIBUTION AT LARGE SNR

Next we derive the optimal input signal distribution in the large SNR limit. Let us consider the mutual information \( \text{[3]} \). Firstly, using Eq. \( \text{[22]} \) in the output entropy \( H[Y] \), see Eq. \( \text{[5]} \), we can substitute \( P_X[Y] \) instead of \( P_{out}[Y] \) in the leading 1/SNR order:
\[ H[Y] = - \int_0^{2\pi} d\phi (Y) \int_0^\infty d\rho \rho P_X \left[ \rho e^{i\phi(Y)} \right] \times \log P_X \left[ \rho e^{i\phi(Y)} \right]. \] (33)
Secondly, we insert \( P[Y|X] \) normalized to unity and having the fixed average power \( P \) one should solve the variational problem with the functional \( J[P_X] \)
\[ J[P_X] = H[Y] - H[Y|X] - \lambda_1 \left( \int DX P_X[X] - 1 \right) - \lambda_2 \left( \int DXP_X[X]|X|^2 - P \right), \] (35)
where \( \lambda_{1,2} \) are the Lagrangian coefficients. Taking \( H[Y] \) and \( H[Y|X] \) from Eqs. \( \text{[33]} \) and \( \text{[34]} \) and performing variation over \( P_X[X] \) and \( \lambda_{1,2} \) we obtain the angular symmetric solution \( P_{opt}[X] \) referred to as the “optimal” distribution:
\[ P_{opt}[X] = N_0(P) \exp \left\{ -\lambda_0(P)|X|^2 \right\} \sqrt{1 + \gamma L^2 |X|^4/3}, \] (36)
where \( N_0[P] \) and \( \lambda_0[P] \) are the functions of the average power \( P \) and nonlinearity and are determined from the conditions
\[ \int DX P_{opt}[X] = 2\pi N_0(P) \int_0^{\infty} dp \rho e^{-\lambda_0(P)\rho^2} / \sqrt{1 + \gamma L^2 \rho^2/3} = 1, \] (37)
\[ \int DX P_{opt}[X]|X|^2 = 2\pi N_0(P) \int_0^{\infty} dp \rho e^{-\lambda_0(P)\rho^2} / \sqrt{1 + \gamma L^2 \rho^2/3} = P. \] (38)
In a parametric form this dependance reads
\[ \lambda_0(P) = \frac{\gamma L}{\sqrt{3}} \alpha, \quad N_0(P) = \frac{\gamma L}{\pi \sqrt{3} G(\alpha)}, \] (39)
here \( G(\alpha) = \int_0^{\infty} dz e^{-\alpha z} / \sqrt{1+z^2} \) is the Neumann and Struve functions of zero order correspondingly. And parameter \( \alpha(P) > 0 \) emerges as the real solution of the nonlinear equation \( \frac{d}{dP} \log G(\alpha) = -\gamma L P \sqrt{3} \), which comes from Eqs. \( \text{[37]} \) and \( \text{[38]} \). For sufficiently large values of the power \( P \), such that \( \gamma P L \gg 1 \), and \( \log(\gamma PL) \gg 1 \) we can simplify \( \text{[39]} \) by the asymptotical expansion:
\[ \lambda_0(P) \approx 1 - \log \log(C \gamma / \log(C \gamma)), \quad N_0(P) \approx \frac{\gamma}{\pi} \lambda_0(P), \] (40)
where \( C = 2 e^{-\gamma K} \) and \( \gamma = \gamma L P / \sqrt{3} \). At small \( P \), the parameter \( \gamma \ll 1 \), the solution of the Eqs. \( \text{[37]} \) and \( \text{[38]} \) has the form:
\[ \lambda_0(P) = \frac{1}{P} (1 - 2\gamma^2), \quad N_0(P) = \frac{1}{\pi P} (1 - \gamma^2). \] (41)
Now we are ready to calculate the mutual information.

IV. THE MUTUAL INFORMATION

In this Section we present the entropies and the mutual information for $P_\beta[X]$ and $P_{\text{opt}}[X]$ and compare our new results with the already known ones.

In the leading $1/\text{SNR}$ order the mutual information corresponding to the solution (36) reads as

$$I_{P_{\text{opt}}[X]} = P\lambda_0(P) - \log N_0(P) - \log(\pi e Q L),$$

(42)

and it gives the capacity estimation in the wide range of the average power $P$:

$$QL \ll P \ll (QL^3\gamma^2)^{-1}.$$  

(43)

The mutual information (42) is depicted by solid line in Fig. 1 as a function of power $P$ for the following parameters: $Q = 1.5 \times 10^{-7} \, \text{mW km}^{-1}$, $\gamma = 10^{-3} \, \text{mW}^{-1} \text{km}^{-1}$, $L = 1000 \, \text{km}$. There is no simple analytical form for the $N_0(P)$ and $\lambda_0(P)$, but for large and small values of the parameter $\tilde{\gamma}$ we can use solutions Eqs. (40) and (41), respectively. At small $\tilde{\gamma}$ we obtain

$$I_{P_{\text{opt}}[X]} = \log(\text{SNR}) - \tilde{\gamma}^2 + O(\tilde{\gamma}^3) + O\left(\frac{1}{\sqrt{\text{SNR}}}\right),$$

(44)

that is nothing else, but the expression for the Shannon capacity of the linear AWGN channel (1) with first nonlinear correction. Note that the derived expressions (43) and (44) have the correct limit when the parameter $\gamma$ tends to zero (in contrast to the Eq. (35) in Ref. (11)). In the second power sub-interval $Q L^{-1} \ll P \ll (QL^3\gamma^2)^{-1}$, using (40) one can see that the mutual information demonstrates very slow (loglog) increase with $P$

$$I_{P_{\text{opt}}[X]} = \log\left(\frac{1}{QL^2\gamma}\right) - 1 + \frac{\log 3}{2} + \log \left(\frac{C_0 \gamma LP}{\sqrt{3}}\right) + O\left(1/\log(\gamma LP)\right) + O\left(1/\text{SNR}\right),$$

(45)

rather than the constant behavior of the mutual information for Gaussian-like distributions of an input signal (see formulae (49) and (50) below).

In the rest of the present Section we perform analysis of the mutual information, corresponding to the generalized Gaussian distribution $P_X^{(\beta)}[X]$, Eq. (28).

In the rest of the present Section we perform analysis of the mutual information for the distribution $P_X^{(\beta)}[X]$, see (28), generalizing half-Gaussian distribution (9) (see, for example Ref. (11)) and the Gaussian input PDF (29). In the leading order in $1/\text{SNR}$ from (33) we get

$$H_\beta[Y] = \log \left(\frac{P \pi^{2\beta}/\beta}{\Gamma(\beta/2)}\right) + \frac{\beta - 2}{2} \psi\left(\frac{\beta}{2}\right),$$

(46)

where $\psi(z)$ is digamma function $\psi(z) = \Gamma'(z)/\Gamma(z)$, at that, $\psi(1) = -\gamma_E$, $\psi(1/2) = -\gamma_E - 2 \log(2)$. The substitution of Eq. (9) to Eq. (33) gives 1/\text{SNR} accuracy

$$H_\beta[Y][X] = \log \left(\pi e Q L\right) + \frac{1}{2\Gamma\left(\frac{\beta}{2}\right)} \int_0^\infty d\tau e^{-\tau^{\beta/2}-1} \log \left(1 + \frac{4\gamma^2 \tau^2}{\beta^2}\right)$$

(47)

whith at least $O(1/\text{SNR})$ accuracy. The integral in the Eq. (47) can be calculated analytically using Ref. (22), however, the result of integration is cumbersome, therefore we do not present it here. One can easily obtain the mutual information $I_{P_X^{(\beta)}[X]}$ subtracting Eq. (47) from Eq. (46).

$$I_{P_X^{(\beta)}[X]} = \log \text{SNR} + \log \left(\frac{2\Gamma(\beta/2)}{\beta}\right) - \log \left(\pi e Q L\right) + \frac{1}{2\Gamma\left(\frac{\beta}{2}\right)} \int_0^\infty d\tau e^{-\tau^{\beta/2}-1} \log \left(1 + \frac{4\gamma^2 \tau^2}{\beta^2}\right) +$$

$$+ \frac{\beta - 2}{2} \left(1 - \psi\left(\frac{\beta}{2}\right)\right).$$

(48)

The mutual information is depicted in Fig. 1 for the Gaussian distribution by blue dashed line, and the half-Gaussian one by red dashed dotted line. One can see, that at small $P$ the mutual information for the Gaussian distribution is greater than that for the half-Gaussian one, whereas at $P > 11 \, \text{mW}$ the mutual information is greater for the half-Gaussian distribution. Note that as it should be $I_{P_{\text{opt}}[X]}$ is greater than $I_{P_X^{(\beta)}[X]}$ for all values of $P$. At $\tilde{\gamma} \gg 1$ the mutual information $I_{P_{\beta}}[X]$ takes the form

$$I_{P_X^{(\beta)}[X]} = \log \left(\frac{1}{QL^2\gamma}\right) - \frac{2 - \beta}{2} \log 3 - \frac{\beta}{2} \psi\left(\frac{\beta}{2}\right) +$$

$$+ \log(\Gamma(\beta/2)) + O\left(\frac{1}{\text{SNR}^\nu}\right),$$

(49)

where $\nu = \min\{\beta/2, 1\}$. One can see that at large SNR $I_{P_X^{(\beta)}[X]}$ goes to constant in the considered interval, and this constant depends on noise power $QL$. Recall that $I_{P_{\text{opt}}[X]}$ increases as $\log P$ in the region under study. For the half-Gaussian distribution (9) from (49) we have

$$I_{P_X^{(1)}[X]} = \log \left(\frac{1}{QL^2\gamma}\right) +$$

$$+ \log 12\pi - 1 + \gamma_E + O\left(\frac{1}{\sqrt{\text{SNR}}}\right).$$

(50)
Comparing our expression (50) with the result of the paper [11] (see Eq. (40) therein) we have an extra term $+\log 2$ due to more accurate calculation of $H[Y|X]$.

Since we found the $P_{\text{opt}}[X]$ in the leading order in SNR, we present an accurate lower bound for the capacity of the considered per-sample nonlinear channel in the region $QL \ll P \ll (\gamma^2 Q L^3)^{-1}$:

$$C = I_{P_{\text{opt}}[X]} + O \left( \frac{1}{\sqrt{\text{SNR}}} \right).$$  
(51)

The comparison of the capacity low bound (51) with the Shannon capacity of the linear AWGN channel is presented in Fig. 2. One can see that the Shannon capacity of the linear AWGN channel is always greater than the low bound of the capacity (51) for the nondispersive nonlinear fiber channel for the considered region of $P$.

V. CONCLUSION

We have developed a new approach to the conditional probability calculation via the path-integral representation [13] in the limit of large SNR. This may be especially useful technique for complex nonlinear channels in which calculation of the conditional probability is technically challenging. Applying the developed method to per-sample nondispersive nonlinear fiber channel, we derived the compact analytical expressions for the conditional probability, conditional entropy and the entropy of the output signal for varying input signal PDF $F_X[X]$.

Moreover, we solved the variational problem on $P_X[X]$ maximizing the mutual information in the leading in $1/\text{SNR}$ order. That allows us to find the optimal input signal distribution (36) up to $1/\text{SNR}$ corrections and the lower bound on the channel capacity (45) up to $1/\sqrt{\text{SNR}}$ corrections in the power interval $P: QL \ll P \ll (\gamma^2 Q L^3)^{-1}$. We demonstrated that the found lower bound on the considered channel capacity (51) is always greater than the mutual information calculated for Gaussian and half-Gaussian distributions, and lower than the Shannon capacity of the linear AWGN channel.

Acknowledgments A. V. thanks the Dynasty foundation for financial support and the President support program. Support from a grant from the Ministry of Education and Science of the Russian Federation (agreement No. 14.B25.31.0003) and the EPSRC project UNLOC is acknowledged.

REFERENCES

[1] C. Shannon, "A mathematical theory of communication", Bell System Techn. J., vol. 27, no. 3, pp. 379–423, 1948; vol. 27, no. 4, pp. 623–656, 1948.
[2] R.-J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity Limits of Optical Fiber Networks", J. of Lightwave Technol., vol. 28, no. 4, pp. 662–701, 2010.
[3] P. P. Mitra and J. B. Stark, "Nonlinear limits to the information capacity of optical fibre communications", Nature, vol. 411, pp. 1027-1030, April 2001.
[4] A. Mecozzi and M. Shetaif, "On the capacity of intensity modulated systems using optical amplifiers", IEEE Photonics Technol. Lett., vol. 13, no. 9, pp. 1029 - 1031, September 2001.
[5] E. E. Narimanov and P. Mitra, "The channel capacity of a fiber optics communication system: Perturbation theory," J. Lightwave Technol., vol. 20, no. 3, pp. 530–537, March 2002.
[6] J. M. Kahn and K.-P. Ho, "Spectral efficiency limits and modulation detection techniques for DWDM systems," IEEE. J. Sel. Topics Quant. Electron., vol. 10, no. 2, pp. 259–272, March/April 2004.
[7] R. J. Essiambre, G. J. Foschini, G. Kramer, and P. J. Winzer, "Capacity Limits of Information Transport in Fiber-Optic Networks", Phys. Rev. Lett., vol. 101, p. 163901, October 2008.
[8] R. Killey and C. Behrens, "Shannon’s theory in nonlinear systems", J. Mod. Opt., vol. 58, no. 1, pp. 1–10, January 2011.
[9] A. D. Ellis and J. Zhao, "Impact of Nonlinearities on Fiber Optic Communications", Springer, New York, 2011.
[10] K.S. Turitsyn, S.A. Derevyanko, I.V. Yurkevich, and S.K. Turitsyn, "Information Capacity of Optical Fiber Channels with Zero Average Dispersion", Phys. Rev. Lett., vol. 91, p. 203901, November 2003.
[11] M. I. Youssef, R. K. Chisichian, "On the Per-Sample Capacity of Nondispersive Optical Fibers", IEEE transactions on information theory, vol. 57, no. 11, pp. 7522 - 7541, November 2011.
[12] E. Agrell, "The Channel Capacity Increases with Power", arXiv: 1108.0391v3.
[13] E. Agrell, "Nonlinear Fiber Capacity", Eur. Conf. Opt. Commun. London U.K., paper We.4.D.3, 2013.
[14] E. Agrell, A. Alvarado, G. Durisi, M. Karlsson, "Capacity of a Nonlinear Optical Channel with Finite Memory", arXiv:1403.1393.
[15] K. S. Turitsyn, S. K. Turitsyn, "Nonlinear communication channels with capacity above the linear Shannon limit", Opt. Lett., vol. 37, no. 17, pp. 3600-3602, September 2012.
[16] M. A. Sorokina and S. K. Turitsyn, "Regeneration limit of classical Shannon capacity", Nat. Comm., vol. 5, p. 3861, Feb 2014.
[17] I. S. Terekhov, S. S. Vergeles, and S. K. Turitsyn, "Conditional Probability Calculations for the Nonlinear Schrödinger Equation with Additive Noise", Phys. Rev. Lett., vol. 113, p. 230602, December 2014.
[18] A. Mecozzi, "Limits to long-haul coherent transmission set by the Kerr nonlinearity and noise of the in-line amplifiers", J. Lightwave Technol., vol. 12, no. 11, pp. 1993 – 2000, November 1994.
[19] J. Zinn-Justin, "Quantum Field Theory and Critical Phenomena", Oxford University Press, Oxford, 2002.
[20] R. P. Feynman, A. R. Hibbs, "Quantum mechanics and path integrals", McGraw-Hill Book Company, New York, 1965.
[21] Supplementary Materials.
[22] I. S. Gradshtein and I. M. Ryzik, "Table of Integrals and Series, and Products", Academic Press, Orlando, Florida, 2014.