Path-integral approach to mutual information calculation for nonlinear communication channel with small second dispersion at large signal-to-noise power ratio

A. V. Reznichenko[1,2] and V. O. Guba[1]

[1]Theory department, Budker Institute of Nuclear Physics of Siberian Branch Russian Academy of Sciences, Novosibirsk 630090, Russia
[2]Novosibirsk State University, Novosibirsk 630090, Russia

E-mail: A.V.Reznichenko@inp.nsk.su

Abstract

We consider the information fiber optical channel modeled by the nonlinear Schrodinger equation with additive Gaussian noise. Using path-integral approach and perturbation theory for the small dimensionless parameter of the second dispersion, we calculate the conditional probability density functional in the leading and next-to-leading order in the dimensionless second dispersion parameter associated with the input signal bandwidth. Taking into account specific filtering of the output signal by the output signal receiver, we calculate the mutual information in the leading and next-to-leading order in the dispersion parameter and in the leading order in the parameter signal-to-noise ratio (SNR). Further, we find the explicit expression for the mutual information in case of the modified Gaussian input signal distribution taking into account the limited frequency bandwidth of the input signal.

1   Introduction

Analytical description of information transmission through a nonlinear communication channel remains an open problem of information theory, as it significantly depends on the physical model of the channel, receiver and the decoding procedure. Addressing this problem, one usually aims to calculate the channel capacity, which determines the maximum amount of information per symbol that can be transmitted through a channel. The concept of channel capacity

1

was introduced by Shannon in the paper [?]. In the same work he calculated the channel capacity of a linear channel with a Gaussian noise:

$$C \propto \log_2(1 + R_{sn}),\qquad(1)$$

where $R_{sn} = P/N$ is the signal-to-noise ratio (SNR), P is the input signal power and N is the noise power. Relation (??) suggests that to increase the channel capacity one should increase the input signal power. This formula is often used for estimations in cases of nonlinear channels or channels with non-Gaussian noise. However, nonlinear effects become significant in fiber-optic communication channels if the signal power grows large, which makes consideration of nonlinear communication channels necessary.

A model that is widely used to describe signal propagation through a noisy nonlinear optical fiber channel is the nonlinear Schrodinger equation (NLSE) with the additive white Gaussian noise (AWGN) [?]:

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

The term $i\gamma|\psi|^2\psi$, where γ is called the Kerr nonlinearity, in the equation (??), is introduced to describe nonlinear effects (self-phase modulation, cross-phase modulation, etc.) due to the Kerr effect, i.e. the effect of changing of the refractive index of the fiber in proportion to the square of the strength of the electric field applied. The term with the second time derivative is responsible for the dispersion, and β is called the second dispersion. We assume that the noise function $\eta(z,t)$ in the right-hand side of eq. (??) has zero mean and finite frequency bandwidth W' of the correlation function:

$$\langle \eta(z,t) \rangle_\eta = 0,\qquad(3)$$

$$\langle \eta(z,t)\overline{\eta}(z',t') \rangle_\eta = \frac{Q}{\pi(t-t')}\sin\left(\frac{W'(t-t')}{2}\right)\delta(z-z'),\qquad(4)$$

where Q is the noise power per unit length and per unit frequency, and W' is the bandwidth of the noise. The right-hand side of eq. (??) implies that the noise only exists within the frequency interval $[-W'/2, W'/2]$. The brackets $\langle...\rangle_\eta$ mean the averaging over the noise realizations in the channel. The bandwidth W' is introduced to regularize the problem: when $W' \to +\infty$ one has:

$$\langle \eta(z,t)\overline{\eta}(z',t') \rangle_\eta \to Q\delta(t-t')\delta(z-z').\qquad(5)$$

In this simple model described by the eq. (??) the noise comes from the equally spaced amplifiers that compensate for the signal attenuation.

An explicit expression for the capacity of a nonlinear noisy channel with an arbitrary dispersion is not yet known. Considering simplified cases allows one to achieve some insight into the transmission of information through a nonlinear channel. One of such cases is the channel with zero dispersion. When the dispersion is zero, the signals at different time moments propagate independently, which effectively allows one to consider time-independent problem. Such treatment is known as the per-sample model. However, the per-sample model has

several limitations. It does not describe spectral broadening of the propagating signal and limited bandwidth of the receiver. Moreover, in the work [?] it was shown that in the per-sample model with the finite frequency bandwidth of the noise the capacity can reach arbitrarily large values for any input power P. Namely, one achieves large rates of information transmission by sending signals at frequencies that belong to the noise-free part of the spectrum. The work [?] also develops upper bounds for the channel capacity of dispersion-free fiber.

The zero-dispersion case was also considered in papers [?, ?, ?, ?, ?]. In these works the expression for the conditional probability distribution functional (PDF) of the output signal was obtained in the form of functional integral due to the Martin-Siggia-Rose formalism [?, ?]. The conditional PDF, which we denote in our work as $P[Y|X]$, is the probability density to detect the output signal Y if the input signal is X. The path-integral approach allows for a perturbative study of the conditional PDF and other information characteristics. The channel capacity of the per-sample model was shown to grow with the average input signal power P as $\log \log P$ in the so-called intermediate power range $N \ll P \ll (N\gamma^2 L^2)^{-1}$, where L is the channel length and γ is the Kerr nonlinearity.

In order to grasp effects of the dispersion one may consider channels with small nonlinearity and apply the perturbative approach. This approach was used in [?] to assess the influence of the nonlinear interference noise on the capacity. The authors obtained upper and lower bounds for the capacity and demonstrated that the difference between the bounds is independent of any parameter of the system in the first order of the perturbation analysis.

In [?] the perturbative study of the nonlinear channel with arbitrary dispersion and large SNR was carried out using the path-integral approach and by the direct calculation of the output signal correlators.

In this paper we address the problem of dispersive communication channels described by the equation (??) developing perturbative description of the channel with arbitrary Kerr nonlinearity and small second dispersion β. Saying that β is small, we imply in our work smallness of the dimensionless parameter associated with the second dispersion and with the bandwidth of the input signal X.

This paper is organized as follows. In the Second section we use the path-integral representation to perturbatively calculate the conditional PDF in the first order in $1/R_{sn}$ and in the leading and the next-to-leading order in the second dispersion. The expression that we obtain for the conditional PDF is valid in the intermediate power range $N \ll P \ll (N\gamma^2 L^2)^{-1}$. In the Third section we take into account the limited bandwidth of the receiver. It is achieved by integration of the conditional PDF $P[Y|X]$ over the high-frequency Fourier components of the output signal Y which are not distinguished by the receiver. Then we use the normalization condition to determine the normalization factor of the conditional PDF for the observed components of Y. As it was demonstrated in the paper [?], the normalization factor plays a significant role in the calculation of the mutual information. Namely, in the leading order in $1/R_{sn}$ the mutual information is proportional to the logarithm of the normalization factor averaged

3

over realizations of the input signal. The Fourth section contains evaluation of the information characteristics: output signal PDF, conditional entropy, entropy of the output signal and mutual information. The expression for the mutual information is used to evaluate the transmission rate for a specific distribution of the input signal in the Fifth section. Namely, we assumed Gaussian distribution with the limited bandwidth. Such form of the input signal distribution allowed us to use results from the theory of the quantum harmonic oscillator evolving in the imaginary time. The appendices A and B contain details of the calculation of the conditional PDF $P[Y|X]$. The appendix C complements the Fifth section with the details of the calculation of the transmission rate.

2   Model of the channel

2.1   Propagation of the signal

In our work propagation of the signal $\psi(z,t)$ is governed by the nonlinear Schrödinger equation with additive Gaussian noise [?, ?, ?, ?]:

$$\partial_z \psi + i\beta \partial_t^2 \psi - i\gamma |\psi|^2 \psi = \eta(z,t), \tag{6}$$

with the input condition $\psi(z=0,t) = X(t)$. The function $\psi(z,t)$ is related to the modulation of the electric field in the fiber: $\vec{E}(z,t) = \vec{e}_0 \mathrm{Re}\{\psi(z,t + \frac{z}{v_g})e^{ik_0 z - i\omega_0 t}\}$, where ω_0 is the frequency of the carrier wave, $k_0 = k(\omega_0)$ is the corresponding wave vector, $v_g = \frac{d\omega}{dk}(k_0)$ is the group velocity and $\vec{e}_0$ is the polarization vector. The coefficients β and γ are the second dispersion and the Kerr nonlinearity, respectively. The equation (??) does not contain terms responsible for the attenuation of the signal, because in our model we have distributed amplifiers which compensate the attenuation and also cause additive noise $\eta(z,t)$ to appear [?]. In our model of the noise the random function $\eta(z,t)$ has statistical properties of the Gaussian noise with finite frequency bandwidth W':

$$\langle \eta(z,t) \rangle_\eta = 0,$$

$$\langle \eta(z,t) \bar{\eta}(z',t') \rangle_\eta = \frac{Q}{\pi(t-t')} \sin\left(\frac{W'(t-t')}{2}\right) \delta(z-z'), \tag{7}$$

where $\langle ... \rangle_\eta$ is the averaging over the realizations of the function $\eta(z,t)$, Q is the power of the noise per unit length and per unit frequency, and the bar stands for the complex conjugation. Meaning of the correlators (??) is transparent in the frequency domain. Namely, for the Fourier transform of the noise function:

$$\eta(z,\omega) = \int_{-\infty}^{+\infty} dt e^{i\omega t} \eta(z,t) \tag{8}$$

the correlation function reads:

$$\langle \eta(z,\omega) \bar{\eta}(z',\omega') \rangle_\eta = 2\pi Q \delta(\omega - \omega') \theta\left(\frac{W'}{2} - |\omega|\right) \delta(z-z'), \tag{9}$$

where $\theta(x)$ is the Heaviside theta-function. From the eq. (??) one sees that the noise is not zero within the frequency interval $[-W'/2, W'/2]$, and the limit $W' \to +\infty$ corresponds to the case of white noise. In our work the bandwidth W' plays the role of the ultraviolet cutoff. We are not considering effects of the signals which propagate at such frequencies that $|\omega| > W'/2$.

2.2 Receiver model

After the process of propagation the output signal $Y(t) = \psi(z = L, t)$ is detected by a receiver, where L is the distance of propagation. In our work we wish to take into account finite bandwidth W_d of the receiver. There may be, in general, different ways of including the receiver bandwidth W_d into the model. Finite bandwidth of the receiver implies that high-frequency Fourier harmonics of the output signal $Y(\omega)$ are not distinguished. A straightforward way to get rid of the unobservable degrees of freedom is to integrate the conditional PDF over these high-frequency modes $Y(\omega)$, $W_d/2 < |\omega| < W'/2$:

$$P_d[Y(\omega)|X] = \int_{\{Y(\omega):|\omega|>\frac{W_d}{2}\}} DY\, P[Y(\omega)|X]. \tag{10}$$

The new functional $P_d[Y(\omega)|X]$ depends only on the observable degrees of freedom and can be used to calculate information characteristics.

However, one may think of a different detection procedure. For instance, we can choose the receiver which does not distinguish the high-frequency Fourier harmonics of a different function $\tilde{Y}$, where $\tilde{Y}$ is functionally related to the output signal Y. The function $\tilde{Y}$ can be chosen to significantly simplify the analytical consideration, but we get a different functional integrating out the unobservable harmonics of $\tilde{Y}$:

$$\tilde{P}_d[Y(\omega)|X] = \int_{\{\tilde{Y}(\omega):|\omega|>\frac{W_d}{2}\}} D\tilde{Y}\, P[Y(\omega)|X]. \tag{11}$$

For the functional $\tilde{P}_d$ to be a good approximation of the functional (??), it is reasonable to choose $\tilde{Y}$ so that $\tilde{Y}(\omega) \approx Y(\omega)$ at large $|\omega|$. In other words, introduction of $\tilde{P}_d$ can be treated both as the model of a new receiver, or as the analytical trick to perform the approximate integration in the eq. (??).

2.3 Input signal

We consider input signals $X(t)$ with fixed average power P and fixed average bandwidth W_X, defined as:

$$P = \int \frac{dt}{T} \langle |X(t)|^2 \rangle_{P_X}, \tag{12}$$

$$W_X^2 = \int \frac{dt}{T} \langle |\dot{X}(t)|^2 \rangle_{P_X} \bigg/ \int \frac{dt}{T} \langle |X(t)|^2 \rangle_{P_X}, \tag{13}$$

5

where the brackets $\langle ...\rangle_{P_X} = \int DX(...)P_X[X]$ are the average with respect to the input signal distribution P_X. The normalization condition for the input signal distribution reads:

$$\int DX P_X[X] = 1, \qquad (14)$$

where the integration measure is $DX(t) \equiv \prod_i d\mathrm{Re}X(t_i)d\mathrm{Im}X(t_i)$. We consider such input signals $X(t)$ that vanish outside of the time interval $[-T/2, T/2]$. We also assume that the input signal is a slowly varying function of time in the sense that $\beta L W_X^2 \ll 1$. The parameters $\beta L W_X^2$ and $\beta L W_d^2$ play the roles of the small dimensionless parameters of the perturbative expansion in our work.

Having introduced all the bandwidths that we work with, we can specify the hierarchy. We assume the following relation on the receiver bandwidth:

$$\beta L W_d^2 \ll 1, \qquad (15)$$

and the following hierarchy:

$$W_X \lesssim W_d \ll W'. \qquad (16)$$

3 Conditional probability density functional

3.1 The path-integral representation of the conditional PDF

The path-integral approach to the calculation of the conditional PDF $P[Y|X]$ is based on the Martin-Siggia-Rose formalism [?], which allows us to write the functional $P[Y|X]$ in the following form:

$$P[Y|X] = \int_{\substack{\psi(z=0,t)=X(t)}}^{\substack{\psi(z=L,t)=Y(t)}} D\psi \exp\{-S[\psi]/Q\}, \qquad (17)$$

where the effective action $S[\psi]$ is associated with the noise statistics and is the integral of the squared left-hand side of the equation (??):

$$S[\psi] \equiv \int_T dt \int_0^L dz \left| \partial_z \psi + i\beta \partial_t^2 \psi - i\gamma |\psi|^2 \psi \right|^2. \qquad (18)$$

When calculating the path integral (??), one should use the retarded discretization scheme in evolution variable z [?]: $(\partial_z \psi)(z_n, t) = (\psi(z_n, t) - \psi(z_{n-1}, t))/\delta_z$, where $z_n = n\delta_z$, $\delta_z = L/N$ ($z_0 = 0$, $z_N = L$). This choice of discretization scheme is motivated by the causal structure of the propagation process [?, ?]. The discretization step of time grid is $\delta_t = T/M$, and points of the time grid are defined as $t_m = m\delta_t = T/2$ ($t_0 = -T/2$, $t_M = T/2$). The discretization step is larger than $1/W'$. The integration measure $D\psi$ from the eq. (??) is defined as:

$$D\psi = \left(\frac{\delta_t}{\delta_z \pi Q}\right)^{2M} \prod_{i=1}^{N-1} \prod_{j=0}^{M} \left\{ \frac{\delta_t}{\delta_z \pi Q} d\mathrm{Re}\psi(z_i, t_j) d\mathrm{Im}\psi(z_i, t_j) \right\}. \qquad (19)$$

6

For the case of large R_{sn}, which we consider in our work, it is convenient to factorize the contribution of the classical solution to the Euler-Lagrange equation for the effective action $S[\psi]$:

$$P[Y|X] = \Lambda \exp\{-S[\Psi_{cl}]/Q\}, \tag{20}$$

where Ψ_{cl} is the solution of the equation $\delta S[\Psi_{cl}] = 0$ with the fixed boundary conditions $\Psi_{cl}(z = 0, t) = X(t)$ and $\Psi_{cl}(z = L, t) = Y(t)$, and the "normalization factor" Λ is equal to:

$$\Lambda = \int_{\varphi(z=0,t)=0}^{\varphi(z=L,t)=0} D\varphi \exp\left[-\frac{1}{Q}\left\{S[\Psi_{cl} + \varphi] - S[\Psi_{cl}]\right\}\right]. \tag{21}$$

Large R_{sn} allows us to use the quasiclassical approach, which means that the "normalization factor" Λ does not depend on properties of output signal Y(t), but depends only on the input signal $X(t)$ in the leading order in $1/R_{sn}$. Therefore, in order to find P[Y|X] we should first calculate the "classical" solution Ψ_{cl} and then calculate the normalization factor Λ using the normalization condition:

$$\int DY\, P[Y|X] = 1. \tag{22}$$

As we mentioned before, the normalization factor Λ plays a significant role in the calculation of the information characteristics. The variational problem $\delta S[\psi] = 0$ which gives the "classical" solution Ψ_{cl} can be written explicitly as the following boundary problem:

$$\begin{cases} (\partial_z + i\beta\partial_t^2 - 2i\gamma|\Psi_{cl}|^2)\mathcal{L}[\Psi_{cl}] + i\gamma\Psi_{cl}^2\overline{\mathcal{L}[\Psi_{cl}]} = 0, \\ \mathcal{L}[\Psi_{cl}] = (\partial_z + i\beta\partial_t^2 - i\gamma|\Psi_{cl}|^2)\Psi_{cl}, \\ \Psi_{cl}(z = 0, t) = X(t), \\ \Psi_{cl}(z = L, t) = Y(t), \end{cases} \tag{23}$$

where $\overline{\mathcal{L}[\Psi_{cl}]}$ is the complex conjugate of the function $\mathcal{L}[\Psi_{cl}]$. In the large R_{sn} case we can linearize the problem (??). Namely, we present the solution to the problem (??) in the form of a small deviation from the solution $\Phi(z, t)$ to the NLSE (??) without noise:

$$\Psi_{cl} = \Phi(z, t) + \kappa(z, t)\exp\left\{i\mu\frac{z}{L} + i\phi\right\}, \tag{24}$$

where $\mu = \gamma L\rho^2$, $X = \rho e^{i\phi}$, and the function $\Phi(z, t)$ is the solution to the following problem:

$$\begin{cases} \partial_z\Phi + i\beta\partial_t^2\Phi - i\gamma|\Phi|^2\Phi = 0, \\ \Phi(z = 0, t) = X(t). \end{cases} \tag{25}$$

To understand why we can treat the function $\kappa(z, t)$ as a small deviation, one should consider the behaviour of the functional $S[\psi]$ around $\Phi(z, t)$. The action $S[\psi]$ achieves absolute minimum ($S[\Phi] = 0$) on the function $\Phi(z, t)$, so

7

the expansion of the functional $S[\Phi + \kappa \exp\{i\mu\frac{z}{L} + i\phi\}]$ starts from quadratic terms for small κ. Therefore, the conditional PDF $P[Y|X] \propto \exp[-S[\Psi_{cl}]/Q]$ decreases exponentially for such configurations of the output signal $Y(t)$ that make $|\kappa(z,t)|$ much larger than $\sqrt{QLW_X}$. The most significant contribution comes from $\kappa(z,t) \propto \sqrt{QLW_X}$. To find $P[Y|X]$ in the leading order in $1/R_{sn}$ it will be sufficient to solve the linearized version of the problem (??).

Let us start from calculation of $\Phi(z,t)$ in the form of perturbative expansion in the small second dispersion β:

$$\Phi = \Phi_0 + \Phi_1, \tag{26}$$

where $\Phi_k \propto \beta^k$. Substituting this expansion in eq. (??), we get problems that define terms $\Phi_{0,1}$. For the leading order term we get the following per-sample problem:

$$\begin{cases} \partial_z\Phi_0 - i\gamma|\Phi_0|^2\Phi_0 = 0, \\ \Phi_0(z=0,t) = X(t), \end{cases} \tag{27}$$

and then its solution reads:

$$\Phi_0(z,t) = X(t)e^{i\gamma z|X(t)|^2} = X(t)e^{i\mu(t)\frac{z}{L}}, \tag{28}$$

For the next-to-leading order Φ_1 we get the following problem:

$$\begin{cases} \partial_z\Phi_1 - i\gamma\Phi_0^2\overline{\Phi}_1 - 2i\gamma|\Phi_0|^2\Phi_1 = -i\beta\partial_t^2\Phi_0, \\ \Phi_1(z=0,t) = 0. \end{cases} \tag{29}$$

This problem has the following solution:

$$\Phi_1(z,t) = \beta\exp\left\{i\left(\mu\frac{z}{L}+\phi\right)\right\}\left\{z[2\dot{\phi}\dot{\rho}+\rho\ddot{\phi}]+z^2\gamma\rho^2[\ddot{\rho}+3\frac{\dot{\rho}^2}{\rho}]+\right.$$
$$\left. i\left(z^3(\gamma\rho^2)^2\frac{2}{3}[\ddot{\rho}+5\frac{\dot{\rho}^2}{\rho}]+z^2\gamma\rho^2[4\dot{\phi}\dot{\rho}+\rho\ddot{\phi}]+z[\rho\dot{\phi}^2-\ddot{\rho}]\right)\right\}. \tag{30}$$

In order to find the "classical" solution Ψ_{cl} and the action $S[\Psi_{cl}]$ we also use the perturbative approach, expanding the function $\kappa(z,t)$ in terms of the small parameter of the second dispersion:

$$\kappa = \kappa_0 + \kappa_1. \tag{31}$$

The linearized problem (??) in the leading order in β reads:

$$\begin{cases} \partial_z^2\kappa_0 - 2i\gamma\rho^2\kappa_0 - 4(\gamma\rho^2)^2\text{Re}[\kappa_0] = 0, \\ \kappa_0(z=0,t) = 0, \\ \kappa_0(z=L,t) = [Y(t)-\Phi(z=L,t)]\,e^{-i\phi-i\mu} \end{cases} \tag{32}$$

Following [?], we express the solution to the problem (??) defining new real-valued functions $x(t)$ and $y(t)$ as $x(t)+iy(t) = [Y(t)-\Phi(L,t)]\,e^{-i\phi-i\mu}$:

$$\text{Re}[\kappa_0] = \left(\mu\frac{\mu x-y}{1+\mu^2/3}\frac{z}{L}+\frac{(1-2\mu^2/3)x+\mu y}{1+\mu^2/3}\right)\frac{z}{L},$$
$$\text{Im}[\kappa_0] = \left(\frac{\mu x-y}{1+\mu^2/3}\left[\frac{2\mu^2 z^2}{3L^2}-1\right]+\mu\frac{(1-2\mu^2/3)x+\mu y}{1+\mu^2/3}\right)\frac{z}{L}. \tag{33}$$

The function κ_1 is the solution to the problem which is similar to the problem (??), but with a non-zero right-hand side and with zero boundary conditions:

$$\begin{cases} \partial_z^2 \kappa_1 - 2i\gamma\rho^2 \kappa_1 - 4(\gamma\rho^2)^2 \text{Re}[\kappa_1] = F(z,t), \\ \kappa_1(z=0,t) = 0, \\ \kappa_1(z=L,t) = 0, \end{cases} \qquad (34)$$

because the condition $\Psi_{cl}(z=L,t) = Y(t)$ is already satisfied by the leading order contribution κ_0. Expression for right-hand side $F(z,t)$ is rather cumbersome, so we present it in the Appendix A: see eq. (??).

With the expression for Ψ_{cl} we can obtain expressions for the action $S[\Psi_{cl}] = S_{cl}$ in the leading and next to leading orders in parameter β. The leading order contribution reads:

$$S_{cl}^{(0)} = \int_T dt \frac{(1 + 4\mu^2/3)x^2 - 2\mu xy + y^2}{L(1 + \mu^2/3)}. \qquad (35)$$

The first correction to the $S_{cl}^{(0)}$ can be written in the following form:

$$S_{cl}^{(1)} = \beta \int_T dt \Big\{ a_1(t)x^2 + a_2(t)y^2 + a_3(t)xy + a_4(t)x\dot{x} + a_5(t)y\dot{y} +$$
$$a_6(t)x\dot{y} + a_7(t)\dot{x}y + a_8(t)x\ddot{x} + a_9(t)y\ddot{y} + a_{10}(t)x\ddot{y} + a_{11}(t)\ddot{x}y \Big\}, \qquad (36)$$

where we have introduced functions $a_i(t)$ which can be found in the Appendix A.

3.2 The transformation of the receiver harmonics

Now that we know the action S_{cl} in the leading and next-to-leading orders in β, we proceed to take into account the limited bandwidth of the receiver. As we have stated in the section 2.2, our treatment of the receiver implies integrating the functional $P[Y|X]$ over the high-frequency Fourier harmonics of complex-valued function $Y(t)$. In what follows we choose such a model of the receiver, that the function $\check{Y}(t) = y_1(t) + iy_2(t)$, where $y_{1,2}(t)$ are real-valued, is defined by the following equation:

$$\frac{y_1(t)}{\sqrt{\alpha_1}} + i\frac{y_2(t)}{\sqrt{\alpha_2}} = [Y(t) - \Phi(z=L,t)]e^{i\theta - i\mu - i\phi} = [x(t) + iy(t)]e^{i\theta}, \qquad (37)$$

where $\theta = \frac{1}{2}\arctan\frac{2}{3\mu}$, and $\alpha_{1,2}$ are eigenvalues of the symmetric matrix that represents the quadratic form in the variables x and y from the eq. (??):

$$S_{cl}^{(0)} = \int dt \frac{1}{L} (x,y) \begin{pmatrix} \frac{1 + 4\mu^2/3}{1 + \mu^2/3} & \frac{-\mu}{1 + \mu^2/3} \\ \frac{-\mu}{1 + \mu^2/3} & \frac{1}{1 + \mu^2/3} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \qquad (38)$$

or explicitly:

$$\alpha_1 = \frac{3 + 2\mu^2 - \mu\sqrt{9 + 4\mu^2}}{3 + \mu^2}, \ \alpha_2 = \frac{3 + 2\mu^2 + \mu\sqrt{9 + 4\mu^2}}{3 + \mu^2}. \qquad (39)$$

In other words, real and imaginary parts of the function $\tilde{Y}(t)$ are related to functions $x(t)$ and $y(t)$ through the following transformation:

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \tag{40}$$

where the coefficients $A_{ij}(t)$ are

$$A_{11} = \frac{1}{2}\sqrt{2 - \frac{2\mu}{\sqrt{9 + 4\mu^2}}}, \quad A_{12} = \frac{1}{2}\sqrt{2 + \frac{2\mu}{\sqrt{9 + 4\mu^2}}},$$

$$A_{21} = -\sqrt{\frac{1}{2} + \frac{2\mu^2}{3} - \frac{\mu(15 + 8\mu^2)}{6\sqrt{9 + 4\mu^2}}}, \quad A_{22} = \sqrt{\frac{1}{2} + \frac{2\mu^2}{3} + \frac{\mu(15 + 8\mu^2)}{6\sqrt{9 + 4\mu^2}}}. \tag{41}$$

The way we chose the function $\tilde{Y}(t)$ is dictated by two reasons. First of all, the leading order contribution to the "classical" action in terms of the function $\tilde{Y}(t)$ is now diagonal:

$$S_{cl}^{(0)}[\tilde{Y}(t)] = \frac{1}{L}\int_T dt \left[y_1^2(t) + y_2^2(t) \right]. \tag{42}$$

The diagonal form of the action (??) implies that integration of the functional $P[Y|X]$ over Fourier components of the function $\tilde{Y}$ defined by the eq. (??) is straightforward due to the Parseval's theorem. However, one should keep in mind that the first correction in β to the action S_{cl} should be expressed through the functions $y_{1,2}(t)$ as well:

$$S_{cl}^{(1)}[\tilde{Y}(t)] = \beta \int_T dt \{ b_1 y_1^2 + b_2 y_2^2 + b_3 y_1 y_2 + b_4 y_1 \dot{y}_1 + b_5 y_2 \dot{y}_2 +$$

$$b_6 y_1 \dot{y}_2 + b_7 \dot{y}_1 y_2 + b_8 y_1 \ddot{y}_1 + b_9 y_2 \ddot{y}_2 + b_{10} y_1 \ddot{y}_2 + b_{11} \ddot{y}_1 y_2 \}, \tag{43}$$

where coefficients $b_i(t)$ can be expressed through functions $a_i(t)$ (listed in eqs. (??)-(??)) and $A_{ij}(t)$ (see eq. (??)). The expressions for the functions $b_i(t)$ turned out to be rather cumbersome. Fortunately, in order to calculate information characteristics in subsequent sections we only need specific combinations of the functions $b_i(t)$. Namely, we only need expressions for $b_1(t) + b_2(t)$ and $b_8(t) + b_9(t)$, which turned out to be relatively simple, and we list it in the Appendix B, see eqs. (??). In the appendix B we also show how the functions $b_i(t)$ can be expressed in terms of the functions $a_i(t)$ and $A_{ij}(t)$.

The second reason for our choice of the function $\tilde{Y}(t)$ is more subtle. We claim that under our assumptions about the input signal, which we stated in the section 2.3, Fourier components of functions $Y(t)$ and $\tilde{Y}(t)$ become equivalent at high frequencies in the following sense: two conditional PDFs defined by the eq. (??) and by the eq. (??) do not differ significantly if the condition (??) is satisfied.

10

Let us consider the behaviour of Fourier components of functions $Y(t) = Y_1(t) + iY_2(t)$ and $\tilde{Y}(t) = y_1(t) + iy_2(t)$. The relation between these functions can be written as:

$$Y_i(t) = \sum_{j=1,2} B_{ij}(t)y_j(t) + \Phi_i(t), \tag{44}$$

where the matrix $B_{ij}(t)$ is defined by the equation (??) and $\Phi(z = L, t) = \Phi_1(t) + i\Phi_2(t)$ with $\Phi_1(t)$ and $\Phi_2(t)$ being real-valued functions. Explicitly, the matrix $B_{ij}(t)$ reads:

$$B_{ij} = \begin{pmatrix} \alpha_1^{-1/2} \cos\left(\theta - \mu - \phi\right) & \alpha_2^{-1/2} \sin\left(\theta - \mu - \phi\right) \\ -\alpha_1^{-1/2} \sin\left(\theta - \mu - \phi\right) & \alpha_2^{-1/2} \cos\left(\theta - \mu - \phi\right) . \end{pmatrix} \tag{45}$$

An important observation regarding $B_{ij}(t)$ is that this matrix tends to a constant rotation matrix b_{ij} through the angle of $\pi/4$ as time tends to $\pm\infty$, because the input signal $X(t)$ vanishes. Performing the Fourier transform of the equation (??), we get:

$$Y_i(\omega) = \sum_{j=1,2} b_{ij}y_j(\omega) + \sum_{j=1,2} \int_{-\infty}^{+\infty} dt F_{ij}(t)y_j(t)e^{i\omega t} + \Phi_i(\omega), \tag{46}$$

where $F_{ij}(t) = B_{ij}(t) - b_{ij}$, $B_{ij}(t) \underset{t\to\pm\infty}{\to} b_{ij}$, and the rotation matrix b_{ij} reads:

$$(b_{ij}) = \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}. \tag{47}$$

The Fourier transform is defined for a function of time $f(t)$ as:

$$f(\omega) = \int_{-\infty}^{+\infty} dt f(t)e^{i\omega t}. \tag{48}$$

Now we recall that the input signal vanishes outside of the time interval $[-T/2; T/2]$, so we can replace the integral over time in eq. (??) with the integral with finite limits. Integrating by parts, we get asymptotic behaviour of $Y_i(\omega)$ for high frequencies:

$$Y_i(\omega) \approx \sum_{j=1,2} b_{ij}y_j(\omega) + \frac{1}{i\omega} \sum_{j=1,2} \left(F_{ij}(T/2)y_j(T/2)e^{i\omega T/2} + \right.$$
$$\left. F_{ij}(-T/2)y_j(-T/2)e^{-i\omega T/2} \right) + \Phi_i(\omega), \tag{49}$$

while the asymptotic behaviour of $y_i(\omega)$ is:

$$y_i(\omega) \approx \frac{1}{i\omega} \left(y_i(T/2)e^{i\omega T/2} + y_i(-T/2)e^{-i\omega T/2} \right). \tag{50}$$

11

Due to the factor $F_{ij}(\pm T/2)$ being small we state that at high frequencies $Y_i(\omega)$ and $y_i(\omega)$ differ only by the rotation b_{ij} and the additive term $\Phi_i(\omega)$. The Jacobian determinant of such transformation equals to one, so at high frequencies we have:

$$d\mathrm{Re}Y_1(\omega)d\mathrm{Im}Y_1(\omega)d\mathrm{Re}Y_2(\omega)d\mathrm{Im}Y_2(\omega) = d\mathrm{Re}y_1(\omega)d\mathrm{Im}y_1(\omega)d\mathrm{Re}y_2(\omega)d\mathrm{Im}y_2(\omega). \tag{51}$$

However, the fact that at high frequencies we can change the integration variables from $Y_{1,2}(\omega)$ to $y_{1,2}(\omega)$ does not imply that the detection procedures defined by the eqs. (??) and (??) will give the same results. Namely, due to the relation (??) values of $Y_{1,2}(\omega)$ at high frequencies depend on values of $y_{1,2}(\omega)$ at low frequencies as well:

$$Y_i(\omega) = \sum_{j=1,2} \int \frac{d\omega'}{2\pi} B_{ij}(\omega - \omega')y_j(\omega') + \Phi_i(\omega). \tag{52}$$

From the eq. (??) it follows that in general the result of integration depends on whether we keep fixed the low-frequency values of $Y_{1,2}(\omega)$ or $y_{1,2}(\omega)$. These are related through the matrix $B_{ij}(\omega)$, which inherits its frequency bandwidth from the input signal $X(t)$. In our work we consider the input signal to be a slowly varying function with a small bandwidth, which implies that the value of $B_{ij}(\omega)$ decreases rapidly as $|\omega|$ increases. Therefore, in our case the relation between $Y_{1,2}(\omega)$ and $y_{1,2}(\omega')$ is negligible if $|\omega - \omega'|$ is large. Due to this fact, we expect the conditional PDF defined by the eq. (??) to be a good approximation of the conditional defined by a more straightforward procedure (??).

3.3 The result of the calculation of $P_d[Y|X]$

The details of how we calculated the integral from the eq (??) are presented in the Appendix B. Here we only write out the result:

$$\tilde{P}_d[Y_d|X] = \bar{\Lambda}_d^{(0)} \exp\left\{ -\frac{\bar{\delta}_t}{QL} \sum_{j=0}^{M_d-1} \left[y_1^2(t_j) + y_2^2(t_j) \right] \right\} \left(1 + \frac{\bar{\Lambda}_d^{(1)}}{\bar{\Lambda}_d^{(0)}} - \right.$$

$$\frac{\beta \bar{\delta}_t}{Q} \sum_{j=0}^{M_d-1} \{ b_{1,j} y_{1,j}^2 + b_{2,j} y_{2,j}^2 + b_{3,j} y_{1,j} y_{2,j} + b_{4,j} y_{1,j} (\dot{y}_1)_j +$$

$$b_{5,j} y_{2,j} (\dot{y}_2)_j + b_{6,j} y_{1,j} (\dot{y}_2)_j + b_{7,j} (\dot{y}_1)_j y_{2,j} + b_{8,j} y_{1,j} (\bar{y}_1)_j +$$

$$\left. b_{9,j} y_{2,j} (\ddot{y}_2)_j + b_{10,j} y_{1,j} (\ddot{y}_2)_j + b_{11,j} (\ddot{y}_1)_j y_{2,j} \} \right). \tag{53}$$

In the eq. (??) we introduced a discretized time grid with the discretization step $\bar{\delta}_t$ related to the bandwidth of the receiver: $\bar{\delta}_t = 2\pi/W_d$. We call this grid coarse, because its discretization step is much larger than the discretization

12

step $2\pi/W'$ related to the bandwidth of the noise W'. The functions $y_{1,2}(t)$ now contain only observable Fourier components with frequencies inside of the interval $[-W_d/2, W_d/2]$, and we reflect it by writing the argument of the PDF as Y_d. In the normalization factor, which we denote as $\tilde{\Lambda}_d$, we separated leading order and next-to-leading order in β and wrote it as $\tilde{\Lambda}_d = \tilde{\Lambda}_d^{(0)} + \tilde{\Lambda}_d^{(1)}$. We calculated the normalization factor using the condition:

$$\int DY_d \tilde{P}_d[Y_d|X] = 1. \tag{54}$$

The normalization factor depends on the input signal and has the following form:

$$\tilde{\Lambda}_d^{(0)} = \left(\frac{\bar{\delta}_t}{\pi Q L}\right)^{M_d},$$

$$\frac{\tilde{\Lambda}_d^{(1)}}{\tilde{\Lambda}_d^{(0)}} = M_d \frac{\beta L W_d^2}{12} \int_T \frac{4\mu^3}{15(3+\mu^2)} \frac{dt}{T} - M_d \beta L \int_T \left[\frac{4\mu\dot{\mu}\dot{\phi}}{3+\mu^2} + \right.$$

$$\frac{2(3+2\mu^2)\ddot{\phi}}{3+\mu^2} + \frac{\mu\dot{\mu}^2}{15(3+\mu^2)^3(9+4\mu^2)^2}(10206 + 21303\mu^2 +$$

$$\left. 15399\mu^4 + 4644\mu^6 + 496\mu^8) \right] \frac{dt}{T}, \tag{55}$$

where the integration over time should be understood as the summation over points of the coarse grid as in the eq. (??). However, the input signal $X(t)$ does not change significantly over the period of $\bar{\delta}_t$ (because $W_d \gg W_X$), so we can replace the sum with the continuous limit.

As we can see in the eqs. (??) and (??), the first correction to the functional $\tilde{P}_d[Y_d|X]$ is proportional to the small dimensionless parameters of dispersion (which are $\beta L W_X^2$, $\beta L W_d^2$ and $\beta L W_X W_d$) and does not depend on the noise bandwidth W'. One may ask whether we can be sure if the contributions of higher orders of the perturbative expansion in β will not contain terms which grow infinitely with the noise bandwidth W'. To address this question, we estimated the second order correction to the functional $\tilde{P}_d[Y_d|X]$. We found out that the terms from the second order correction are all bounded.

4 Mutual information

In this section we calculate information characteristics using the PDF $\tilde{P}_d[\tilde{Y}_d|X]$. First of all, we can rewrite the formula (??) in a more compact way:

$$\tilde{P}_d[Y_d|X] = \left(\tilde{\Lambda}_d^{(0)} + \tilde{\Lambda}_d^{(1)}\right)\left(1 - \frac{S_{coarse}^{(1)}}{Q}\right)\exp\left\{-\frac{S_{coarse}^{(0)}}{Q}\right\} \approx$$

$$\left(\tilde{\Lambda}_d^{(0)} + \tilde{\Lambda}_d^{(1)}\right)\exp\left\{-\frac{S_{coarse}^{(0)} + S_{coarse}^{(1)}}{Q}\right\}, \tag{56}$$

where the second equality is true if we neglect all the terms with powers of β higher than one. We denoted the action functional on the coarse grid as S_{coarse}:

$$S_{coarse}^{(0)} = \frac{\bar{\tilde{\delta}}_t}{L} \sum_{j=0}^{M_d-1} \left[y_1^2(t_j) + y_2^2(t_j) \right],$$ (57)

$$S_{coarse}^{(1)} = \beta\bar{\tilde{\delta}}_t \sum_{j=0}^{M_d-1} \left\{ b_{1,j} y_{1,j}^2 + b_{2,j} y_{2,j}^2 + b_{3,j} y_{1,j} y_{2,j} + b_{4,j} y_{1,j} (\dot{y}_1)_j + \right.$$
$$b_{5,j} y_{2,j} (\dot{y}_2)_j + b_{6,j} y_{1,j} (\dot{y}_2)_j + b_{7,j} (\dot{y}_1)_j y_{2,j} + b_{8,j} y_{1,j} (\ddot{y}_1)_j +$$
$$\left. b_{9,j} y_{2,j} (\ddot{y}_2)_j + b_{10,j} y_{1,j} (\ddot{y}_2)_j + b_{11,j} (\ddot{y}_1)_j y_{2,j} \right\}.$$ (58)

We begin the calculation of the information characteristics with the conditional entropy of our channel, defined as follows:

$$H_{Y|X} = -\int DX P_X[X] \left\{ \int D\tilde{Y}_d \tilde{P}_d[\tilde{Y}_d|X] \ln \tilde{P}_d[\tilde{Y}_d|X] \right\}.$$ (59)

The conditional entropy is the measure of the noise impact on the propagation of the signal — it is a negative contribution to the information transmitted through the channel. The integral over $\tilde{Y}_d$ in the eq. (??) consists of two terms:

$$\int D\tilde{Y}_d \tilde{P}_d[\tilde{Y}_d|X] \ln \left[\left(\tilde{\Lambda}_d^{(0)} + \tilde{\Lambda}_d^{(1)} \right) \exp \left\{ -\frac{S_{coarse}^{(0)} + S_{coarse}^{(1)}}{Q} \right\} \right] =$$
$$= \ln[\tilde{\Lambda}_d^{(0)} + \tilde{\Lambda}_d^{(1)}] - \int D\tilde{Y}_d \left(\frac{S_{coarse}^{(0)} + S_{coarse}^{(1)}}{Q} \right) \tilde{P}_d[\tilde{Y}_d|X],$$ (60)

where we have used the normalization condition $\int D\tilde{Y}_d \tilde{P}_d[\tilde{Y}_d|X] = 1$ and the fact that the normalization factor does not depend on the output signal $\tilde{Y}_d$. The remaining integral in the eq. (??) is gaussian and can be evaluated by using the Wick's theorem for gaussian integrals:

$$\int D\tilde{Y}_d \left(\frac{S_{coarse}^{(0)} + S_{coarse}^{(1)}}{Q} \right) \tilde{P}_d[\tilde{Y}_d|X] = M_d.$$ (61)

Thus, the conditional entropy is expressed through the logarithm of the normalization constant averaged over P_X:

$$H_{Y|X} = -\langle \ln[\tilde{\Lambda}_d^{(0)} + \tilde{\Lambda}_d^{(1)}] \rangle_{P_X} + M_d.$$ (62)

In order to calculate the mutual information, which is defined as:

$$I_{P_X} = H_Y - H_{Y|X},$$ (63)

we also need the entropy of the output signal H_Y, which is expressed through the output signal PDF:

$$P_{out}[\tilde{Y}_d] = \int DX P_X[X] \tilde{P}_d[\tilde{Y}_d|X],$$ (64)

$$H_Y = -\int D\tilde{Y}_d P_{out}[\tilde{Y}_d] \ln P_{out}[\tilde{Y}_d].$$ (65)

14

Since the average noise power is much less than the average input signal power $(P \gg QLW_X)$, we can use the Laplace method to obtain the following result in the leading order in $1/R_{sn}$ [?]:

$$P_{out}[\tilde{Y}_d] \approx \mathbb{J}_d[\Phi^{-1}(Y_d)] P_X[\Phi^{-1}(Y_d)], \qquad (66)$$

here $\Phi^{-1}(Y)$ is the nonlinear function which recovers the input condition $\Phi(z = 0, t) = \Phi^{-1}(Y)$ from the corresponding solution of the NLSE without noise taken at the point $z = L$, $Y(t) = \Phi(z = L, t)$. We denoted the Jacobian of the transformation (??) on the coarse grid by $\mathbb{J}_d[X]$:

$$\mathbb{J}_d[X] = \prod_{j=0}^{M_d-1} \sqrt{1 + \mu_j^2/3}. \qquad (67)$$

Now we return to the eq. (??) and make the following substitution:

$$X = \Phi^{-1}(Y). \qquad (68)$$

The Jacobian of the substitution (??) equals one due to the Liouville's theorem for Hamiltonian systems [?]. In the leading order in $1/R_{sn}$ for the entropy of the output signal we get:

$$H_Y = H_X - \langle \ln \mathbb{J}_d[X] \rangle_{P_X} =$$
$$= H_X - M_d \left\langle \int_T \frac{dt}{T} \ln \sqrt{1 + \mu^2/3} \right\rangle_{P_X}. \qquad (69)$$

We get the mutual information subtracting the conditional entropy (??) from the entropy of the output signal (??):

$$I_{P_X} = H_X - M_d \left\langle \int_T \frac{dt}{T} \ln \sqrt{1 + \mu^2/3} \right\rangle_{P_X} -$$
$$M_d + \langle \ln[\bar{\Lambda}_d^{(0)} + \tilde{\Lambda}_d^{(1)}] \rangle_{P_X}. \qquad (70)$$

After substituting the explicit expressions for $\bar{\Lambda}_d^{(0)}$ and $\bar{\Lambda}_d^{(1)}$ from the eq. (??) we get the following result:

$$I_{P_X} = H_X - M_d \left\langle \int_T \frac{dt}{T} \ln \sqrt{1 + \mu^2/3} \right\rangle_{P_X} + M_d \left\langle \left\{ \ln \left[\frac{\tilde{\delta}_t e^{-1}}{\pi Q L} \right] + \right. \right.$$
$$\frac{\beta L W_d^2}{12} \int_T \frac{dt}{T} \frac{4\mu^3}{15(3 + \mu^2)} - \beta L \int_T \frac{dt}{T} \left[\frac{4\mu\dot{\mu}\dot{\phi}}{3 + \mu^2} + \frac{2(3 + 2\mu^2)\ddot{\phi}}{3 + \mu^2} + \right.$$
$$\frac{\mu\dot{\mu}^2}{15(3 + \mu^2)^3(9 + 4\mu^2)^2} (10206 + 21303\mu^2 + 15399\mu^4 +$$
$$\left. \left. \left. 4644\mu^6 + 496\mu^8) \right] \right\} \right\rangle_{P_X}. \qquad (71)$$

15

5 Mutual information for the Gaussian input signal distribution with the restrained signal bandwidth

In this section to get understanding of the behaviour of the mutual information from the eq. (??) we wish to evaluate it with respect to the Gaussian distribution of the input signal P_X, which is the optimal distribution of a linear channel:

$$P_X^{(0)}[X] = \Lambda_X^{(0)} \exp\left[-\frac{1}{P}\int_T \frac{dt}{T}|X|^2\right], \tag{72}$$

where P is assumed to be fixed:

$$\int_T \frac{dt}{T}\langle|X|^2\rangle_{P_X^{(0)}} = P. \tag{73}$$

However, we wish to consider input signals of the fixed average bandwidth W_X, while the exponent in the eq. (??) input does not contain any terms which suppress fast realizations of the input signal (i.e. with large W_X). For this reason we modify the distribution (??) by adding the term proportional to the squared derivative of the input signal:

$$P_X[X] = \Lambda_X \exp\left[-\frac{1}{P}\int_T \frac{dt}{T}\left\{\left(\frac{T}{2\xi}\right)^2|\dot{X}|^2 + |X|^2\right\}\right], \tag{74}$$

where ξ is the real number chosen so that the average power remains to be equal to P:

$$\int_T \frac{dt}{T}\langle|X|^2\rangle_{P_X} = P. \tag{75}$$

In other words, the condition (??) gives us the equation on ξ which reads:

$$3 = 2\xi \operatorname{cth} 2\xi. \tag{76}$$

The equation (??) can be solved numerically. The value of ξ is roughly 1.49. The factor Λ_X is the normalization factor which ensures that the normalization condition is satisfied:

$$\int DX P_X[X] = 1. \tag{77}$$

The average bandwidth of the input signal, as we defined it in the second section, reads:

$$W_X^2 = \frac{1}{P}\int_T \frac{dt}{T}\langle|\dot{X}|^2\rangle_{P_X}. \tag{78}$$

We also recall that the input signal vanishes at the boundaries of the time interval $[-T/2, T/2]$:

$$X(-T/2) = X(T/2) = 0. \tag{79}$$

16

The exponent in the eq. (??) is essentially the action of the quantum oscillator with imaginary time, which can be made explicit if we rename the coefficients in the eq. (??):

$$\frac{1}{\xi P} = m\Omega, \ 2\xi = \Omega T. \tag{80}$$

Then the input signal distribution reads:

$$P_X[X] = \Lambda_X e^{-S[X]}, \tag{81}$$

and for the analogy to be completely evident we also mention that the oscillator is considered in such a system of units that $\hbar = 1$. The action $S[X]$ is:

$$S[X] = \int_T dt \left(\frac{m|\dot{X}|^2}{2} + \frac{m\Omega^2 |X|^2}{2} \right). \tag{82}$$

With the action (??) we will be able to use well-known expressions from statistical mechanics for averages with respect to the distribution (??).

Calculating averages over P_X in the eq. (??) with respect to the distribution (??), one has to choose a discretization scheme for derivatives. In the Appendix C we use the following discretization for the derivative operator D_{ij}:

$$\dot{g}_i = \sum_{j=0}^{M} D_{ij} g_j = \frac{1}{\delta_t} (g_{i+1} - g_i), \ 0 \le i \le M - 1, \tag{83}$$

where $g_i = g(t_i)$, $\dot{g}_i = \dot{g}(t_i)$ and t_i is a point of the time grid with the discretization step δ_t: $t_i = -T/2 + i\delta_t$. Evaluation of the average power and average bandwidth with the discretization scheme (??) gives:

$$P = \frac{1}{m\Omega} \frac{\Omega T \operatorname{cth} \Omega T - 1}{\Omega T}, \tag{84}$$

$$W_X^2 = \frac{2\delta_t}{PT} \sum_{\alpha=1}^{M-1} \frac{\nu_\alpha}{Q_\alpha} = \frac{2(M-1)}{mPT} - \frac{2\Omega^2}{mPT} \sum_{\alpha=1}^{M-1} \frac{1}{\nu_\alpha + \Omega^2},$$

where we denoted by ν_α eigenvalues of the matrix $(D^T D)_{ij}$ (see eq. (??) for the explicit expression for ν_α). We remind that the operator D_{ij} acts on the functions with boundary conditions (??). The expression for W_X^2 turns out to be divergent if we try to pass to the continuous limit, but the same singularity will appear in the evaluation of the mutual information (??) when we will express the mutual information through the average bandwidth W_X and the average power P to obtain finite result. In other words, the equation (??) plays the role of the renormalization condition.

First of all, let us consider averages from the eq. (??) which contain derivatives of the phase ϕ of the input signal. If we substitute $X(t) = \rho(t) e^{i\phi(t)}$ into the action (??), we get:

$$S[X] = \int_T dt \left(\frac{m|\dot{X}|^2}{2} + \frac{m\Omega^2 |X|^2}{2} \right) = \int_T dt \left(\frac{m}{2} (\dot{\rho}^2 + \rho^2 \dot{\phi}^2) + \frac{m\Omega^2 \rho^2}{2} \right). \tag{85}$$

We see that average of any expression linear in ϕ vanishes, because the action is quadratic in ϕ.

Now we have two terms which make up the first order correction in β to the mutual information:

$$\int_T \frac{dt}{T} \left\langle \frac{4\mu^3}{15(3+\mu^2)} \right\rangle_{P_X} \tag{86}$$

and

$$\int_T \frac{dt}{T} \langle f(\mu)\dot{\mu}^2 \rangle_{P_X}, \tag{87}$$

where we denoted

$$f(\mu) = \frac{\mu}{15(3+\mu^2)^3(9+4\mu^2)^2}(10206 + 21303\mu^2 + 15399\mu^4 + \tag{88}$$
$$4644\mu^6 + 496\mu^8).$$

The details of how we evaluated the integrals from the eqs. (??) and (??) are presented in the appendix C. Eventually, we obtain following results:

$$\int_T \frac{dt}{T} \langle f(\mu)\dot{\mu}^2 \rangle_{P_X} \approx (\gamma LP)^2 W_X^2 \int_0^{+\infty} dy\, y f(\gamma LP\xi y) \tag{89}$$
$$K_0\left(y/\mathrm{sh}\,2\xi\right) e^{-y\,\mathrm{cth}\,2\xi},$$

$$\int_T \frac{dt}{T} \left\langle \frac{4\mu^3}{15(3+\mu^2)} \right\rangle_{P_X} = \frac{2}{15}\xi^2(\gamma LP)^3 \int_0^{+\infty} dy \frac{y^3 K_0\left(y/\mathrm{sh}\,2\xi\right)}{3+(\gamma LP\xi)^2 y^2} e^{-y\,\mathrm{cth}\,2\xi},$$

where $K_0(x)$ is the modified Bessel function of the second kind. In the first integral in the eq. (??) we omitted terms which are proportional to $1/T^2$ instead of W_X^2, because we assume that $W_X^2 \gg 1/T^2$.

Substituting the averages (??) into the eq. (??) (??) obtain the mutual information with the first-order correction in β:

$$I_{P_X} = H_X - M_d \left\langle \int \frac{dt}{T} \ln \sqrt{1+\mu^2/3} \right\rangle_{P_X} + M_d \ln \left[\frac{\bar{\delta}_t e^{-1}}{\pi Q L} \right]$$
$$+M_d \beta L W_d^2 I_d(\bar{\gamma}) - M_d \beta L W_X^2 I_X(\bar{\gamma}), \tag{90}$$

where we denoted the dimensionless nonlinearity parameter as $\bar{\gamma} \equiv \gamma LP$. The functions I_d and I_X are defined as:

$$I_d(\bar{\gamma}) = \frac{1}{12} \int_T \frac{dt}{T} \left\langle \frac{4\mu^3}{15(3+\mu^2)} \right\rangle_{P_X}, \tag{91}$$

and

$$I_X(\bar{\gamma}) = \frac{1}{W_X^2} \int_T \frac{dt}{T} \langle f(\mu)\dot{\mu}^2 \rangle_{P_X}. \tag{92}$$

18

Now we wish to investigate the behaviour of the first-order correction to the mutual information as a function of the dimensionless nonlinearity parameter $\bar{\gamma}$ and the ratio W_d/W_X. For that goal, it is convenient to introduce some more notation, denoting the first-order correction divided by M_d as ΔI_{P_X}:

$$\Delta I_{P_X} = \beta L W_d^2 I_d(\bar{\gamma}) - \beta L W_X^2 I_X(\bar{\gamma}). \tag{93}$$

It is also convenient to introduce a dimensionless function $G(\bar{\gamma}, v)$, which we define by the following relation:

$$\Delta I_{P_X} = \beta L W_X^2 G(\bar{\gamma}, W_d/W_X), \tag{94}$$

so the function $G(\bar{\gamma}, v)$ reads:

$$G(\bar{\gamma}, v) = v^2 I_d(\bar{\gamma}) - I_X(\bar{\gamma}). \tag{95}$$

As we shall see shortly, the behaviour of $G(\bar{\gamma}, v)$ as a function of $\bar{\gamma}$ significantly depends on the value of v. However, we begin with demonstrating functions $I_d(\bar{\gamma})$ and $I_X(\bar{\gamma})$ separately in fig. 1 and fig. 2, respectively. To show the contributions each of these functions gives to the mutual information, we have to choose such values for the parameters β, L, W_d and W_X that the dimensionless parameters $\beta L W_d^2$ and $\beta L W_X^2$ are both small. Let us put, for demonstration, $\beta = 2 \cdot 10^{-23}\, s^2/km$, $L = 800\, km$, $W_X = 0.1\, GHz$, $W_d = 1\, GHz$. The corresponding dimensionless parameters of dispersion are $\beta L W_X^2 = 0.00016$ and $\beta L W_d^2 = 0.016$. The contributions of $\beta L W_d^2 I_d(\bar{\gamma})$ and $\beta L W_X^2 I_X(\bar{\gamma})$ are shown in the fig. 3. We also show the first-order correction ΔI_{P_X} in the fig. 4.
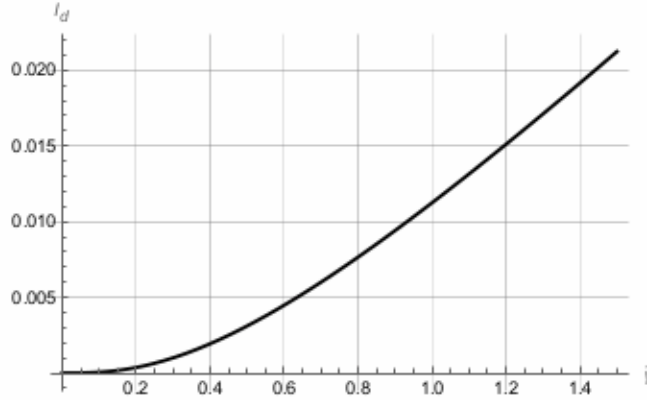

Figure 1: The function $I_d(\bar{\gamma})$ defined in the eq. (??).

In the fig. 4 we observe that the correction reaches a maximum value at a point around $\bar{\gamma} = 1.5$. However, such a point only exists when the ratio W_d/W_X belongs to the specific region, which can be obtained from a close inspection of the function $G(\bar{\gamma}, v)$. The maximum point disappears when the value of v
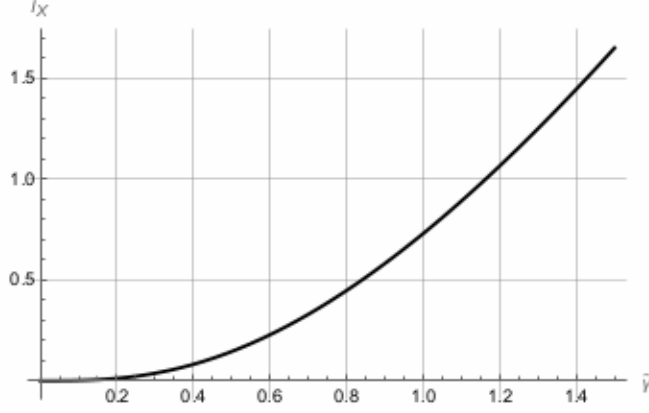
19

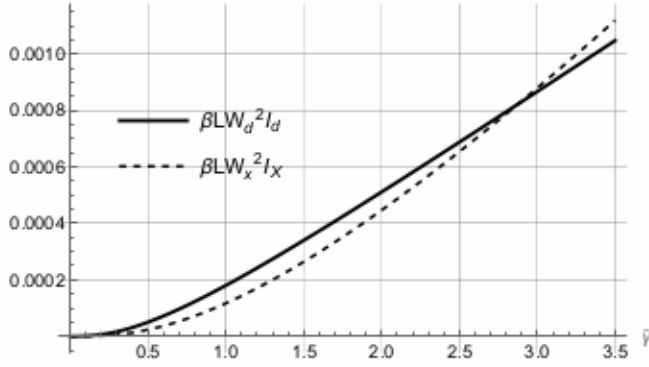Figure 2: The function $I_X(\bar{\gamma})$ defined in the eq. (??).

Figure 3: Two contributions to the mutual information divided by M_d. The solid line corresponds to the positive contribution which is proportional to the function $I_d(\bar{\gamma})$, and the dashed line corresponds to the negative contribution which is proportional to the function $I_X(\bar{\gamma})$.

becomes smaller than the number r_1, which is defined as:

$$r_1 = \sqrt{\left(\frac{I_X(y)}{y^2}\right)' \Big/ \left(\frac{I_d(y)}{y^3}\right)}\Bigg|_{y=0} \tag{96}$$

and which is roughly equal to 4.84. The maximum point also disappears when v becomes larger than $r_2 = \sqrt{186}$. We show all of the three cases in the fig. 5.

Eventually, we see that there are three different regimes of the behaviour of the function $G(\bar{\gamma}, v)$ depending on the value of $v = W_d/W_X$. When the ratio W_d/W_X is large (that is, $W_d/W_X > \sqrt{186}$), the function $G(\gamma LP, v)$ always grows if $\bar{\gamma}$ is increasing. Our interpretation is that for large bandwidth of the

of the harmonics of the output signal.

We remind that our results are valid for the intermediate power range, i.e. for such P that $N \ll P \ll (N\gamma^2 L^2)^{-1}$.

6   Conclusion

In the present paper we consider a channel described by the NLSE with additive Gaussian noise, arbitrary Kerr nonlinearity and small dimensionless second dispersion parameter. We also consider the signal-to-noise ratio (SNR) to be large. Using the path-integral approach and the saddle-point approximation we derive the conditional PDF $P[Y|X]$ in the leading order in 1/SNR. The expression for the PDF $P[Y|X]$ obtained in this way can be expanded in powers of the second dispersion β. We thus obtain the first-order correction in β to the conditional PDF in the eq. (??). To account for the finite bandwidth of the receiver W_d we integrate $P[Y|X]$ over high-frequency Fourier components of the output signal Y, which are not distinguished by the receiver, and denote the conditional PDF for the observable components of the output signal as $P_d[Y|X]$. For the perturbative expansion of $P_d[Y|X]$ to be valid, the bandwidth of the input signal W_X, the bandwidth of the receiver W_d and the second dispersion β must obey the conditions $\beta L W_X^2 \ll 1$, $\beta L W_d^2 \ll 1$ and $W_X \lesssim W_d$. Using the conditional PDF $P_d[Y|X]$ we calculate the conditional entropy, probability density functional for the output signal, entropy of the output signal and mutual information.

The expression for the mutual information can be used to find the optimal distribution of the input signal or to evaluate the transmission rate for some particular choice of the distribution of the input signal. We calculate the transmission rate for the Gaussian input signals with fixed average power and fixed average bandwidth W_X. We choose the input signal distribution in the form (??), which is essentially the case of two independent quantum harmonic oscillators evolving with imaginary time. Using well-known expressions for the correlators, we evaluate the averages in the expression for the mutual information (??).

The behaviour of the first-order correction in dimensionless dispersion parameter to the mutual information, which we considered as a function of the dimensionless nonlinearity parameter $\gamma L P$, turned to have three different regimes depending on the value of the ratio W_d/W_X. In the first case, when $W_d/W_X > \sqrt{186}$, the correction is positive for all the values of $\gamma L P$. In the second case, when W_d/W_X is less than the number $r_1 \approx 4.84$, which is defined in the eq. (??), the correction is negative for all the values of $\gamma L P$. The third regime is achieved when $r_1 < W_d/W_X < \sqrt{186}$. In the third case the correction to the mutual information grows for small values of $\gamma L P$, then reaches a maximum, and then falls for all values of $\gamma L P$ larger than the value of the maximum point.

7 Appendix A

In this section we present explicitly the cumbersome expressions one may meet trying to find the function Ψ_{cl} which makes the action functional from the eq. (??) stationary. We recall that we expanded the function Ψ_{cl} around the solution to the NLSE without noise:

$$\Psi_{cl} = \Phi + \kappa \exp\left\{i\mu\frac{z}{L} + i\phi\right\}, \qquad (97)$$

where $\mu = \gamma L |X|^2$ and $X = |X|e^{i\phi}$. We obtained the following problem for the first-order correction in β to the function $\kappa = \kappa_0 + \kappa_1 + \ldots$, where κ_0 is defined in eqs. (??):

$$\begin{cases} \partial_z^2 \kappa_1 - 2i\gamma\rho^2\kappa_1 - 4(\gamma\rho^2)^2\mathrm{Re}[\kappa_1] = F(z,t), \\ \kappa_1(z=0,t) = 0, \\ \kappa_1(z=L,t) = 0, \end{cases} \qquad (98)$$

where $\rho = |X|$. The cumbersome function $F(z,t)$ in the right-hand side reads:

$$\begin{aligned}
F(z,t) &= -\partial_z\delta L_1 - 2\gamma\rho^2\mathrm{Im}[\delta L_1] - i\beta e^{-i(z\gamma\rho^2+\phi)}\partial_t^2(e^{i(z\gamma\rho^2+\phi)}L_0) + \\
&\quad 2i\gamma\rho^2\left(2L_0\mathrm{Re}[\Phi_1 e^{-i(z\gamma\rho^2+\phi)}] - \overline{L}_0\Phi_1 e^{-i(z\gamma\rho^2+\phi)}\right), \\
\delta L_1 &= i\beta e^{-i(z\gamma\rho^2+\phi)}\partial_t^2(e^{i(z\gamma\rho^2+\phi)}\kappa_0) - \\
&\quad 2i\gamma\rho\left(2\kappa_0\mathrm{Re}[\Phi_1 e^{-i(z\gamma\rho^2+\phi)}] + \overline{\kappa}_0\Phi_1 e^{-i(z\gamma\rho^2+\phi)}\right), \\
L_0 &= \partial_z\kappa_0 - 2i\gamma\rho^2\mathrm{Re}[\kappa_0].
\end{aligned} \qquad (99)$$

The solution to the boundary problem (??) can be written using the Green's function method in the following form:

$$\kappa_1(z,t) = -\frac{1}{L}\int_0^L dz'\left[G_F(z,t;z')F(z',t) + G_{\bar{F}}(z,t;z')\bar{F}(z',t)\right], \qquad (100)$$

23

8 Appendix B

This section contains detailed calculation of the conditional PDF $\tilde{P}_d[Y_d|X]$ from the eq. (??). We start from the PDF from the eq. (??) which contains information about all of the harmonics of the output signal $Y(t)$ up to the first order β it reads:

$$P[Y|X] = \Lambda \exp\left\{-\frac{S_{cl}^{(0)}[\tilde{Y}]}{Q}\right\}\left(1 - \frac{S_{cl}^{(1)}[\tilde{Y}]}{Q}\right), \tag{115}$$

where we write the classical action $S_{cl} \equiv S[\Psi_{cl}]$ as a functional with the variable $\tilde{Y}(t) = y_1(t) + iy_2(t)$ defined in the eq. (??). The leading contribution in β to the action is diagonal in $y_{1,2}(t)$:

$$S_{cl}^{(0)}[\tilde{Y}] = \frac{1}{L}\int_T dt \left[y_1^2(t) + y_2^2(t)\right], \tag{116}$$

and the first-order correction can be written as:

$$S_{cl}^{(1)}[\tilde{Y}] = \beta \int_T dt\{b_1 y_1^2 + b_2 y_2^2 + b_3 y_1 y_2 + b_4 y_1\dot{y}_1 + b_5 y_2\dot{y}_2 +$$
$$b_6 y_1\dot{y}_2 + b_7\dot{y}_1 y_2 + b_8 y_1\ddot{y}_1 + b_9 y_2\ddot{y}_2 + b_{10} y_1\ddot{y}_2 + b_{11}\ddot{y}_1 y_2\}, \tag{117}$$

where the functions $b_i(t)$ can be expressed through $a_i(t)$ (see eqs. (??)-(??)) and $A_{ij}(t)$ (see eqs. (??)):

$$b_1(t) = a_4 A_{11}\dot{A}_{11} + a_6 A_{11}\dot{A}_{21} + a_7 A_{21}\dot{A}_{11} + a_5 A_{21}\dot{A}_{21} + a_8 A_{11}\ddot{A}_{11} +$$
$$a_{10} A_{11}\ddot{A}_{21} + a_{11} A_{21}\ddot{A}_{11} + a_9 A_{21}\ddot{A}_{21} + a_1 A_{11}^2 + a_3 A_{21}A_{11} + a_2 A_{21}^2, \tag{118}$$

$$b_2(t) = a_4 A_{12}\dot{A}_{12} + a_6 A_{12}\dot{A}_{22} + a_7 A_{22}\dot{A}_{12} + a_5 A_{22}\dot{A}_{22} + a_8 A_{12}\ddot{A}_{12} +$$
$$a_{10} A_{12}\ddot{A}_{22} + a_{11} A_{22}\ddot{A}_{12} + a_9 A_{22}\ddot{A}_{22} + a_1 A_{12}^2 + a_3 A_{22}A_{12} + a_2 A_{22}^2, \tag{119}$$

$$b_3(t) = a_4\left(A_{12}\dot{A}_{11} + A_{11}\dot{A}_{12}\right) + a_7\left(A_{22}\dot{A}_{11} + A_{21}\dot{A}_{12}\right) +$$
$$a_5\left(A_{22}\dot{A}_{21} + A_{21}\dot{A}_{22}\right) + a_6\left(A_{11}\dot{A}_{22} + A_{12}\dot{A}_{21}\right) + a_{10}\left(A_{12}\ddot{A}_{21} + A_{11}\ddot{A}_{22}\right) +$$
$$a_{11}\left(A_{22}\ddot{A}_{11} + A_{21}\ddot{A}_{12}\right) + a_8\left(A_{11}\ddot{A}_{12} + A_{12}\ddot{A}_{11}\right) + a_9\left(A_{22}\ddot{A}_{21} + A_{21}\ddot{A}_{22}\right) +$$
$$a_3\left(A_{21}A_{12} + A_{11}A_{22}\right) + 2a_2 A_{21}A_{22} + 2a_1 A_{11}A_{12}, \tag{120}$$

$$b_4(t) = 2a_8 A_{11}\dot{A}_{11} + 2a_{10} A_{11}\dot{A}_{21} + 2a_{11} A_{21}\dot{A}_{11} + 2a_9 A_{21}\dot{A}_{21} + a_4 A_{11}^2 +$$
$$a_6 A_{21}A_{11} + a_7 A_{21}A_{11} + a_5 A_{21}^2, \tag{121}$$

$$b_5(t) = 2a_8 A_{12}\dot{A}_{12} + 2a_{10} A_{12}\dot{A}_{22} + 2a_{11} A_{22}\dot{A}_{12} + 2a_9 A_{22}\dot{A}_{22} + a_4 A_{12}^2 +$$
$$a_6 A_{22}A_{12} + a_7 A_{22}A_{12} + a_5 A_{22}^2, \tag{122}$$

due to the discrete version of the Parseval's theorem:

$$\frac{1}{M} \sum_{n=0}^{M-1} \Psi(t_n)\bar{\Phi}(t_n) = \sum_{n'=0}^{M-1} \hat{\Psi}(\omega_{n'})\bar{\hat{\Phi}}(\omega_{n'}). \tag{134}$$

It is also worth mentioning that due to the functions $y_i(t)$ being real-valued, the Fourier-transform of $y_i(t)$ has the following property:

$$\bar{\hat{y}}_i(\omega) = \hat{y}_i(-\omega). \tag{135}$$

In the case of the first order correction $S_{cl}^{(1)}$, we will adopt the matrix notation:

$$S_{cl}^{(1)} = \beta M \delta_t \{ \hat{y}_1^\dagger \hat{b}_1 \hat{y}_1 + \hat{y}_2^\dagger \hat{b}_2 \hat{y}_2 + \hat{y}_1^\dagger \hat{b}_3 \hat{y}_2 + \hat{y}_1^\dagger \hat{b}_4 \hat{D}_1 \hat{y}_1 + \hat{y}_2^\dagger \hat{b}_5 \hat{D}_1 \hat{y}_2 +$$
$$\hat{y}_1^\dagger \hat{b}_6 \hat{D}_1 \hat{y}_2 + \hat{y}_2^\dagger \hat{b}_7 \hat{D}_1 \hat{y}_1 + \hat{y}_1^\dagger \hat{b}_8 \hat{D}_2 \hat{y}_1 + \hat{y}_2^\dagger \hat{b}_9 \hat{D}_2 \hat{y}_2 + \hat{y}_1^\dagger \hat{b}_{10} \hat{D}_2 \hat{y}_2 +$$
$$\hat{y}_2^\dagger \hat{b}_{11} \hat{D}_2 \hat{y}_1 \}, \tag{136}$$

where the components of the matrices $\hat{b}_i$ and vectors $\hat{y}_{1,2}$ are defined as:

$$(\hat{y}_{1,2})_n = \hat{y}_{1,2}(\omega_n), \tag{137}$$

$$(\hat{b}_i)_{nn'} = \hat{b}_i(\omega_n - \omega_{n'}). \tag{138}$$

The matrices $\hat{D}_1$ and $\hat{D}_2$ correspond to the operators of the first and the second derivatives, respectively. We write it in that way to emphasize that for now we do not restrict our consideration to a specific choice of the discretization scheme for the time derivatives. The first time derivative of a function $g(t)$ after discretization in this notation reads:

$$\dot{g}(t_n) = \sum_{m=0}^{M-1} (D_1)_{nm} g(t_m) = \sum_{m,n'=0}^{M-1} (\hat{D}_1)_{n'm} \hat{g}(\omega_m) e^{\frac{2\pi i n n'}{M}}, \tag{139}$$

and the second derivative is defined in a similar way by replacing D_1 by D_2.

Now that we developed the notation we can move to integrating the PDF from the eq. (??) over the high-frequency components of the vectors $\hat{y}_{1,2}$. According to the discretization of frequencies defined by the eq. (??), the vectors $\hat{y}_{1,2}$ can be divided into three parts:

$$\hat{y}_{1,2} = \begin{pmatrix} \vec{\varphi}_+^{1,2} \\ \vec{0} \\ \vec{\varphi}_-^{1,2} \end{pmatrix} + \begin{pmatrix} \vec{0} \\ \vec{L}_{1,2} \\ \vec{0} \end{pmatrix}. \tag{140}$$

The vectors $\vec{L}_{1,2}$ consist of the high-frequency harmonics ($|\omega| > \frac{W_t}{2}$) which we wish to get rid of by the integration (??). On the other hand, the vectors $\vec{\phi}_\pm^{1,2}$ consist of the observable harmonics with low frequencies ($|\omega| < \frac{W_t}{2}$). Due to the diagonal form of the functional $S_{cl}^{(0)}$:

$$S_{cl}^{(0)} = \frac{M\delta_t}{L} \left[\vec{\chi}_1^\dagger \vec{\chi}_1 + \vec{\chi}_2^\dagger \vec{\chi}_2 \right] + \frac{M\delta_t}{L} \left[\vec{L}_1^\dagger \vec{L}_1 + \vec{L}_2^\dagger \vec{L}_2 \right], \tag{141}$$

28

the integration over the components of the vectors $\vec{L}_{1,2}$ can be done using the following formulas for Gaussian integrals:

$$\int D\vec{L}_1 D\vec{L}_2 \exp\left\{-\frac{M\delta_t}{QL}\left[\vec{L}_1^\dagger \vec{L}_1 + \vec{L}_2^\dagger \vec{L}_2\right]\right\} = \left(\frac{\pi QL}{M\delta_t}\right)^{M-M_d}, \qquad (142)$$

$$\int D\vec{L}_1 D\vec{L}_2 \vec{L}_\alpha^\dagger \hat{A} \vec{L}_\beta \exp\left\{-\frac{M\delta_t}{QL}\left[\vec{L}_1^\dagger \vec{L}_1 + \vec{L}_2^\dagger \vec{L}_2\right]\right\} =$$

$$= \left(\frac{\pi QL}{M\delta_t}\right)^{M-M_d} \frac{QL}{M\delta_t} Tr[\hat{A}]\delta_{\alpha\beta}, \qquad (143)$$

where in the integration measure $D\vec{L}_i = \prod_n d\mathrm{Re}\hat{y}_i(\omega_n)d\mathrm{Im}\hat{y}_i(\omega_n)$ the product is taken over such frequencies that $|\omega_n| > W_d/2$. We denoted by M_d the amount of the frequencies ω_n which correspond to the observable low-frequency harmonics. These harmonics are contained in the vectors $\vec{\chi}_{1,2}$:

$$\vec{\chi}_{1,2} = \begin{pmatrix} \vec{\varphi}_+^{1,2} \\ \vec{\varphi}_-^{1,2} \end{pmatrix}. \qquad (144)$$

The result of the integration can be written as follows:

$$\tilde{P}_d[\tilde{Y}_d|X] \equiv \int D\vec{L}_1 D\vec{L}_2 \tilde{P}[\tilde{Y}(t)|X] =$$

$$\Lambda \left(\frac{\pi QL}{M\delta_t}\right)^{M-M_d} \exp\left\{-\frac{M\delta_t}{QL}\left[\vec{\chi}_1^\dagger \vec{\chi}_1 + \vec{\chi}_2^\dagger \vec{\chi}_2\right]\right\} \times$$

$$\left(1 - \frac{\beta M\delta_t}{Q}\left[\vec{\chi}_1^\dagger \hat{b}_1^d \vec{\chi}_1 + \vec{\chi}_2^\dagger \hat{b}_2^d \vec{\chi}_2 + \vec{\chi}_1^\dagger \hat{b}_3^d \vec{\chi}_2 + \vec{\chi}_1^\dagger \hat{b}_4^d \hat{d}_1 \vec{\chi}_1 + \right.\right.$$

$$\vec{\chi}_2^\dagger \hat{b}_5^d \hat{d}_1 \vec{\chi}_2 + \vec{\chi}_1^\dagger \hat{b}_6^d \hat{d}_1 \vec{\chi}_2 + \vec{\chi}_2^\dagger \hat{b}_7^d \hat{d}_1 \vec{\chi}_1 + \vec{\chi}_1^\dagger \hat{b}_8^d \hat{d}_2 \vec{\chi}_1 + \vec{\chi}_2^\dagger \hat{b}_9^d \hat{d}_2 \vec{\chi}_2 +$$

$$\left. \vec{\chi}_1^\dagger \hat{b}_{10}^d \hat{d}_2 \vec{\chi}_2 + \vec{\chi}_2^\dagger \hat{b}_{11}^d \hat{d}_2 \vec{\chi}_1\right] - \beta L Tr\left[\hat{b}_1^L + \hat{b}_2^L + \hat{b}_4^L \hat{D}_1^L + \right.$$

$$\left.\left. \hat{b}_5^L \hat{D}_1^L + \hat{b}_8^L \hat{D}_2^L + \hat{b}_9^L \hat{D}_2^L\right]\right). \qquad (145)$$

In the eq. (??) we have introduced some more notation with matrices $\hat{b}_i^{d,L}, \hat{d}_{1,2}$ and $\hat{D}_{1,2}^L$ to describe what happens to the matrices $\hat{b}_i$ and $\hat{D}_{1,2}$ after we divide vectors $\hat{y}_{1,2}$ into low- and high-frequency parts according to the eq. (??). To explain the definition of the new matrices, let us first write $\hat{b}_i$ and $\hat{D}_{1,2}$ as the following block matrices:

$$\hat{b}_i = \begin{pmatrix} (\hat{b}_i)_{11} & (\hat{b}_i)_{12} & (\hat{b}_i)_{13} \\ (\hat{b}_i)_{21} & (\hat{b}_i)_{22} & (\hat{b}_i)_{23} \\ (\hat{b}_i)_{31} & (\hat{b}_i)_{32} & (\hat{b}_i)_{33} \end{pmatrix}, \qquad (146)$$

$$\hat{D}_{1,2} = \begin{pmatrix} (\hat{D}_{1,2})_{11} & (\hat{D}_{1,2})_{12} & (\hat{D}_{1,2})_{13} \\ (\hat{D}_{1,2})_{21} & (\hat{D}_{1,2})_{22} & (\hat{D}_{1,2})_{23} \\ (\hat{D}_{1,2})_{31} & (\hat{D}_{1,2})_{32} & (\hat{D}_{1,2})_{33} \end{pmatrix}. \qquad (147)$$

29

Using the block representation from the eqs. (??) and (??), the scalar products which one meets computing $\tilde{P}_d[\tilde{Y}_d|X]$ can be written as follows:

$$\left(\vec{\varphi}_+^{1,2}, \vec{0}, \vec{\varphi}_-^{1,2}\right)^\dagger \hat{b}_i \hat{D}_{1,2} \begin{pmatrix} \vec{\varphi}_+^{1,2} \\ \vec{0} \\ \vec{\varphi}_-^{1,2} \end{pmatrix} \approx \vec{\chi}_{1,2}^\dagger \hat{b}_i^d \hat{d}_{1,2} \vec{\chi}_{1,2}, \tag{148}$$

$$\left(\vec{0}, \vec{L}_{1,2}, \vec{0}\right)^\dagger \hat{b}_i \hat{D}_{1,2} \begin{pmatrix} \vec{0} \\ \vec{L}_{1,2} \\ \vec{0} \end{pmatrix} = \vec{L}_{1,2}^\dagger \hat{b}_i^L \hat{D}_{1,2}^L \vec{L}_{1,2}, \tag{149}$$

where the new matrices are:

$$\hat{b}_i^d = \begin{pmatrix} (\hat{b}_i)_{11} & (\hat{b}_i)_{13} \\ (\hat{b}_i)_{31} & (\hat{b}_i)_{33} \end{pmatrix}, \quad \hat{d}_{1,2} = \begin{pmatrix} (\hat{D}_{1,2})_{11} & (\hat{D}_{1,2})_{13} \\ (\hat{D}_{1,2})_{31} & (\hat{D}_{1,2})_{33} \end{pmatrix},$$

$$\hat{L}_i^d = (\hat{b}_i)_{22}, \quad \hat{D}_{1,2}^L = (\hat{D}_{1,2})_{22}. \tag{150}$$

Thus, the matrices $\hat{d}_{1,2}$ and $\hat{D}_{1,2}^L$ correspond to the low- and high-frequency parts of the derivative operators $\hat{D}_{1,2}$ respectively. The same goes for the matrices $\hat{b}_i^L$ and $\hat{b}_i^d$ - the superscript L means that the matrix $\hat{b}_i^L$ consists of such components $(\hat{b}_i)_{nm}$ that the corresponding frequencies $\omega_{n,m}$ are large: $|\omega_n| > W_d/2$ and $|\omega_m| > W_d/2$. In turn, the matrix $\hat{b}_i^d$ consists of low-frequency components of $(\hat{b}_i)_{nm}$. We wrote approximate equality in the eq. (??) because in that step we neglected contribution of the blocks $(\hat{b}_i)_{12}$ and $(\hat{b}_i)_{32}$. These blocks contain high-frequency Fourier harmonics of the functions $b_i(t)$. But we consider slow functions $b_i(t)$ (which means that $\beta L W_X^2 \ll 1$), so the high-frequency harmonics are much smaller than the low-frequency ones.

What is left now is finding the normalization factor Λ, which we will do by using the normalization condition for the PDF $\tilde{P}_d[\tilde{Y}_d|X]$:

$$\int D\tilde{Y}_d \tilde{P}_d[\tilde{Y}_d|X] \equiv \int D\vec{\chi}_1 D\vec{\chi}_2 \tilde{P}_d[\tilde{Y}_d|X] = 1. \tag{151}$$

The integrals in the eq. (??) are Gaussian ones and can be calculated using the formulas for the Gaussian integrals (see eqs. (??) and (??)) again. Splitting the normalization factor into the leading contribution ($\Lambda^{(0)}$) and the first correction in β ($\Lambda^{(1)}$) we get for each order:

$$\Lambda^{(0)} = \left(\frac{\pi Q L}{M \delta_t}\right)^{-M} \tag{152}$$

and

$$\begin{aligned} \frac{\Lambda^{(1)}}{\Lambda^{(0)}} &= \beta L Tr\left[\hat{b}_1^L + \hat{b}_2^L + \hat{b}_4^L \hat{D}_1^L + \hat{b}_5^L \hat{D}_1^L + \right. \\ &\quad \left. \hat{b}_8^L \hat{D}_2^L + \hat{b}_9^L \hat{D}_2^L\right] + \beta L Tr\left[\hat{b}_1^d + \hat{b}_2^d + \hat{b}_4^d \hat{d}_1 + \right. \\ &\quad \left. \hat{b}_5^d \hat{d}_1 + \hat{b}_8^d \hat{d}_2 + \hat{b}_9^d \hat{d}_2\right]. \end{aligned} \tag{153}$$

After substituting contributions to the normalization factor from the eqs. (??) and (??) into the eq. (??), contribution of the high-frequency parts of the restric and $\hat{D}_{1,2}$ vanishes and we can write the conditional PDF in the following way:

$$\tilde{P}_d[\tilde{Y}_d(\omega)|X] = \bar{\Lambda}_{d,\omega}^{(0)} \exp\left\{-\frac{M\delta_t}{QL}[\vec{\chi}_1^\dagger \vec{\chi}_1 + \vec{\chi}_2^\dagger \vec{\chi}_2]\right\}\left(1 + \frac{\bar{\Lambda}_{d,\omega}^{(1)}}{\bar{\Lambda}_{d,\omega}^{(0)}} - \right.$$
$$\frac{\beta M \delta_t}{Q}\left[\vec{\chi}_1^\dagger \hat{b}_1^d \vec{\chi}_1 + \vec{\chi}_2^\dagger \hat{b}_2^d \vec{\chi}_2 + \vec{\chi}_1^\dagger \hat{b}_3^d \vec{\chi}_2 + \vec{\chi}_1^\dagger \hat{b}_4^d \hat{d}_1 \vec{\chi}_1 + \vec{\chi}_2^\dagger \hat{b}_5^d \hat{d}_1 \vec{\chi}_2 + \right.$$
$$\vec{\chi}_1^\dagger \hat{b}_6^d \hat{d}_1 \vec{\chi}_2 + \vec{\chi}_2^\dagger \hat{b}_7^d \hat{d}_1 \vec{\chi}_1 + \vec{\chi}_1^\dagger \hat{b}_8^d \hat{d}_2 \vec{\chi}_1 + \vec{\chi}_2^\dagger \hat{b}_9^d \hat{d}_2 \vec{\chi}_2 + \vec{\chi}_1^\dagger \hat{b}_{10}^d \hat{d}_2 \vec{\chi}_2 + $$
$$\left.\left.\vec{\chi}_2^\dagger \hat{b}_{11}^d \hat{d}_2 \vec{\chi}_1\right]\right), \tag{154}$$

where we defined the normalization factor $\bar{\Lambda}_{d,\omega} = \bar{\Lambda}_{d,\omega}^{(0)} + \bar{\Lambda}_{d,\omega}^{(1)}$ of the new PDF $\tilde{P}_d[\tilde{Y}_d(\omega)|X]$ as:

$$\bar{\Lambda}_{d,\omega}^{(0)} = \left(\frac{\pi QL}{M\delta_t}\right)^{-M_d}, \tag{155}$$

$$\frac{\bar{\Lambda}_{d,\omega}^{(1)}}{\bar{\Lambda}_{d,\omega}^{(0)}} = \beta L Tr\left[\hat{b}_1^d + \hat{b}_2^d + \hat{b}_4^d \hat{d}_1 + \hat{b}_5^d \hat{d}_1 + \hat{b}_8^d \hat{d}_2 + \hat{b}_9^d \hat{d}_2\right]. \tag{156}$$

We wrote the argument of the PDF in the eq. (??) as $\tilde{Y}_d(\omega)$ and added the subscript ω to the normalization factor in order to specify that the PDF $\tilde{P}_d[\tilde{Y}_d(\omega)|X]$ is normalized by the condition (??), where we integrated over the low-frequency Fourier harmonics of the functions $y_{1,2}(t)$. We can move to the time domain by introducing the DFT on a coarse grid with the number of points M_d

$$\hat{y}_i(\omega_{n'}) = \frac{1}{M_d}\sum_{n=0}^{M_d-1} y_i(t_n)e^{-\frac{2\pi i n n'}{M_d}}. \tag{157}$$

The discretization step $\bar{\delta}_t$ of the coarse grid is defined by the following relation:

$$T = M\delta_t = M_d \bar{\delta}_t. \tag{158}$$

After substituting Fourier-harmonics from the eq. (??) in the eq. (??), we get the conditional PDF in terms of the functions $y_{1,2}(t)$ defined on the coarse grid:

$$\tilde{P}_d[\tilde{Y}_d(\omega)|X] = \bar{\Lambda}_{d,\omega}^{(0)}\left(1 + \frac{\bar{\Lambda}_{d,\omega}^{(1)}}{\bar{\Lambda}_{d,\omega}^{(0)}}\right)\exp\left\{-\frac{S_{coarse}^{(0)} + S_{coarse}^{(1)}}{Q}\right\}, \tag{159}$$

where the functionals $S_{coarse}^{(0),(1)}$ are expressed through the values of the functions

$b_i(t)$ and $y_{1,2}(t)$ on the coarse grid:

$$S_{coarse}^{(0)} = \frac{\bar{\delta}_t}{L} \sum_{j=0}^{M_d-1} \left[y_1^2(t_j) + y_2^2(t_j) \right], \tag{160}$$

$$S_{coarse}^{(1)} = \beta \bar{\delta}_t \sum_{j=0}^{M_d-1} \{ b_{1,j} y_{1,j}^2 + b_{2,j} y_{2,j}^2 + b_{3,j} y_{1,j} y_{2,j} + b_{4,j} y_{1,j} (\dot{y}_1)_j +$$

$$b_{5,j} y_{2,j} (\dot{y}_2)_j + b_{6,j} y_{1,j} (\dot{y}_2)_j + b_{7,j} (\dot{y}_1)_j y_{2,j} + b_{8,j} y_{1,j} (\ddot{y}_1)_j +$$

$$b_{9,j} y_{2,j} (\ddot{y}_2)_j + b_{10,j} y_{1,j} (\ddot{y}_2)_j + b_{11,j} (\ddot{y}_1)_j y_{2,j} \}. \tag{161}$$

However, the PDF from the eq. (??) is still normalized with respect to the Fourier harmonics of the functions $y_{1,2}(t)$. If we want our variables to be the values of the functions $y_{1,2}(t)$ on the coarse time grid, we also need to multiply the PDF $\tilde{P}_d[\tilde{Y}_d(\omega)|X]$ by the Jacobian determinant of the substitution (??):

$$\tilde{P}_d[\tilde{Y}_d(t)|X] = \left| \frac{\partial(\hat{y}_1(\omega), \hat{y}_2(\omega))}{\partial(y_1(t), y_2(t))} \right| \tilde{P}_d[\tilde{Y}(\omega)|X] = M_d^{-M_d} \tilde{P}_d[\tilde{Y}(\omega)|X]. \tag{162}$$

Thus, we obtain the conditional PDF for the variables $y_{1,2}(t_i)$:

$$\tilde{P}_d[\tilde{Y}_d(t)|X] = \bar{\Lambda}_d^{(0)} \left(1 + \frac{\bar{\Lambda}_d^{(1)}}{\bar{\Lambda}_d^{(0)}} \right) \exp \left\{ -\frac{S_{coarse}^{(0)} + S_{coarse}^{(1)}}{Q} \right\}, \tag{163}$$

where the normalization factor reads:

$$\bar{\Lambda}_d^{(0)} = \left(\frac{\bar{\delta}_t}{\pi Q L} \right)^{M_d}, \quad \frac{\bar{\Lambda}_d^{(1)}}{\bar{\Lambda}_d^{(0)}} = \frac{\bar{\Lambda}_{d,\omega}^{(1)}}{\bar{\Lambda}_{d,\omega}^{(0)}}. \tag{164}$$

The first correction to the normalization factor from the eq. (??) can be expressed through the functions $b_i(t)$. We recall that the matrix components $(\hat{b}_i)_{nn'}$ were defined in the eq. (??) as the Fourier transform of the function $b_i(t)$ evaluated at the frequency $\omega_n - \omega_{n'}$. Therefore we can immediately express the three terms which do not contain derivative operators $(\hat{b}_{1,2,3}^d)$:

$$Tr\left[\hat{b}_i^d \right] = \sum_{n=n_d} (\hat{b}_i)_{nn} = M_d \hat{b}_i(\omega = 0) = \frac{M_d}{M \delta_t} \sum_{n=0}^{M-1} b_i(t_n) \delta_t \xrightarrow[\delta_t \to 0]{}$$

$$M_d \int_T b_i(t) \frac{dt}{T}, \tag{165}$$

where we denoted by $\sum_{n=n_d}$ the summation over such indices n which correspond to the low frequencies $|\omega_n| \leq W_d/2$. Due to the fact that the trace is taken over the low-frequency indices, we can replace the elements of the matrices $\hat{d}_{1,2}$ with the elements of the derivative operator in the continuous case.

Therefore, traces from the eq. (??) reduce to:

$$Tr\left[\hat{b}_i^d \hat{d}_1\right] \rightarrow \frac{M_d}{W_d}\int_T b_i(t)dt \times \int_{-\frac{W_d}{2}}^{\frac{W_d}{2}} i\omega d\omega = 0, \tag{166}$$

$$Tr\left[\hat{b}_i^d \hat{d}_2\right] \rightarrow -\frac{M_d}{W_d}\int_T b_i(t)dt \times \int_{-\frac{W_d}{2}}^{\frac{W_d}{2}} \omega^2 d\omega = -M_d\frac{W_d^2}{12}\int_T b_i(t)\frac{dt}{T}. \tag{167}$$

Explicit expressions for the functions $b_i(t)$ are rather cumbersome. Fortunately, because of the eq. (??), we only need to know expressions for $b_1(t) + b_2(t)$ and $b_8(t) + b_9(t)$ in order to get the normalization factor. Expressions for these quantities are relatively simple:

$$b_1(t) + b_2(t) = -\frac{\mu\dot{\mu}^2}{15(3+\mu^2)^3(9+4\mu^2)^2}(10206 + 21303\mu^2 + 15399\mu^4 +$$

$$4644\mu^6 + 496\mu^8) - \frac{4\mu\dot{\mu}\dot{\phi}}{3+\mu^2} - \frac{2(3+2\mu^2)\ddot{\phi}}{3+\mu^2},$$

$$b_8(t) + b_9(t) = -\frac{4\mu^3}{15(3+\mu^2)}. \tag{168}$$

Finally, with the eq. (??) we get:

$$\frac{\tilde{\Lambda}_d^{(1)}}{\tilde{\Lambda}_d^{(0)}} = M_d\frac{\beta L W_d^2}{12}\int_T \frac{4\mu^3}{15(3+\mu^2)}\frac{dt}{T} - M_d\beta L\int\left[\frac{4\mu\dot{\mu}\dot{\phi}}{3+\mu^2} +\right.$$

$$\frac{2(3+2\mu^2)\ddot{\phi}}{3+\mu^2} + \frac{\mu\dot{\mu}^2}{15(3+\mu^2)^3(9+4\mu^2)^2}(10206 + 21303\mu^2 +$$

$$\left.15399\mu^4 + 4644\mu^6 + 496\mu^8)\right]\frac{dt}{T}. \tag{169}$$

Averaging the eq. (??) with respect to the distribution of the input signal will give two contributions: the one proportional to the dimensionless parameter $\beta L W_d^2$ and the one proportional to $\beta L W_X^2$ (we get W_X^2 from averaging derivatives of the input signal).

9   Appendix C

In this section we explain in detail how we evaluated two averages from the section 5:

$$\int_T \frac{dt}{T}\langle f(\mu)\dot{\mu}^2\rangle_{P_X} = \int_T \frac{dt}{T}\int DX P_X[X]f(\mu)\dot{\mu}^2, \tag{170}$$

$$\int_T \frac{dt}{T}\left\langle\frac{4\mu^3}{15(3+\mu^2)}\right\rangle_{P_X} = \int_T \frac{dt}{T}\int DX P_X[X]\frac{4\mu^3}{15(3+\mu^2)}.$$

In what follows we will work with real and imaginary parts of the input signal $X(t) = a(t) + ib(t)$, where a and b are the real functions of time. In the action $S[X]$ we have two independent oscillators in imaginary time with the same frequency:

$$S[X] = \int_T dt \left(\frac{m|\dot{X}|^2}{2} + \frac{m\Omega^2|X|^2}{2} \right) = \int_T dt \left(\frac{m\dot{a}^2}{2} + \frac{m\Omega^2 a^2}{2} \right)$$
$$+ \int_T dt \left(\frac{m\dot{b}^2}{2} + \frac{m\Omega^2 b^2}{2} \right). \tag{171}$$

Two-point correlation functions for the action (??) with the boundary conditions $X(T/2) = X(-T/2) = 0$ are well-known [?], and we will use it evaluating the averages from the eq. (??):

$$\langle a(t)a(s) \rangle_{P_X} = \langle b(t)b(s) \rangle_{P_X} = \begin{cases} \dfrac{\operatorname{sh}\Omega(T/2+s)\operatorname{sh}\Omega(T/2-t)}{m\Omega\operatorname{sh}\Omega T}, & t > s \\[3mm] \dfrac{\operatorname{sh}\Omega(T/2+t)\operatorname{sh}\Omega(T/2-s)}{m\Omega\operatorname{sh}\Omega T}, & t < s \end{cases},$$
$$\langle a(t)b(s) \rangle_{P_X} = 0. \tag{172}$$

To evaluate path-integrals from the eq. (??) we introduce time discretization and the discrete version of the time derivative:

$$\dot{g}_i = \sum_{j=0}^{M} D_{ij} g_j = \frac{1}{\delta_t} \left(g_{i+1} - g_i \right), \, 0 \le i \le M - 1, \tag{173}$$

where $g_i = g(t_i)$, $\dot{g}_i = \dot{g}(t_i)$ and t_i is a point of the time grid with the discretization step δ_t: $t_i = -T/2 + i\delta_t$. Due to the boundary condition $X(-T/2) = X(T/2) = 0$ we are considering functions $g(t)$ which vanish at the endpoints, i.e. $g_M = g_0 = 0$.

We can write action from the eq. (??) on the discretized time axis as:

$$S[X] = \frac{1}{2} \sum_{j,k=1}^{M-1} a_j Q_{jk} a_k + \frac{1}{2} \sum_{j,k=0}^{M-1} b_j Q_{jk} b_k, \tag{174}$$

$$Q_{jk} = m\delta_t \sum_{i=0}^{M-1} D_{ij} D_{ik} + m\Omega^2 \delta_t \delta_{jk}, \tag{175}$$

where δ_t is the discretization step. In terms of the discretized action the correlators from the eq. (??) are kernels of the integral operator Q^{-1}:

$$\langle a(t_i)a(t_j) \rangle_{P_X} = \langle b(t_i)b(t_j) \rangle_{P_X} = (Q^{-1})_{ij}. \tag{176}$$

Knowledge of the correlators from the eq. (??) and of the action from the eq. (??) is enough to evaluate average of the form $\langle f(\mu)\dot{\mu}^2 \rangle_{P_X}$. First we recall how μ is related to the input signal: $\mu = \gamma L(a^2 + b^2)$, so $\dot{\mu}^2 = 4(\gamma L)^2(a^2\dot{a}^2 + b^2\dot{b}^2 + 2ab\dot{a}\dot{b})$, and we have three terms:

$$\langle f(\mu)\dot{\mu}^2 \rangle_{P_X} = 4(\gamma L)^2 \langle f(\mu)a^2\dot{a}^2 \rangle_{P_X} + 4(\gamma L)^2 \langle f(\mu)b^2\dot{b}^2 \rangle_{P_X} + \\ 8(\gamma L)^2 \langle f(\mu)ab\dot{a}\dot{b} \rangle_{P_X}. \tag{177}$$

The first one reads:

$$\langle f(a_i, b_i)a_i^2\dot{a}_i^2 \rangle_{P_X} = \sum_{j,k=1}^{M-1} D_{ij}D_{ik}\langle f(a_i, b_i)a_i^2 a_j a_k \rangle_{P_X}, \tag{178}$$

where we have moved to the discrete notation. In the eq. (??) we do not imply summation over the repeated index i, and in what follows we always write sums explicitly. From the eq. (??) we proceed by adding two additional integrals over delta-functions, which allows us to extract the function $f(a_i, b_i)$ from the average:

$$\langle f(a_i, b_i)a_i^2 a_j a_k \rangle_{P_X} = \int dx\,dy\,f(x,y)x^2 \langle \delta(x - a_i)\delta(y - b_i)a_j a_k \rangle_{P_X}. \tag{179}$$

Next we use integral representation of the delta-function:

$$\langle \delta(x - a_i)\delta(y - b_i)a_j a_k \rangle_{P_X} = \int \frac{dk_x dk_y}{(2\pi)^2} e^{ik_x x + ik_y y} \langle e^{-ik_x a_i - ik_y b_i} a_j a_k \rangle_{P_X}. \tag{180}$$

Recalling the definition of the brackets $\langle ... \rangle_{P_X}$, we see that in case of the quadratic action from the eq. (??) the average from the eq. (??) can be calculated explicitly. Namely, we change variables in the following integral:

$$\langle e^{-ik_x a_i - ik_y b_i} a_j a_k \rangle_{P_X} = \Lambda_X \int \left(\prod_n da_n db_n \right) a_j a_k e^{-S[X] - ik_x a_i - ik_y b_i} \tag{181}$$

from a_j and b_j to $\alpha_j = a_j + ik_x(Q^{-1})_{ii}\delta_{ij}$ and $\beta_j = b_j + ik_y(Q^{-1})_{ii}\delta_{ij}$. After such substitution the integral from the eq. (??) reads:

$$\Lambda_X \int \left(\prod_n d\alpha_n d\beta_n \right) \left[\alpha_j \alpha_k - k_x^2(Q^{-1})_{ji}(Q^{-1})_{ki} \right] e^{-S[\alpha,\beta] - \frac{1}{2}(Q^{-1})_{ii}(k_x^2 + k_y^2)}, \tag{182}$$

where $S[\alpha, \beta] = \frac{1}{2}\sum_{i,j=1}^{M-1} \alpha_i Q_{ij}\alpha_j + \frac{1}{2}\sum_{i,j=1}^{M-1} \beta_i Q_{ij}\beta_j$. Performing the Gaussian integration over α_n and β_n, we get:

$$\langle e^{-ik_x a_i - ik_y b_i} a_j a_k \rangle_{P_X} = \left[(Q^{-1})_{jk} - k_x^2(Q^{-1})_{ji}(Q^{-1})_{ki} \right] e^{-\frac{1}{2}(Q^{-1})_{ii}(k_x^2 + k_y^2)}, \tag{183}$$

where we also have used the expression for the normalization constant $\Lambda_X = \det\left(\frac{Q}{2\pi}\right)$. Now we return to the eq. (??) and integrate over k_x and k_y to get the following expression for the average $\langle \delta(x - a_i)\delta(y - b_i)a_j a_k \rangle_{P_X}$:

$$\frac{1}{2\pi(Q^{-1})_{ii}} \left[(Q^{-1})_{jk} - \frac{(Q^{-1})_{ji}(Q^{-1})_{ki}}{(Q^{-1})_{ii}} + x^2\frac{(Q^{-1})_{ji}(Q^{-1})_{ki}}{(Q^{-1})_{ii}^2} \right] e^{-\frac{x^2 + y^2}{2(Q^{-1})_{ii}}}. \tag{184}$$

Finally, we arrive at the following result for one of the three terms which contribute to $\langle f(\mu)\dot{\mu}^2\rangle_{P_X}$:

$$\langle f(a_i, b_i)a_i^2 a_j a_k\rangle_{P_X} = \frac{1}{2\pi(Q^{-1})_{ii}} \int dx\, dy\, f(x,y) x^2 e^{-\frac{x^2+y^2}{2(Q^{-1})_{ii}}}$$

$$\left[(Q^{-1})_{jk} - \frac{(Q^{-1})_{ji}(Q^{-1})_{ki}}{(Q^{-1})_{ii}} + x^2 \frac{(Q^{-1})_{ji}(Q^{-1})_{ki}}{(Q^{-1})_{ii}^2}\right]. \tag{185}$$

Evaluation of the other two terms ($\langle f(a,b)b^2\dot{b}^2\rangle_{P_X}$ and $2\langle f(a,b)ab\dot{a}\dot{b}\rangle_{P_X}$) can be done in the exactly same way by extracting the function $f(a,b)$ from the average sign by the virtue of introducing additional delta-functions. Moreover, we can get the answer for $\langle f(a_i, b_i)b_i^2 b_j b_k\rangle_{P_X}$ by replacing x with y in the eq. (??):

$$\langle f(a_i, b_i)b_i^2 b_j b_k\rangle_{P_X} = \frac{1}{2\pi(Q^{-1})_{ii}} \int dx\, dy\, f(x,y) y^2 e^{-\frac{x^2+y^2}{2(Q^{-1})_{ii}}}$$

$$\left[(Q^{-1})_{jk} - \frac{(Q^{-1})_{ji}(Q^{-1})_{ki}}{(Q^{-1})_{ii}} + y^2 \frac{(Q^{-1})_{ji}(Q^{-1})_{ki}}{(Q^{-1})_{ii}^2}\right]. \tag{186}$$

Calculation of the last term $\langle f(a_i, b_i)a_i b_i a_j b_k\rangle_{P_X}$ gives us:

$$\langle f(a_i, b_i)a_i b_i a_j b_k\rangle_{P_X} = \frac{1}{2\pi(Q^{-1})_{ii}} \int dx\, dy\, f(x,y) x^2 y^2 e^{-\frac{x^2+y^2}{2(Q^{-1})_{ii}}}$$

$$\frac{(Q^{-1})_{ji}(Q^{-1})_{ki}}{(Q^{-1})_{ii}^2}. \tag{187}$$

Now we combine all the three terms and restore the derivative operators D_{ij}, which we were omitting, to obtain $\langle f(\mu)\dot{\mu}^2\rangle_{P_X}$:

$$\langle f(\mu_i)\dot{\mu}_i^2\rangle_{P_X} = \sum_{j,k=1}^{M-1} \frac{4(\gamma L)^2}{2\pi(Q^{-1})_{ii}} \int dx\, dy\, f(x,y)(x^2+y^2) e^{-\frac{x^2+y^2}{2(Q^{-1})_{ii}}} D_{ij} D_{ik}$$

$$\left[(Q^{-1})_{jk} - \frac{(Q^{-1})_{ji}(Q^{-1})_{ki}}{(Q^{-1})_{ii}} + (x^2+y^2) \frac{(Q^{-1})_{ji}(Q^{-1})_{ki}}{(Q^{-1})_{ii}^2}\right]. \tag{188}$$

We were able to express the average $\langle f(\mu)\dot{\mu}^2\rangle_{P_X}$ through the matrix Q^{-1}, which, as we recall from the eq. (??), is essentially the two-point correlation function of our input signal. Next we should take the continuous limit of the expression (??), which amounts to the computation of derivatives of the correlation functions from the eq. (??). However, if we try to evaluate the second derivative of the correlator at equal times (the term which contains $\sum_{j,k} D_{ij} D_{ik}(Q^{-1})_{jk}$), we will get a divergent result. It can be seen from the continuous expression for the second derivative of the correlator:

$$\frac{d}{ds}\frac{d}{dt}\langle a(t)a(s)\rangle_{P_X} = \frac{\delta(t-s)}{m} -$$

$$\frac{\Omega}{m\,\mathrm{sh}\,\Omega T}\Big(\mathrm{ch}\,\Omega(T/2+s)\,\mathrm{ch}\,\Omega(T/2-t) + \mathrm{ch}\,\Omega(T/2+t)\,\mathrm{ch}\,\Omega(T/2-s)\Big). \tag{189}$$

36

To get the continuous limit of $D_{ij}D_{ik}(Q^{-1})_{jk}$ we are supposed to put $t = s$ in the eq. (??). We see that due to the presence of the delta-function $\delta(t - s)$ we acquire a divergence trying to take the continuous limit.

We will deal with this divergence staying in discrete time by observing that the same singular term appears in the calculation of W_X^2. Therefore, we can express the singular term through W_X^2 and substitute it in $\langle f(\mu_i)\dot{\mu}_i^2\rangle_{P_X}$, thus obtaining an expression which does not contain singular terms and allows us to take the continuous limit. First of all, we should find an explicit expression for the matrix Q^{-1}. It can be done if we diagonalize the matrix $(D^T D)_{jk} \equiv \sum_{i=0}^{M-1} D_{ij}D_{ik}$ (here D^T stands for the transpose of the matrix D), which has the following form:

$$(D^T D)_{jk} = \frac{1}{\delta_t^2}\begin{pmatrix} 2 & -1 & 0 & \cdots & 0 & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 & 0 \\ \vdots & & & & & & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 2 & -1 \\ 0 & 0 & 0 & \cdots & 0 & -1 & 2 \end{pmatrix}. \tag{190}$$

The eigenvectors of the matrix from the eq. (??) read:

$$v_j^{(\alpha)} = \sqrt{\frac{2}{M}}\sin\frac{\pi j\alpha}{M},\ 1 \le \alpha \le M - 1, \tag{191}$$

where each vector corresponds to the eigenvalue ν_α:

$$\nu_\alpha = \frac{2}{\delta_t^2}\left(1 - \cos\frac{\pi\alpha}{M}\right) = \frac{4}{\delta_t^2}\sin^2\frac{\pi\alpha}{2M}. \tag{192}$$

The normalization factor $\sqrt{2/M}$ in the eq. (??) ensures that the vectors $v_j^{(\alpha)}$ satisfy the following orthogonality and completeness relations:

$$\sum_{j=1}^{M-1} v_j^{(\alpha)}v_j^{(\beta)} = \delta_{\alpha\beta},$$

$$\sum_{j=1}^{M-1} v_j^{(\alpha)}v_k^{(\alpha)} = \delta_{jk}. \tag{193}$$

We can expand the matrix Q in terms of the vectors $v_j^{(\alpha)}$:

$$Q_{jk} = m\delta_t(D^T D)_{jk} + m\Omega^2\delta_t\delta_{jk} = \sum_{\alpha=1}^{M-1} Q_\alpha v_j^{(\alpha)}v_k^{(\alpha)}, \tag{194}$$

where the eigenvalues Q_α are:

$$Q_\alpha = m\delta_t\nu_\alpha + m\Omega^2\delta_t. \tag{195}$$

Therefore, the inverse of the matrix Q is:

$$Q_{jk}^{-1} = \sum_{\alpha=1}^{M-1} \frac{1}{Q_\alpha} v_j^{(\alpha)} v_k^{(\alpha)}. \tag{196}$$

Now we can compute the result of applying two derivative operators D_{ij} to the matrix Q^{-1}:

$$\sum_{j,k=1}^{M-1} D_{ij} D_{ik} Q_{jk}^{-1} = \sum_{\alpha=1}^{M-1} \frac{1}{Q_\alpha} \left(\sum_{j=1}^{M-1} D_{ij} v_j^{(\alpha)} \right)^2. \tag{197}$$

Recalling the expressions for the components of the vectors $v_j^{(\alpha)}$ (see the eq. (??)), one can derive how D_{ij} acts on these vectors:

$$\sum_{j=1}^{M-1} D_{ij} v_j^{(\alpha)} = \frac{2}{\delta_t} \sqrt{\frac{2}{M}} \sin\frac{\pi\alpha}{2M} \cos\frac{\pi\alpha(2i+1)}{2M}, \, 0 \le i \le M-1, \tag{198}$$

and substitute the result of (??) into the eq. (??):

$$\sum_{j,k=1}^{M-1} D_{ij} D_{ik} Q_{jk}^{-1} = \frac{8}{M\delta_t^2} \sum_{\alpha=1}^{M-1} \frac{1}{Q_\alpha} \sin^2\frac{\pi\alpha}{2M} \cos^2\frac{\pi\alpha(2i+1)}{2M} = \tag{199}$$

$$\frac{2}{M} \sum_{\alpha=1}^{M-1} \frac{\nu_\alpha}{Q_\alpha} \cos^2\frac{\pi\alpha(2i+1)}{2M} = \frac{2}{mT} \sum_{\alpha=1}^{M-1} \cos^2\frac{\pi\alpha(2i+1)}{2M} -$$

$$\frac{2\Omega^2}{mT} \sum_{\alpha=1}^{M-1} \frac{1}{\nu_\alpha + \Omega^2} \cos^2\frac{\pi\alpha(2i+1)}{2M} = \frac{M-1}{mT} - \frac{2\Omega^2}{mT} \sum_{\alpha=1}^{M-1} \frac{1}{\nu_\alpha + \Omega^2} \cos^2\frac{\pi\alpha(2i+1)}{2M}. \tag{200}$$

The squared bandwidth W_X^2 can be evaluated in the same way, which gives:

$$W_X^2 = \frac{1}{P} \int_T \frac{dt}{T} \langle |\dot{X}|^2 \rangle_{Px} = \frac{1}{P} \sum_{i=0}^{M-1} \frac{\delta_t}{T} \langle \dot{a}_i^2 + \dot{b}_i^2 \rangle_{Px} =$$

$$\frac{2\delta_t}{PT} \sum_{i=0}^{M-1} \sum_{j,k=1}^{M-1} D_{ij} D_{ik} Q_{jk}^{-1} = \frac{2\delta_t}{PT} \sum_{\alpha=1}^{M-1} \frac{\nu_\alpha}{Q_\alpha} = \frac{2(M-1)}{mPT} - \frac{2\Omega^2}{mPT} \sum_{\alpha=1}^{M-1} \frac{1}{\nu_\alpha + \Omega^2}. \tag{201}$$

We see that the expressions (??) and (??) both contain the same singularity $M/T = 1/\delta_t$. Expressing the singular term through the squared bandwidth and substituting it in the eq. (??), we obtain:

$$\sum_{j,k=1}^{M-1} D_{ij} D_{ik} Q_{jk}^{-1} = \frac{1}{2} P W_X^2 - \frac{\Omega^2}{mT} \sum_{\alpha=1}^{M-1} \frac{1}{\nu_\alpha + \Omega^2} \cos\frac{\pi\alpha(2i+1)}{M}. \tag{202}$$

38

Now we can substitute the right-hand side of the eq. (??) in $\langle f(\mu_i)\dot{\mu}_i^2\rangle_{P_X}$:

$$\langle f(\mu_i)\dot{\mu}_i^2\rangle_{P_X} = \frac{4(\gamma L)^2}{2\pi(Q^{-1})_{ii}}\int dxdy f(x,y)(x^2+y^2)\mathrm{e}^{-\frac{x^2+y^2}{2(Q^{-1})_{ii}}}$$

$$\left[\frac{1}{2}PW_X^2 - \frac{\Omega^2}{mT}\sum_{\alpha=1}^{M-1}\frac{1}{\nu_\alpha+\Omega^2}\cos\frac{\pi\alpha(2i+1)}{M} - \right.$$

$$\left.\sum_{j,k=1}^{M-1}D_{ij}D_{ik}\frac{(Q^{-1})_{ji}(Q^{-1})_{ki}}{(Q^{-1})_{ii}}\left(1-\frac{x^2+y^2}{(Q^{-1})_{ii}}\right)\right]. \tag{203}$$

Before taking the continuous limit of the eq. (??), we should note that most of the terms in the eq. (??) can be neglected, because we consider the amount of points $M\sim TW_X$ needed to represent the input signal to be large. Let us first consider the following sum:

$$\frac{\Omega^2}{mT}\sum_{\alpha=1}^{M-1}\frac{1}{\nu_\alpha+\Omega^2}\cos\frac{\pi\alpha(2i+1)}{M}, \tag{204}$$

and compare it to PW_X^2. The contribution of the sum (??) can be estimated as follows:

$$\frac{\Omega^2}{mT}\left|\sum_{\alpha=1}^{M-1}\frac{1}{\nu_\alpha+\Omega^2}\cos\frac{\pi\alpha(2i+1)}{M}\right| \leq \frac{1}{mT}\sum_{\alpha=1}^{M-1}\frac{1}{1+(M/\xi)^2\sin^2\frac{\pi\alpha}{2M}} \leq$$

$$\frac{\xi^2}{mTM^2}\sum_{\alpha=1}^{M-1}\frac{1}{\sin^2\frac{\pi\alpha}{2M}} \leq \frac{\xi^2}{mT}\sum_{\alpha=1}^{M-1}\frac{1}{\alpha^2} \leq \frac{2P\xi^4}{T^2}\frac{\pi^2}{6}, \tag{205}$$

where we used the relations between m, Ω and P, T:

$$\frac{1}{P\xi} = m\Omega,\ 2\xi = \Omega T, \tag{206}$$

and also the following inequality:

$$\sin x \geq \frac{2}{\pi}x, \qquad |x|\leq \pi/2. \tag{207}$$

Now one can see that the sum (??) is negligible, because $W_X\gg 1/T$. The second term which should be considered reads:

$$\sum_{j,k=1}^{M-1}D_{ij}D_{ik}\frac{(Q^{-1})_{ji}(Q^{-1})_{ki}}{(Q^{-1})_{ii}} = \frac{4PM^2}{T^2}\left(\sum_{\alpha=1}^{M-1}\frac{1}{1+(M/\xi)^2\sin^2\frac{\pi\alpha}{2M}}\sin\frac{\pi\alpha}{2M}\right.$$

$$\left.\cos\frac{\pi\alpha(2i+1)}{2M}\sin\frac{\pi i\alpha}{M}\right)^2\bigg/\left(\sum_{\alpha=1}^{M-1}\frac{1}{1+(M/\xi)^2\sin^2\frac{\pi\alpha}{2M}}\sin^2\frac{\pi i\alpha}{M}\right) \leq$$

$$\frac{APM}{T^2} \sim \frac{APW_X}{T}, \tag{208}$$

where A is some real positive constant. We see again that the contribution from the eq. (??) can be neglected due to the presence of the factor $1/T \ll W_X$.

Finally, neglecting the small terms and changing variables in the integral to $x = \rho \cos\phi$ and $y = \rho \sin\phi$, we write the continuous limit of the eq. (??):

$$\langle f(\mu(t))\dot\mu(t)^2 \rangle_{P_X} \approx \frac{2PW_X^2(\gamma L)^2 m\Omega\,\mathrm{sh}\,\Omega T}{\mathrm{sh}\,\Omega(T/2+t)\,\mathrm{sh}\,\Omega(T/2-t)} \int_0^{+\infty} d\rho\rho^3 f(\mu) \qquad (209)$$
$$\exp -\frac{m\Omega\rho^2}{2} \frac{\mathrm{sh}\,\Omega T}{\mathrm{sh}\,\Omega(T/2+t)\,\mathrm{sh}\,\Omega(T/2-t)}.$$

Integration in the eq. (??) over the time interval $[-T/2, T/2]$ will give us a contribution to the mutual information. The integral over time can be calculated by first shifting the integration interval: $t = t' - T/2$ and then changing the variable from t' to $x = (\mathrm{cth}\,\Omega(T-t') - \mathrm{cth}\,\Omega T)^{-1}$. Then we have the following integral:

$$\int_0^T \frac{dt'}{T} \langle f(\mu)\dot\mu^2 \rangle_{P_X} \approx \frac{2mPW_X^2(\gamma L)^2}{T} \int_0^{+\infty} d\rho f(\mu)\rho^3 \int_0^{+\infty} \frac{dx}{x} \quad (210)$$
$$\exp\left\{ -\frac{m\Omega\rho^2}{2}\left(\frac{1}{x} + \frac{x}{\mathrm{sh}^2\,\Omega T}\right) - \mathrm{cth}\,\Omega T\, m\Omega\rho^2 \right\},$$

which can be calculated by using the integral representation of the modified Bessel function $K_\nu(x)$:

$$\int_0^{+\infty} dx\, x^{\nu-1} \exp\left(-\frac{p}{x} - qx\right) = 2\left(\frac{p}{q}\right)^{\nu/2} K_\nu(2\sqrt{pq}),\ p > 0,\ q > 0. \quad (211)$$

Eventually, we get the following result:

$$\int_{-T/2}^{T/2} \frac{dt}{T} \langle f(\mu)\dot\mu^2 \rangle_{P_X} \approx \frac{4(\gamma L)^2 mPW_X^2}{T} \int_0^{+\infty} d\rho f(\mu)$$
$$\rho^3 \mathrm{e}^{-m\Omega\,\mathrm{cth}\,\Omega T\,\rho^2} K_0\left(\frac{m\Omega\rho^2}{\mathrm{sh}\,\Omega T}\right). \quad (212)$$

Expressing m and Ω through the parameters of the input signal P and W_X^2 according to the relations (??) we arrive at the formula:

$$\int_{-T/2}^{T/2} \frac{dt}{T} \langle f(\mu)\dot\mu^2 \rangle_{P_X} \approx (\gamma LP)^2 W_X^2 \mathrm{sh}^2\,2\xi \int_0^{+\infty} dy\, y f\left(\gamma LP\xi\,\mathrm{sh}\,2\xi\,y\right) \quad (213)$$
$$K_0(y)\,\mathrm{e}^{-y\,\mathrm{ch}\,2\xi}.$$

The second integral which we have to evaluate can be treated in the exactly same way as the first integral. The result reads:

$$\int_{-T/2}^{T/2} \frac{dt}{T} \left\langle \frac{4\mu^3}{15(3+\mu^2)} \right\rangle_{P_X} = \frac{2}{15}\xi^2(\gamma LP)^3 \mathrm{sh}^4\,2\xi$$
$$\int_0^{+\infty} dy \frac{y^3 K_0(y)}{3 + (\gamma LP\xi)^2 \mathrm{sh}^2\,2\xi\,y^2} \mathrm{e}^{-y\,\mathrm{ch}\,2\xi} \quad (214)$$

40