Quantum nondemolition measurements of photon number in monolithic microcavities

S. N. Balybin,[1,2,*] A. B. Matsko,[3] F. Ya. Khalili,[1,4] D. V. Strekalov,[3] V. S. Ilchenko,[3] A. A. Savchenkov,[5] N. M. Lebedev,[1,2] and I. A. Bilenko[1,2]

[1]*Russian Quantum Center, Skolkovo IC, Bolshoy Bulvar 30, building 1, Moscow, 121205, Russia*
[2]*M.V. Lomonosov Moscow State University, Faculty of Physics, Leninskie Gory, 1 bld. building 2, 119991, Russia*
[3]*Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, California 91109-8099, USA*
[4]*NTI Center for Quantum Communications, National University of Science and Technology MISiS, Leninsky Prospekt 4, Moscow 119049, Russia*
[5]*Rockley Photonics, 234 E Colorado Boulevard Suite 600, Pasadena, California 91101, USA*

(Received 7 December 2021; revised 5 April 2022; accepted 6 July 2022; published 28 July 2022)

We revisit the idea of quantum nondemolition measurement (QND) of optical quanta via a resonantly enhanced Kerr nonlinearity taking into account quantum backaction. We show that the monolithic microcavities enable QND measurement of the number of quanta in a weak signal field using a classical probe field spatially overlapping with the signal. The phase of the probe field acquires information about the signal number of quanta without altering it due to the cross-phase modulation effect. We find the exact solution to the Heisenberg equations of motion of this system and calculate the measurement error, accounting for the optical losses in the measurement path. We identify a realistic approximation to obtain the explicit form of the final conditional quantum state of the signal field, accounting for the undesirable self-phase modulation effect and designing the optimal homodyne measurement of the probe beam to evade this effect. We show that the best modern monolithic microcavities allow achieving the measurement imprecision several times better than the standard quantum limit.

DOI: 10.1103/PhysRevA.106.013720

I. INTRODUCTION

Quantum nondemolition measurement (QND) measurement schemes involving coupling of two optical waves by means of the cross-phase modulation (XPM) effect resulting from the cubic nonlinearity of an optical material were proposed more than 40 yr ago [1,2]. The experimental work based on the idea started nearly at the same time [3]. Whispering gallery mode (WGM) optical resonators were considered as one of optimal venues for the QND realization because of their high quality factors and small mode volumes [4].

Self-phase modulation (SPM) limits the measurement sensitivity in the proposed resonant schemes. A method eliminating the SPM was proposed in Ref. [5]. The technique utilizes the optimized detuning of both the signal and the probe fields from the corresponding resonance frequencies of a nonlinear microcavity. As it follows from that study, this approach cancels not only the SPM in the probe beam, but also the perturbation of the signal mode phase, which ultimately contradicts the Heisenberg uncertainty relation (14).

In Ref. [6], a theoretical analysis of the optical modes coupled by means of the XPM effect was performed, showing that strongly non-Gaussian quantum states of light can be prepared using such systems. However, that work was aimed only at the preparation of the signal mode quantum state for the particular case of the initial coherent state with low intensity. Also, the SPM effect was not taken into account, and only linearized treatment was performed.

In this paper we study a feasibility of the QND measurement of a photon number in a modern nonlinear microcavity. Following previous developments in the field, we: (i) provide a consistent quantum analysis of the two-mode QND measurement scheme based on the $\chi^{(3)}$-nonlinearity, accounting for the SPM effect, and (ii) evaluate the prospects of experimental implementation of this scheme using the recent achievements in fabrication of the high-Q WGM resonators. As the result of our analysis it becomes possible to conclude that the photon number in the probe mode should exceed the photon number in the signal mode to beat the standard quantum limit (SQL) if the optical measurement strategy is selected.

We also have performed a detailed analysis of an imperfect measurement scheme, taking into account a limited efficiency of the photon detection and attenuation of light. We analyze these effects using a closed quantum system ansatz as well as a more rigorous Langevin approach. Both considerations let us arrive at the equivalent results.

This article is organized as follows. Section II describes the basic principles of QND used throughout our paper. The basic principles and salient problems of the cross-phase modulation based QND techniques are discussed in Sec. III. In Sec. IV, we start with the simplified semiclassical treatment of the measurement scheme and find the measurement sensitivity. In Sec. V, we derive the exact solution of the corresponding Heisenberg equations of motion and identify an important from the practical point of view the asymptotic

*sn.balybin@physics.msu.ru

case of this solution. In Sec. VI, we find the explicit forms of the final quantum state of the signal mode and of the probability distribution for the measurement results for this asymptotic case, accounting for the SPM. We show that this distribution could be strongly non-Gaussian. In Sec. VII, we estimate the sensitivity, achievable with the modern WGM microresonators used for the QND measurements. Finally, in Sec. VIII we summarize the results of this paper.

II. CRITERIA OF QND

According to von Neumann's reduction postulate [7], any ideal (that is precise and free from technical imperfections) quantum measurement leaves the object in the eigenstate of the measured observable corresponding the eigenvalue obtained as the measured result. If the object is already in such an eigenstate before the measurement, then this state remains unchanged after the measurement. The majority of the real world measurement devices does not obey this rule and perturb the measured observable.

Let us consider a problem of a nondisturbing detection of a number of photons in a mode of a lossless optical resonator as an example. A standard tool for the measurement of the number of photons, a photocounter absorbs all the counted photons, leaving the optical cavity field in the ground state. An output of a phase-preserving linear amplifier [8], that also can be utilized for the photon counting, depends on the two noncommuting quadratures of an input mode and due to the Heisenberg uncertainty relation disturbs them. It cannot measure the number of quanta n in this mode with the precision better than the SQL,

$$\Delta n_{\mathrm{SQL}} = \sqrt{\bar{n}}, \tag{1}$$

where $\bar{n}$ is the mean number of quanta. Due to the same reasons, it also perturbs n by, at least, the same value. Therefore, obtaining the sensitivity better than the limit (1) with the photons number perturbation also below this limit, can be considered as the minimal requirement for the "true quantum" measurement.

A sufficient condition for implementing the ideal von Neumann's measurement was explicitly formulated by Bohm [9]. He showed that the eigenstates of the measured observable $\hat{q}$ are not affected by the measurement if this observable commutes with the Hamiltonian $\hat{H}$ of the combined system,

$$[\hat{q}, \hat{H}] = 0, \tag{2}$$

where

$$\hat{H} = \hat{H}_S + \hat{H}_A + \hat{H}_I, \tag{3}$$

$\hat{H}_S$ and $\hat{H}_A$ are, respectively, the Hamiltonians of the object and the meter, and $\hat{H}_I$ is the interaction Hamiltonian (it was assumed in Ref. [9], that during the measurement, $\hat{H}_{S,A} \to 0$, but this assumption does not affect the main conclusion). The term (QND measurements was proposed for this class of quantum measurements in the late 1970s [10,11] and has become generally accepted since then.

The number of quanta in an electromagnetic mode of a linear cavity is an integral of motion, commuting with the Hamiltonian of the mode $[\hat{n}, \hat{H}_S] = 0$ (the *QND observable*). In this case, the condition (2) can be simplified to the commutativity with the interaction Hamiltonian,

$$[\hat{n}, \hat{H}_I] = 0. \tag{4}$$

In other words, the coupling of the mode with the meter has to be nonlinear in the mode generalized coordinate represented by the field strength.

III. ADVANCES AND PROBLEMS OF XPM-BASED QND

Following the initial semigedanken proposals [12–14], a realistic scheme of QND measurement of electromagnetic photons number was proposed in Refs. [1,2]. It uses two spatially overlapping optical modes, the signal (object) and the meter (probe) ones, interacting by means of the cubic optical nonlinearity $\chi^{(3)}$ arising due to the Kerr effect. During the interaction, the phase of the probe mode is shifted by the value proportional to the photon number in the signal mode. This effect is called XPM. The photon number of the signal mode is preserved during the interaction in the ideal lossless case. The phase of this mode is perturbed by the number of quanta uncertainty of the probe mode due to the same XPM mechanism. This perturbation ensures the fulfillment of the Heisenberg uncertainty principle meaning that reduction of the uncertainty of a signal observable should lead to increase of the uncertainty of an observable conjugated to the signal.

Later on, a significant amount of experimental work based on this idea was performed, starting from the pioneering work [3], see the reviews [15–17] for more detail. The sensitivity exceeding the SQL was demonstrated in these experiments, but the single-photon accuracy limit was not reached.

In parallel, another class of the QND schemes, which uses single atoms [18–20] or superconductive nonlinear circuits (artificial atoms) [21] as nonlinear elements, was actively developed during the past decades. These, in essence, lumped devices are capable of providing resonant cubic nonlinearity many orders of magnitude larger than the electronic nonlinearity of the transparent dielectrics, which led to successful measurements of a single photon.

An important disadvantage of the atom-based QND measurements is their complexity. It is desirable to perform the measurements on a chip without involvement of bulky equipment needed for the atomic systems. The superconductive circuits can be, and usually are, implemented on chip, but they operate in a microwave band and require cryogenic cooling. In addition, both these measurement classes, whereas sensitive to a single or a few quanta, do not scale well to bright (multiquanta) states.

The main problem with the pure optical implementations of QND is that the high optical nonlinearity is typically associated with the high absorption. The promising way to overcome this problem is the usage of WGM optical microresonators [4,22], which combine very high quality ($Q-$) factors, exceeding 10^{11} in crystalline microresonators [23] and 10^9 in on-chip ones [24,25] with a high concentration of the optical energy in the small volume of the optical modes.

The WGM resonators have also a series of practical advantages, which makes them perhaps the most promising QND platform. Because of a broad-band nature of the total internal reflection, their Q factors remain very high within an optical wavelength range far exceeding an octave. Therefore, their QND applications is not tied to any specific wavelength as in atoms or resonators using dielectric mirrors, which allow for a greater versatility. Furthermore, the WGM resonators have a continuously tunable coupling rate. This allows for a fine control over a parameter responsible for various QND regimes.

Another major problem, specific to the $\chi^{(3)}$ nonlinearity, is associated with the SPM effect resulting in perturbation of the phases of both the probe and the signal modes by the energy uncertainties of the corresponding modes [26]. It is not so crucial for the signal mode because, whereas distorting (squeezing) its final quantum state, it does not affect the number of quanta in this mode. At the same time, it introduces an additional uncertainty into the phase of probe mode, proportional to the number of quanta uncertainty in this mode, thus, limiting the measurement precision (see details in Sec. IV). This is the so-called quantum backaction effect.

Two straightforward methods of cancellation of this effect were proposed in Ref. [26]: either using a resonant $\chi^{(3)}$ medium or passing the probe beam through a negative $\chi^{(3)}$ medium before the detection. More recently, implementations of optical QND measurements using rubidium atoms in a magneto-optical trap were studied experimentally [27]. It was also noted that semiconductor quantum dots can provide the negative nonlinearity of proper magnitude to compensate for the SPM in experiments with quantum solitons [28].

Unfortunately, these methods cannot be considered as simple ones. A more practical method based on the measurement of the optimal quadrature of the output probe field instead of the phase one was proposed in Ref. [29]. This measurement allows one to eliminate the major linear part of the SPM and can be made using the ordinary homodyne detectors.

In what follows we develop a technique of the backaction avoiding QND measurement based on a WGM system. The proposed here QND implementation is free of the problems encountered by the previous propositions involving WGM resonators.

IV. SIMPLIFIED ANALYSIS

Let us start with a simple intuitive semiclassical approach, considering the classical equations of motion but assuming that the initial values of the involved observables have quantum uncertainties. The validity of the approach will be justified in the next section.

Within this approach, the evolution of the phases ϕ_p and ϕ_s of the probe (p) and the signal (s) waves propagating in a nonlinear media with a cubic (Kerr) nonlinearity in the rotating-wave frame is described by the following equations:

$$\phi_p(t) = \phi_p + \Gamma_S n_p + \Gamma_X n_s, \tag{5a}$$

$$\phi_s(t) = \phi_s + \Gamma_S n_s + \Gamma_X n_p, \tag{5b}$$

where $n_{p,s}$'s are the photon numbers in these modes, which are preserved during the interactions, $\phi_{p,s}$'s are the initial values of the phases,

$$\Gamma_{S,X} = \gamma_{s,x}\tau, \tag{6}$$

γ_s and γ_x are the SPM and XPM nonlinearity factors, and τ is the effective duration of the interaction. The last two terms in Eqs. (5), proportional to Γ_X, describe, respectively, the signal phase shift in the probe mode and the perturbation of the signal mode phase,

$$\Delta\phi_{s\,\text{pert}} = \Gamma_X \Delta n_p, \tag{7}$$

where Δn_p is the initial uncertainty of n_p.

Suppose that the output phase of the probe mode $\phi_p(t)$ is measured by a phase-sensitive detector. In this case, initial uncertainties of both the phase and the number of quanta of the probe mode contribute to the measurement error. The signal photon number n_s can be estimated with the uncertainty,

$$\Delta n_{s\,\text{meas}} = \frac{1}{\Gamma_X}\sqrt{(\Delta\phi_p)^2 + \Gamma_S^2(\Delta n_p)^2} \geqslant \frac{\Gamma_S}{\Gamma_X}\Delta n_p. \tag{8}$$

Since usually $\Gamma_X \sim \Gamma_S$, we find that

$$\Delta n_{s\,\text{meas}} \gtrsim \Delta n_p. \tag{9}$$

For the coherent initial quantum state of the probe mode,

$$\Delta\phi_p = \frac{1}{2\sqrt{\bar{n}_p}}, \qquad \Delta n_p = \sqrt{\bar{n}_p}, \tag{10}$$

where $\bar{n}_p$ is the expectation number of probe quanta, Eq. (8) results in

$$\Delta n_{s\,\text{meas}} = \frac{1}{\Gamma_X}\sqrt{\frac{1}{4\bar{n}_p} + \Gamma_S^2\bar{n}_p} > \frac{\Gamma_S}{\Gamma_X}\sqrt{\bar{n}_p}. \tag{11}$$

As it follows from this inequality in order to overcome the SQL [see Eq. (1)], there should be $\bar{n}_s \gtrsim \bar{n}_p$, which makes a high-precision QND measurement of a small number of quanta impossible.

Let us consider now a measurement of the linear combination of the probe phase and photon number (see Ref. [26]),

$$\phi_p(t) - \Gamma_S n_p(t) = \phi_p + \Gamma_X n_s. \tag{12}$$

In this case, the sensitivity is affected only by initial uncertainty of the probe phase,

$$\Delta n_{s\,\text{meas}} = \frac{\Delta\phi_p}{\Gamma_X}, \tag{13}$$

which could be arbitrarily small, provided sufficiently big nonlinearity factor Γ_X and the probe photon number.

It follows also from Eqs. (7) and (13) that

$$\Delta n_{s\,\text{meas}}\Delta\phi_{s\,\text{pert}} = \Delta\phi_p\Delta n_p \geqslant \tfrac{1}{2}, \tag{14}$$

that is, the uncertainty relation for the number of quanta and phase of the probe mode directly translates to the uncertainty relation for $\Delta n_{s\,\text{meas}}$ and $\Delta\phi_{s\,\text{pert}}$. The general form of the number of quanta and phase uncertainty relations is much more involved, see, e.g., Ref. [30]. The simplified form of the uncertainty relation (14) has good precision for the 'practical quantum states with $\Delta\phi \ll 1$ considered here.

In the case of the coherent quantum state of the probe mode (10),

$$\Delta n_{s\,\text{meas}} = \frac{1}{2\Gamma_X\sqrt{\bar{n}_p}}, \tag{15a}$$

$$\Delta\phi_{s\,\text{pert}} = \Gamma_X\sqrt{\bar{n}_p}. \tag{15b}$$

Finally, the necessary condition for a successful sub-SQL measurement can be presented as $\Delta n_{s\,\rm meas} < \sqrt{\bar{n}_s}$. With account of Eq. (15a), it corresponds to the following inequality:

$$2\Gamma_X\sqrt{\bar{n}_p}\sqrt{\bar{n}_s} > 1. \tag{16}$$

V. MEASUREMENT IMPRECISION

Using the rotating-wave approximation, the Hamiltonian of the two-mode system, considered in the previous section, can be presented as follows:

$$\hat{\mathcal{H}} = -\frac{\hbar\gamma_S}{2}\sum_{x=s,p}\hat{n}_x(\hat{n}_x - 1) - \hbar\gamma_X\hat{n}_p\hat{n}_s, \tag{17}$$

where $\hbar$ is the reduced Plank constant,

$$\hat{n}_{s,p} = \hat{a}^\dagger_{s,p}\hat{a}_{s,p}, \tag{18}$$

and $\hat{a}_{s,p}$, $\hat{a}^\dagger_{s,p}$ are the annihilation and creation operators of the signal and the probe modes (the peculiar form of the first term is the normal-ordered one). It can be seen from this Hamiltonian, that the numbers of quanta in both modes (in the Heisenberg picture) are integrals of motion of the system,

$$\hat{n}_{s,p}(t) = \hat{n}_{s,p}. \tag{19}$$

The corresponding Heisenberg equations of motion for the annihilation operators are

$$\frac{d\hat{a}_p(t)}{dt} = i[\gamma_S\hat{n}_p(t) + \gamma_X\hat{n}_s(t)]\hat{a}_p(t), \tag{20a}$$

$$\frac{d\hat{a}_s(t)}{dt} = i[\gamma_S\hat{n}_s(t) + \gamma_X\hat{n}_p(t)]\hat{a}_s(t). \tag{20b}$$

Due to the photon number conservation (19) the closed form of the solution of the set of equations can be easily found

$$\hat{a}_p(t) = e^{i(\Gamma_S\hat{n}_p + \Gamma_X\hat{n}_s)}\hat{a}_p, \tag{21a}$$

$$\hat{a}_s(t) = e^{i(\Gamma_S\hat{n}_s + \Gamma_X\hat{n}_p)}\hat{a}_s. \tag{21b}$$

.

Let us consider now the homodyne measurement of the quadrature $\hat{X}_\zeta$ of the probe mode, defined by the homodyne angle ζ,

$$\hat{X}_\zeta = \frac{1}{\sqrt{2}}[\hat{a}_p(t)e^{i\zeta} + \text{H.c.}] = \frac{1}{\sqrt{2}}[e^{i(\Gamma_S\hat{n}_p + \Gamma_X\hat{n}_s + \zeta)}\hat{a}_p + \text{H.c.}], \tag{22}$$

where H.c. stands for "Hermitian conjugate." The measurement error for the number of quanta in the signal mode can be calculated by standard error propagation formula,

$$(\Delta n_s)^2 = \frac{(\Delta X_\zeta)^2}{G^2}, \tag{23}$$

where

$$G = \frac{\partial\langle\hat{X}_\zeta\rangle}{\partial n_s} \tag{24}$$

is the transfer function,

$$(\Delta X_\zeta)^2 = \langle\hat{X}^2_\zeta\rangle - \langle\hat{X}_\zeta\rangle^2, \tag{25}$$

and momenta $\langle\hat{X}_\zeta\rangle$ and $\langle\hat{X}^2_\zeta\rangle$ are calculated for a given value of n_S, that is, for the Fock state $|n_s\rangle$ of the signal mode.

If the probe mode state is prepared in a coherent state $|\alpha\rangle_p$, the expectation value of the probe amplitude can be selected to be real,

$$\alpha = \sqrt{\bar{n}_p}. \tag{26}$$

The exact value of the measurement error Δn_s for this case is calculated in Appendix A, see Eqs. (A5).

In the case of weak nonlinearity and strong probe field the solution can be simplified. Let us assume that

$$|\Gamma_S| \to 0, \qquad \bar{n}_p \to \infty, \qquad \text{but} \quad \Gamma_S\bar{n}_p \text{ remains finite.} \tag{27}$$

This approximation is well satisfied for the realistic WGM resonators (see Sec. VII). In this case (see Appendix A),

$$(\Delta X)^2 = \tfrac{1}{2} - \Gamma_S\bar{n}_p\sin 2\varphi + 2\Gamma_S^2\bar{n}_p^2\sin^2\varphi, \tag{28a}$$

$$G = -\sqrt{2}\alpha\Gamma_X\sin\varphi, \tag{28b}$$

where

$$\varphi = \Gamma_S\bar{n}_p + \Gamma_X n_s + \zeta. \tag{29}$$

These equations correspond to the ideal exact measurement of $\hat{X}_\zeta$. The losses in the measurement channel can be taken into account by introducing its unified quantum efficiency η (which includes, in particular, the finite quantum efficiency of the homodyne detector) as

$$(\Delta X)^2_\eta = \eta(\Delta X)^2 + \frac{1-\eta}{2}, \tag{30a}$$

$$G_\eta = \sqrt{\eta}G, \tag{30b}$$

resulting in an expression for the measurement error for the number of quanta in the signal mode,

$$(\Delta n_s)^2 = \frac{(\Delta X_\zeta)^2_\eta}{G^2_\eta} = \frac{1}{\Gamma_X^2}\left[\frac{1 + (\cot\varphi - 2\eta\Gamma_S\bar{n}_p)^2}{4\eta\bar{n}_p} + (1-\eta)\Gamma_S^2\bar{n}_p\right]. \tag{31}$$

Following the reasoning of Sec. IV, we assume that $\cot\varphi = 0$, which corresponds to the maximum of the transfer function as well as to the measurement of the phase quadrature of the output probe beam. Following this path we arrive at:

$$(\Delta n_s)^2 = \frac{1}{\Gamma_X^2}\left(\frac{1}{4\eta\bar{n}_p} + \Gamma_S^2\bar{n}_p\right). \tag{32}$$

In the ideal case of $\eta = 1$, this equation reduces to Eqs. (11).

The minimum of the measurement error of the optimized detection procedure described by (31),

$$(\Delta n_{s,\min})^2 = \frac{1}{\Gamma_X^2}\left[\frac{1}{4\eta\bar{n}_p} + (1-\eta)\Gamma_S^2\bar{n}_p\right] \tag{33}$$

is achieved at the optimum angle φ given by

$$\cot\varphi = 2\eta\Gamma_S\bar{n}_p. \tag{34}$$

In the ideal case of $\eta = 1$, the additional term in $(\Delta n_s)^2$ vanishes, giving Eq. (15a).

Our reasoning contains a "vicious loop:" the value of ζ, defined by Eq. (34), depends on the measured value n_s, which is unknown before the measurement. Let us consider an important from the practical point of view case of

$$|n_s - \bar{n}_s| \ll \bar{n}_s. \tag{35}$$

It includes the coherent *a priory* quantum state of the signal mode. The *a priory* condition means that the expectation value $\bar{n}_s$ is known by definition. In this case, Eq. (34) can be replaced with the following condition:

$$\cot\bar{\varphi} = 2\eta\Gamma_S\bar{n}_p, \tag{36a}$$

where

$$\bar{\varphi} = \Gamma_S\bar{n}_p + \Gamma_X\bar{n}_s + \zeta. \tag{36b}$$

It is shown in Appendix B, that under reasonable assumptions, the condition (36) leads only to a minor correction to Eq. (33), see Eq. (B6).

It follows from Eq. (33) that the optimal number of probe mode quanta, equal to

$$\bar{n}_p^{\rm opt} = \frac{1}{2\Gamma_S\sqrt{\eta(1-\eta)}} \tag{37}$$

exists, which gives the following minimized measurement error:

$$\left(\Delta n_{s,\min}^{\rm opt}\right)^2 = \frac{\Gamma_S}{\Gamma_X^2}\sqrt{\frac{1-\eta}{\eta}}. \tag{38}$$

To validate the results presented above we study the lossy system using the Langevin approach in Appendix D. The result obtained with both methods led to the same conclusion: the QND of the photon number is feasible if the imperfections of the measurement system are small enough.

The Langevin approach also resulted in an interesting observation. It is possible to envision a scheme in which the signal mode and the probe mode are coupled in different ways. The signal is strongly undercoupled, so its attenuation is minimized to a value defined by the intrinsic Q factor of the signal mode. The probe mode can be overcoupled to minimize the attenuation of the probe light passing the cavity. Such a loading difference is feasible if the signal and probe waves have significantly different carrier frequencies. In that way one can consider sequential QND measurements in the cavity.

VI. CONDITIONAL STATE OF THE SIGNAL MODE

Let us consider the wave function of the final quantum state of the joint two-mode systems and assume for simplicity that $\eta = 1$. Using the Hamiltonian (17) we find for the final state,

$$|\Psi\rangle = \hat{\mathcal{U}}|\alpha\rangle_p \otimes |\psi\rangle_s, \tag{39}$$

where

$$\hat{\mathcal{U}} = \exp\frac{\hat{\mathcal{H}}t}{i\hbar} = \exp\left[\frac{i\Gamma_S}{2}\sum_{x=p,s}\hat{n}_x(\hat{n}_x - 1) + i\Gamma_X\hat{n}_p\hat{n}_s\right] \tag{40}$$

is the evolution operator.

Measurement of the probe mode quadrature $\hat{X}_\zeta$ reduces the signal mode into the following quantum state:

$$|\psi(X)\rangle = \frac{\hat{\Omega}(X)|\psi\rangle_s}{\sqrt{W(X)}}, \tag{41}$$

where X is the measurement result,

$$\hat{\Omega}(X) = {}_p\langle X,\zeta|\hat{\mathcal{U}}|\alpha\rangle_p = \sum_{n_s=0}^{\infty} e^{i\Gamma_S n(n-1)/2}\Omega(X,n_s)|n_s\rangle_{s\,s}\langle n_s| \tag{42}$$

is the reduction (Kraus) operator, $|X,\zeta\rangle_p$ is the eigenstate of $\hat{X}_\zeta$ with the eigenvalue X,

$$W(X) = {}_s\langle\psi|\hat{\Pi}(X)|\psi\rangle_s \tag{43}$$

is the *a priory* probability distribution of X,

$$\hat{\Pi}(X) = \hat{\Omega}^\dagger(X)\hat{\Omega}(X) = \sum_{n_s=0}^{\infty}|\Omega(X,n_s)|^2|n_s\rangle_{s\,s}\langle n_s| \tag{44}$$

is the positive operator-valued measure (POVM) [31] for this measurement, and

$$\Omega(X,n_s) = {}_p\langle X,\zeta|\exp\left[\frac{i\Gamma_S}{2}\hat{n}_p(\hat{n}_p-1) + i\Gamma_X n_s\hat{n}_p\right]|\alpha\rangle_p. \tag{45}$$

The explicit form of the reduction operator $\hat{\Omega}(X)$ for the asymptotic case of (27) is calculated in Appendix C, see Eq. (C6). It follows from this result, that the conditional probability distribution of X for a given number of quanta in the signal mode n is equal to

$$|\Omega(X,n)|^2 = \frac{1}{\sqrt{2\pi(\Delta X)^2}}\exp\left[-\frac{(X-\sqrt{2}\alpha\cos\varphi)^2}{2(\Delta X)^2}\right], \tag{46}$$

with the values of $(\Delta X)^2$ and φ are given by Eqs. (28a) and (B1).

The *a posteriori* probability distribution for n_s conditioned on the measured value of X, can be obtained form Eq. (46) using Bayes' theorem,

$$W_{\rm apost}(n_s|X) = \frac{1}{\mathcal{W}(X)}|\Omega(X,n_s)|^2 W_{\rm apr}(n_s), \tag{47}$$

where $W_{\rm apr}$ is the *a priori* probability distribution and

$$\mathcal{W}(X) = \sum_{n=0_s}^{\infty}|\Omega(X,n)|^2 W_{\rm apr}(n_s) \tag{48}$$

is the normalization factor, equal to the unconditional probability distribution for X.

It is interesting to note, that the function (46) is Gaussian in X, but non-Gaussian in n_s due to the dependence of φ on n_s, see Eq. (29). Therefore, the probability distribution (47) is also non-Gaussian (see also the brief discussion in the end of Appendix B).

In Fig. 1, the probability distribution is plotted as a function of n_s. That picture was plotted using the distribution (47) and assuming the condition (36). The initial probability distribution, which we assume to be the Poissonian one (corresponding to the coherent initial state of the signal

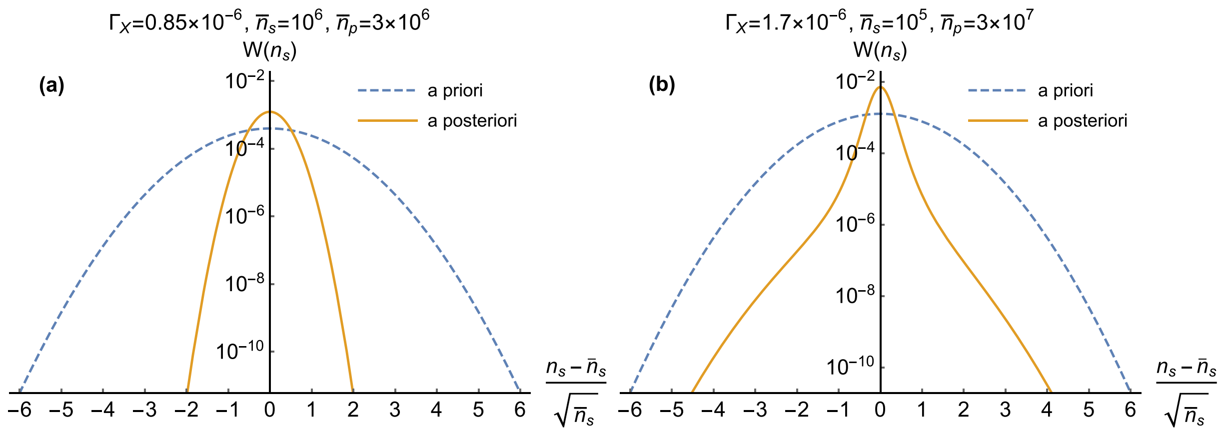

FIG. 1. The *a priori* (49) and *a posteriori* (47) probability distributions for the number of quanta in the signal mode. (a) The parameters, discussed in Sec. VII. (b) Slightly increased values of Γ_X, $\bar{n}_p$, giving the non-Gaussian shape of the *a posteriori* distribution. In both cases, the initial quantum state of the signal mode is assumed to be a coherent one.

mode),

$$W_{\rm apr}(n) = \frac{e^{-\bar{n}_s}\bar{n}_s^n}{n!} \tag{49}$$

is also shown for the comparison. Panel (a) illustrates the result of the QND measurement with the parameters close to the realistic experimental values discussed in Sec. VII. In panel (b) we used higher values of Γ_X and $\bar{n}_p$ which give non-Gaussian shape of the *a posteriori* distribution.

VII. DISCUSSION

Let us evaluate the efficiency and requirements of the QND measurements performed with high-Q WGM resonators. The factors Γ_X and Γ_S [see Eq. (6)] for the XPM and SPM effects based on the electronic nonlinearity of the material can be estimated as

$$\Gamma_X = 2\Gamma_S = 2Q_{\rm load}\frac{n_2}{n_0}\frac{\hbar\omega_0 c}{V_{\rm eff}}, \tag{50}$$

where c is the speed of light, ω_0 is the optical frequency, n_0 is the refractive index of the material, n_2 is the cubic nonlinearity coefficient, $V_{\rm eff}$ is the effective volume of the mode, and $Q_{\rm load} = \omega_0\tau$ is the loaded quality factor. Note that one of the factors, which constitute the unified quantum efficiency η, is equal to

$$\eta_{\rm load} = 1 - \frac{Q_{\rm load}}{Q_{\rm intr}}, \tag{51}$$

where $Q_{\rm intr}$ is the intrinsic quality factor. Therefore, in order to overcome the SQL by a significant margin, $Q_{\rm load}$ should be smaller that $Q_{\rm intr}$ by one to two orders of magnitude.

We select CaF_2 as the resonator host material in which the highest quality factor $Q_{\rm intr} = 3\times 10^{11}$ was achieved so far [23]. We assume the that the (vacuum) wavelengths are close to $\lambda = 2\pi c/\omega_0 = 1.55\ \mu$m for both the signal and the probe modes. At this wavelength, CaF_2 is characterized by the refractive index $n_0 = 1.44$ and the nonlinearity factor $n_2 = 3.2\times 10^{-20}$ m^2/W. We assume also that the resonator has 100 μm in diameter. The circumference of the resonator is shaped in a sharp edge resulting in $2\times 3\ \mu$m^2 mode cross section and, correspondingly, $V_{\rm eff} \simeq 2\times 10^{-15}$ m^3 mode volume.

For these parameters, the factors Γ_X and Γ_S can be estimated as follows:

$$\Gamma_X = 2\Gamma_S \approx 0.85\times 10^{-6}\times\frac{Q_{\rm load}}{10^9}. \tag{52}$$

For a reasonably optimistic value of $\eta = 0.9$, these parameters translate to the following values of the optimal number of the probe quanta and the corresponding measurement error, see Eqs. (37) and (38):

$$\bar{n}_p^{\rm opt} \approx 4\times 10^6\times\frac{10^9}{Q_{\rm load}}, \tag{53}$$

$$\left(\Delta n_{s,\min}^{\rm opt}\right)^2 \approx 2\times 10^5\times\frac{10^9}{Q_{\rm load}}. \tag{54}$$

The pump power, which is necessary to excite the intracavity number of quanta (53), can be estimated as follows:

$$P_p = \frac{\hbar\omega_0^2\bar{n}_p}{2Q_{\rm load}} \approx 0.3\ \mu{\rm W}\times\left(\frac{10^9}{Q_{\rm load}}\right)^2. \tag{55}$$

This estimate shows that we overcome the SQL by almost one order of magnitude using a several times smaller number of the signal photons (for example, $\bar{n}_s \approx 10^6$), than the number of probe photons.

Possible conceptual implementations of the proposed measurement are illustrated by Fig. 2(a). We assume that wavelengths of the probe and signal waves are dissimilar enough to allow their separation outside of the resonator. Probe wave (laser output in a coherent state) is injected to the resonator by means of a prism coupler. Part of the laser output is optimally phase shifted and used as a reference. The signal wave is coupled to the resonator using the same prism. A quadrature component of the probe emitted from the resonator is measured using a balanced homodyne detector. In order to preserve quantum states of the signal and the probe the resonator modes are overcoupled. Alternatively, classical probe wave can be injected through an additional weakly coupled channel, see Fig. 2(b). This later approach is preferable

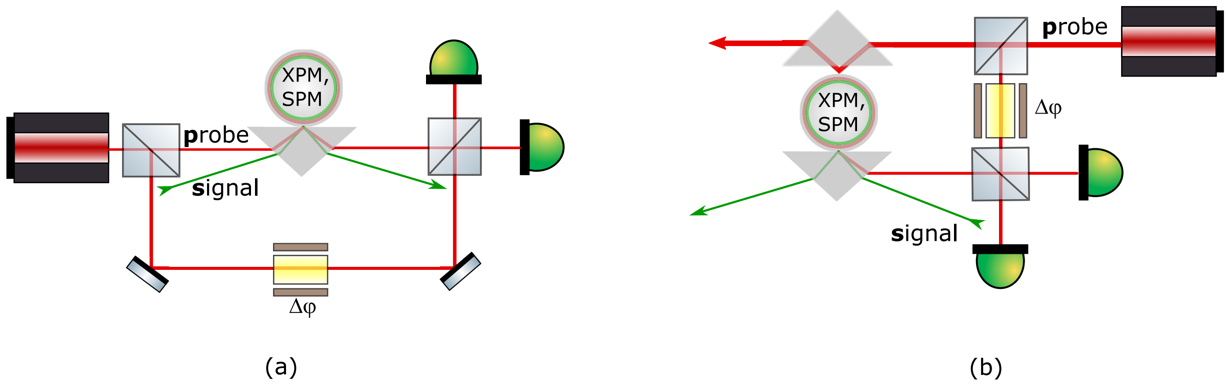

FIG. 2. Possible experimental implementations of QND detection using WGM monolithic microresonators equipped with (a) a single evanescent field coupler and (b) two couplers. The probing wave can be excited either in the same and in the opposite direction with the signal.

because the resonator filters out the unwanted frequency as well as spatial components from the probe.

VIII. CONCLUSION

Sensing an optical signal by means of a probe wave via cross-phase modulation in a nonlinear media paves a way to a nonabsorbing measurement of the signal quanta number with the precision beating under certain conditions the standard quantum limit. In this paper we obtained an exact solution to the Heisenberg nonlinear equations of motion of the system and presented the careful formulation of the important from the practical point of view linearized approximation of this solution. We found the explicit form of the conditional final state of the signal mode and shown that it could have a non-Gaussian shape. Our analysis indicates that the self-phase modulation does not limit the measurement accuracy. Though these results are applicable to various systems both resonant and nonresonant measurable response can be obtained if parameter Γ_X is reasonably high [see Eq. (15).] This could limit the sensitivity because the nonlinearity of transparent materials is low whereas the measurement time τ is limited by the losses. However, the state-of-the-art whispering gallery microresonators with Kerr nonlinearity and record quality factor may overcome this limitation.

We found that using a WGM resonator having 100 μm in diameter, made of pure calcium fluorite, and interrogated with microwatt level probe wave in the 1.5-μm telecom band allows us to obtain the sensitivity several times smaller than the limit $\sqrt{n_s}$ if the mean number of quanta n_s exceeds about 10^6.

It is worth noting that the progress of complementary metal-oxide semiconductor-foundry fabrication of Si_3N_4 microresonators can make them an alternative platform for the QND measurement. Whereas the best-achieved $Q \sim 10^9$ [25,32] is still inferior to crystalline resonators, it was improved by orders of magnitude during the past 3 yr.

ACKNOWLEDGMENTS

The work of S.N.B., F.Y.K., and I.A.B. was supported by the Russian Science Foundation (Project No. 20-12-00344). The work of S.N.B. was also supported by the Foundation for the Advancement of Theoretical Physics and Mathematics BASIS. The research performed by A.B.M., D.V.S., and V.S.I was carried out at the Jet Propulsion Laboratory, California Institute of Technology under a contract with the National Aeronautics and Space Administration (Contract No. 80NM0018D0004).

APPENDIX A: CALCULATION OF THE MEASUREMENT ERROR

The straightforward calculation gives that for the initial state of the two-mode system, equal to $|\alpha\rangle_p \otimes |n\rangle_s$, the first two momenta of X_ζ are equal to

$$\langle \hat{X}_\zeta \rangle = \sqrt{2}\alpha\ \mathrm{Re}[e^{i(\Gamma_X n_s+\zeta)}{}_p\langle\alpha|e^{i\Gamma_S \hat{n}_p}|\alpha\rangle_p], \tag{A1a}$$

$$\langle \hat{X}_\zeta^2 \rangle = \alpha^2\ \mathrm{Re}[e^{2i(\Gamma_X n_s+\zeta+\Gamma_S/2)}{}_p\langle\alpha|e^{2i\Gamma_S \hat{n}_p}|\alpha\rangle_p] + \alpha^2 + \tfrac{1}{2}. \tag{A1b}$$

Taking into that for any factor λ,

$$\langle\alpha|e^{i\lambda\hat{n}}|\alpha\rangle = \exp[|\alpha|^2(e^{\lambda}-1)], \tag{A2}$$

see Ref. [33], we obtain

$$\langle \hat{X}_\zeta \rangle = \sqrt{2}\alpha E_1 \cos\varphi, \tag{A3a}$$

$$\langle \hat{X}_\zeta^2 \rangle = \alpha^2 E_2 \cos(2\varphi+\Delta) + \alpha^2 + \tfrac{1}{2}, \tag{A3b}$$

where

$$E_1 = \exp[\alpha^2(\cos\Gamma_S - 1)], \tag{A4a}$$

$$E_2 = \exp[\alpha^2(\cos 2\Gamma_S - 1)], \tag{A4b}$$

$$\varphi = \alpha^2 \sin\Gamma_S + \Gamma_X n_s + \zeta, \tag{A4c}$$

$$\Delta = \alpha^2(\sin 2\Gamma_S - 2\sin\Gamma_S) + \Gamma_S. \tag{A4d}$$

Therefore,

$$G = -\sqrt{2}\alpha\Gamma_X E_1 \sin\varphi, \tag{A5a}$$

$$(\Delta X_\zeta)^2 = \langle \hat{X}_\zeta^2 \rangle - \langle \hat{X}_\zeta \rangle^2 = A + B\cos 2\varphi - C\sin 2\varphi, \tag{A5b}$$

$$(\Delta n_s)^2 = \frac{(\Delta X_\zeta)^2}{G^2} \tag{A5c}$$

$$= \frac{1}{2\alpha^2\Gamma_X^2 E_1^2}[(A+B)\cot^2\varphi - 2C\cot\varphi + A - B], \tag{A5d}$$

where

$$A = \tfrac{1}{2} + \alpha^2\left(1 - E_1^2\right), \tag{A6a}$$

$$B = \alpha^2\left(E_2 \cos\,\Delta - E_1^2\right), \tag{A6b}$$

$$C = \alpha^2 E_2 \sin\,\Delta. \tag{A6c}$$

In the asymptotic case (27),

$$A \to \tfrac{1}{2} + \alpha^4\Gamma_S^2,\ B \to -\alpha^4\Gamma_S^2,\ C \to \alpha^2\Gamma_S, \tag{A7}$$

which results in Eqs. (28) and (29).

APPENDIX B: ACCOUNTING FOR THE INITIAL UNCERTAINTY OF n_s

It follows from Eqs. (29) and (36b), that

$$\varphi = \bar{\varphi} + \Gamma_X(n_s - \bar{n}_s). \tag{B1}$$

Let us assume that

$$\Gamma_X \sim \Gamma_S, \qquad \bar{n}_s \lesssim \bar{n}_p. \tag{B2}$$

Taking into account the approximation (27) and the inequality (35), one can obtain

$$\Gamma_X|n_s - \bar{n}_s| \ll 1. \tag{B3}$$

Therefore,

$$\cot\,\varphi \approx \cot\,\bar{\varphi} - \frac{\Gamma_X(n_s - \bar{n}_s)}{\sin^2\bar{\varphi}} = 2\eta\Gamma_S\bar{n}_p - \epsilon, \tag{B4}$$

where

$$\epsilon = \left(1 + 4\eta^2\Gamma_S^2\bar{n}_p^2\right)\Gamma_X(n_s - \bar{n}_s). \tag{B5}$$

Substitution of (B4) into Eq. (31) leads to

$$(\Delta n_s)^2 = \frac{1}{\Gamma_X^2}\left[\frac{1+\epsilon^2}{4\eta\bar{n}_p} + (1-\eta)\Gamma_S^2\bar{n}_p\right]. \tag{B6}$$

If the number of the probe mode quanta is equal to the optimal one (37), then Eq. (B5) reduces to

$$\epsilon = \frac{\Gamma_X(n_s - \bar{n}_s)}{1-\eta}. \tag{B7}$$

For the values of Γ_X, $\hat{n}_s$, and η, introduced in Sec. VII, this corresponds to

$$\epsilon \approx 0.01\frac{n_s - \bar{n}_s}{\sqrt{\bar{n}_s}}. \tag{B8}$$

Therefore, if the assumptions (27), (35), and (B2) are fulfilled and the initial quantum state of the signal mode is close to the coherent one, then $\epsilon^2 \ll 1$, which reduces Eqs. (B6) to (33). However, if the measured value n_s deviates strongly form the mean value $\bar{n}_s$, then the factor ϵ becomes significant, introducing dependence on n_s into Δn_s and, therefore, making the *a posteriori* distribution (47) non-Gaussian.

APPENDIX C: CALCULATION OF STATISTICS OF THE MEASUREMENT RESULTS

Absorbing the regular phase shifts into $|X,\zeta\rangle$, the kernel (45) can be presented as follows:

$$\Omega(X,n) = {}_p\langle X,\varphi|\exp\left\{\frac{i\Gamma_S}{2}[\hat{n}_p(\hat{n}_p - 1) - 2\alpha^2\hat{n}_p]\right\}|\alpha\rangle, \tag{C1}$$

where

$$|X,\varphi\rangle_p = e^{-i(\Gamma_S\alpha^2 + \Gamma_X n)\hat{n}_p}|X,\zeta\rangle_p, \tag{C2}$$

and φ is given by Eq. (29). Then rewrite it as follows:

$$\Omega(X,n) = {}_p\langle X,\varphi|\hat{\mathcal{D}}(\alpha)\exp\left\{\frac{i\Gamma_S}{2}[(\hat{a}_p^\dagger + \alpha)^2(\hat{a}_p + \alpha)^2 - 2\alpha^2(\hat{a}_p^\dagger + \alpha)(\hat{a}_p + \alpha)]\right\}\hat{\mathcal{D}}^\dagger(\alpha)|\alpha\rangle_p, \tag{C3}$$

where $\hat{\mathcal{D}}$ is the unitary displacement operator defined by

$$\hat{\mathcal{D}}^\dagger(\alpha)\hat{a}_p\hat{\mathcal{D}}(\alpha) = \hat{a}_p + \alpha. \tag{C4}$$

Assume the approximation (27). In this case, omitting the nonphysical factor $e^{-i\Gamma_S\alpha^4/2}$, we obtain

$$\Omega(X_\zeta,n) \approx {}_p\langle X,\varphi|\mathcal{D}(\alpha)\exp\left[\frac{i\Gamma_S\alpha^2}{2}(\hat{a}_p + \hat{a}_p^\dagger)^2\right]\mathcal{D}^\dagger(\alpha)|\alpha\rangle_p = {}_p\langle X,\varphi|e^{i\Gamma_S\alpha^2(\hat{X}-\alpha\sqrt{2})^2}|\alpha\rangle_p. \tag{C5}$$

The kernel $\Omega(X_\zeta,n)$ can be calculated explicitly using the position representation,

$$\begin{aligned}\Omega(X_\zeta,n) &= \int_{-\infty}^{\infty} {}_p\langle X,\varphi|X\rangle_p e^{i\Gamma_S\alpha^2(X-\alpha\sqrt{2})^2}{}_p\langle X|\alpha\rangle_p\,dX\\ &= \frac{1}{\sqrt{\pi^{1/2}(\kappa + i\,\cot\,\varphi)|\sin\,\varphi|}}\exp\left[\frac{1}{\kappa + i\cot\varphi}\left(-\frac{1+i\kappa\,\cot\varphi}{2}X_\varphi^2 + \frac{\sqrt{2}i\alpha\kappa}{\sin\,\varphi}X_\varphi - i\alpha^2\kappa\,\cot\,\varphi\right)\right],\end{aligned} \tag{C6}$$

where

$$\kappa = 1 - 2i\Gamma_S\alpha^2, \tag{C7}$$

see Ref. [34], Sec. 4.4.2.

APPENDIX D: SENSITIVITY OF THE OPEN SYSTEM

A practical implementation of the QND involves detection of the probe photons outside of the resonator mode as well as attenuation of the pump and probe photons due to imperfection of the system. One has to utilize a constant external pumping of the mode to support the constant number of the probe photons. The impact of these processes can be estimated using the Langevin approach.

Since the exact solution of the problem is rather involved, we introduce several simplifications. First, we expect that the signal mode is isolated from the environment in such a way that the decay from the mode can be neglected during the measurement time. This is reasonable as the signal mode can have a much higher Q factor than the loaded probe mode, and the value of the intrinsic Q factor of a WGM resonator can exceed 10^{11}, whereas the loaded Q factor can be as small as 10^9. Second, considering the probe mode due to the same difference between the intrinsic and loaded Q factors, we assume that the attenuation due to the controlled coupling with the external pump prevails over the unwanted attenuation of the probe. Third, we expect that the pump and the signal have significantly dissimilar frequencies so that the signal is not impacted by the probe coupling. This is reasonable as the coupling element can be optimized for a particular wavelength. Accounting for the aforesaid, we can assume that $\hat{n}_s =$ const, and we have to consider dynamics of the probe wave only.

Taking the assumptions listed above into account and using the rotating with the external pump frequency ω_c picture, equation of motion for the probe field can be presented as follows [compare with Eq. (20a)]:

$$\dot{\hat{a}}_p(t) = -[\kappa + i(\omega_p - \omega_c)]\hat{a}_p(t) + i\gamma_S\hat{a}_p^\dagger(t)\hat{a}_p^2(t) + i\gamma_X\hat{n}_s\hat{a}_p(t) + \sqrt{2\kappa_c}A + \sqrt{2\kappa}\hat{a}_{\rm in}(t). \tag{D1}$$

Here A and ω_c are the amplitude and the frequency of the external pump,

$$\kappa = \kappa_c + \kappa_l \tag{D2a}$$

is the the half-bandwidth of the probe mode,

$$\hat{a}_{\rm in} = \frac{1}{\sqrt{2\kappa}}(\sqrt{2\kappa_c}\hat{a}_c + \sqrt{2\kappa_l}\hat{a}_l) \tag{D2b}$$

is the corresponding Langevine force, κ_c and κ_l are the components of κ imposed, respectively, by the coupling with the transmission line and the losses, $\hat{a}_c$ and $\hat{a}_l$ are the corresponding components of $\hat{a}_{\rm in}$.

We present the operators as sums of large classical components and small zero mean value quantum uncertainties, in particular,

$$\hat{a}_p := \alpha + \hat{a}_p, \tag{D3a}$$

$$\hat{n}_s := N_s + \hat{n}_s, \tag{D3b}$$

with

$$\gamma_S\sqrt{\langle\hat{n}_s^2\rangle} \ll \kappa. \tag{D4}$$

Without loss of generality, we assume validity of the condition (26).

Using Eq. (D1) we derive for α,

$$\alpha(\kappa + i\delta - i\gamma_X N_s) = \sqrt{2\kappa_c}A. \tag{D5}$$

Here,

$$\delta = \omega_p - \omega_c - i\gamma_S\bar{n}_p \tag{D6}$$

is the probe mode frequency detuning.

In the linear approximation with respect of the noise terms of the probe light we get, using the Fourier picture,

$$L(\Omega)\hat{a}(\Omega) - i\gamma_S\alpha^2\hat{a}^\dagger(-\Omega) = \sqrt{2\kappa}\hat{a}_{\rm in}(\Omega) + i\gamma_X\alpha\hat{n}_s, \tag{D7}$$

where

$$L(\Omega) = \kappa + i\delta - i\Omega - i\gamma_S\bar{n}_p. \tag{D8}$$

The solution to this equation is

$$\hat{a}(\Omega) = \frac{1}{D(\Omega)}\{\sqrt{2\kappa}[L^*(-\Omega)\hat{a}_{\rm in}(\Omega) + i\gamma_S\bar{n}_p\hat{a}_{\rm in}^\dagger(-\Omega)] + i\gamma_X\alpha[L^*(-\Omega) - i\gamma_S\bar{n}_p]\hat{n}_s\}, \tag{D9}$$

where

$$D(\Omega) = L(\Omega)L^*(-\Omega) - \gamma_S^2\bar{n}_p^2. \tag{D10}$$

Correspondingly, for the output light we obtain

$$\hat{a}_{\rm out}(\Omega) = \sqrt{2\kappa_c}\hat{a}_p(\Omega) - \hat{a}_c(\Omega) = \sqrt{\eta}\hat{a}_{\rm out\,0}(\Omega) + \sqrt{1-\eta}\hat{f}(\Omega), \tag{D11}$$

where

$$\eta = \frac{\kappa_c}{\kappa} \tag{D12}$$

is the quantum efficiency,

$$\sqrt{2\kappa}\hat{f} = \sqrt{2\kappa_c}\hat{a}_l - \sqrt{2\kappa_l}\hat{a}_c \tag{D13}$$

is the effective noise, and

$$\hat{a}_{\rm out\,0}(\Omega) = \sqrt{2\kappa}\hat{a}_p(\Omega) - \hat{a}_{\rm in}(\Omega) = \frac{1}{D(\Omega)}\big\{\big[L^*(\Omega)L^*(-\Omega) + \gamma_S^2\bar{n}_p^2\big]\hat{a}_{\rm in}(\Omega) + 2i\kappa\gamma_S\bar{n}_p\hat{a}_{\rm in}^\dagger(\Omega) + i\sqrt{2\kappa}\gamma_X\alpha[L^*(-\Omega) - i\gamma_S\bar{n}_p]\hat{n}_s\big\} \tag{D14}$$

is the value of $\hat{a}_{\rm out}$ in the absence of losses ($\eta = 0$). Note that if $\hat{a}_c$ and $\hat{a}_l$ correspond to two independent vacuum fields, then the same is true for $\hat{a}_{\rm in}$ and $\hat{f}$.

Taking into account that within our approximation $\hat{n}_s =$ const, the optimal data processing procedure is the DC filtration of the output signal over the time $\gg 1/\kappa$, which corresponds to the frequency band $|\Omega| \ll \kappa$. We suppose also that the pump is tuned in resonance with the modified eigenfrequency of the probe mode,

$$\delta = 0. \tag{D15}$$

In this case, Eq. (D14) simplifies to

$$\hat{a}_{\rm out\,0}(\Omega) = (1 + i\Gamma_S\bar{n}_p)\hat{a}_{\rm in}(\Omega) + i\Gamma_S\bar{n}_p\hat{a}_{\rm in}^\dagger(\Omega) + i\sqrt{\frac{\kappa}{2}}\Gamma_X\alpha\hat{n}_s, \tag{D16}$$

where

$$\Gamma_{S,X} = \frac{2\gamma_{S,X}}{\kappa}. \tag{D17}$$

Let the quadrature of $\hat{a}_{\rm out}$ defined by the homodyne angle ζ,

$$\hat{a}^{\zeta}_{\rm out}(\Omega) = \frac{1}{\sqrt{2}}[\hat{a}_{\rm out}(\Omega)e^{i\zeta} + \hat{a}^{\dagger}_{\rm out}(-\Omega)e^{-i\zeta}] = \hat{a}^{\zeta}_{\rm out\,noise}(\Omega) + G\hat{n}_s \tag{D18}$$

is monitored by the homodyne detector. Here

$$\hat{a}^{\zeta}_{\rm out\,noise}(\Omega) = \sqrt{\eta}\big[(\cos\zeta - 2\Gamma_S\bar{n}_p \sin\zeta)\hat{a}^{c}_{\rm in}(\Omega) - \hat{a}^{s}_{\rm in}(\Omega)\sin\zeta\big] + \sqrt{1-\eta}\hat{f}^{\zeta}(\Omega) \tag{D19}$$

is the noise part of $\hat{a}^{\zeta}_{\rm out}$,

$$G = -\sqrt{\eta\kappa\bar{n}_p}\Gamma_X \sin\zeta \tag{D20}$$

is the gain factor,

$$\hat{a}^{c}_{\rm in}(\Omega) = \frac{\hat{a}_{\rm in}(\Omega) + \hat{a}^{\dagger}_{\rm in}(-\Omega)}{\sqrt{2}},\ \hat{a}^{s}_{\rm in}(\Omega) = \frac{\hat{a}_{\rm in}(\Omega) - \hat{a}^{\dagger}_{\rm in}(-\Omega)}{i\sqrt{2}} \tag{D21}$$

are the cosine and sine quadratures of $\hat{a}_{\rm in}$ and

$$\hat{f}^{\zeta}(\Omega) = \frac{1}{\sqrt{2}}[\hat{f}(\Omega)e^{i\zeta} + \hat{f}^{\dagger}(-\Omega)e^{-i\zeta}]. \tag{D22}$$

Taking into account our approximations, $\hat{a}^{\zeta}_{\rm out\,noise}$ is a white noise with the spectral density, we obtain

$$S = \tfrac{1}{2}\big(1 - 4\eta\Gamma_S\bar{n}_p\cos\zeta\ \sin\zeta + 4\eta\Gamma_S^2\bar{n}_p^2\sin^2\zeta\big). \tag{D23}$$

Therefore, the measurement error is equal to

$$(\Delta n_s)^2 = \frac{S}{\tau G^2} = \frac{1}{2\kappa\tau\Gamma_X^2}\left[\frac{1 + (\cot\zeta - 2\eta\Gamma_S\bar{n}_p)^2}{4\eta\bar{n}_p} + (1-\eta)\Gamma_S^2\bar{n}_p\right], \tag{D24}$$

where $\tau \gg 1/\kappa$ is the measurement time. Comparing Eq. (D24) with Eq. (31) we conclude that the Langevin approach presented in this Appendix results in the same expression as was derived in the main body of the paper.

[1] V. B. Braginsky and S.P. Vyatchanin, Nondestructive measurement of the energy of optical quanta, Sov. Phys. Dokl. **26**, 686 (1981).

[2] G. J. Milburn and D. F. Walls, Quantum nondemolition measurements via quadratic coupling, Phys. Rev. A **28**, 2065 (1983).

[3] M. D. Levenson, R. M. Shelby, M. Reid, and D. F. Walls, Quantum Nondemolition Detection of Optical Quadrature Amplitudes, Phys. Rev. Lett. **57**, 2473 (1986).

[4] V. B.Braginsky, M. L.Gorodetsky, and V. S. Ilchenko, Quality-factor and nonlinear properties of optical whispering-gallery modes, Phys. Lett. A **137**, 393 (1989).

[5] Y.-F. Xiao, Ş. K. Özdemir, V. Gaddam, C.-H. Dong, N. Imoto, and L. Yang, Quantum nondemolition measurement of photon number via optical kerr effect in an ultra-high-q microtoroid cavity, Opt. Express **16**, 21462 (2008).

[6] T. Tyc and N. Korolkova, Highly non-gaussian states created via cross-kerr nonlinearity, New J. Phys. **10**, 023041 (2008).

[7] J. von Neumann, *Mathematical Foundations of Quantum Mechanics: New Edition*, Princeton Landmarks in Mathematics and Physics (Princeton University Press, Princeton, 2018).

[8] H. Heffner, The fundamental noise limit of linear amplifiers, Proc. IRE **50**, 1604 (1962).

[9] D. Bohm, *Quantum Theory* (Prentice-Hall, New York, 1951), p. 646.

[10] V. B. Braginsky and Y. I. Vorontsov, Quantum-mechanical limitations in macroscopic experiments and modern experimental technique, Sov. Phys. Usp. **17**, 644 (1975).

[11] V. B. Braginsky, Y. I. Vorontsov, and K. S. Thorne, Quantum nondemolition measurements, Science **209**, 547 (1980).

[12] V. B. Braginsky, Y. I. Vorontsov, and F. Y. Khalili, Optimal quantum measurements in detectors of gravitation radiation, JETP Lett. **27**, 276 (1978).

[13] W. G. Unruh, Analysis of quantum-nondemolition measurement, Phys. Rev. D **18**, 1764 (1978).

[14] V. B. Braginskii and F. Ya. Khalili, Optico-magnetic effects in nondestructive quantum counting, Sov. Phys. JETP **51**, 859 (1980).

[15] J. F. Roch, G. Roger, P. Grangier, J.-M. Courty, and S. Reynaud, Quantum non-demolition measurements in optics: a review and some recent experimental results, Appl. Phys. B: Photophys. Laser Chem. **55**, 291 (1992).

[16] V. B. Braginsky and F. Y. Khalili, Quantum nondemolition measurements: the route from toys to tools, Rev. Mod. Phys. **68**, 1 (1996).

[17] P. Grangier, J. A. Levenson, and J.-P. Poizat, Quantum nondemolition measurements in optics, Nature (London) **396**, 537 (1998).

[18] J. M. Raimond, M. Brune, and S. Haroche, Manipulating quantum entanglement with atoms and photons in a cavity, Rev. Mod. Phys. **73**, 565 (2001).

[19] A. Reiserer, and G. Rempe, Cavity-based quantum networks with single atoms and optical photons, Rev. Mod. Phys. **87**, 1379 (2015).

[20] D. Niemietz, P. Farrera, S. Langenfeld, and G. Rempe, Nondestructive detection of photonic qubits, Nature (London) **591**, 570 (2021).

[21] X. Gu, A. F. Kockum, A. Miranowicz, Y. x. Liu, and F. Nori, Microwave photonics with superconducting quantum circuits, Phys. Rep. **718-719**, 1 (2017), microwave photonics with superconducting quantum circuits.

[22] D. V. Strekalov, C. Marquardt, A. B. Matsko, H. G. L. Schwefel, and G. Leuchs, Nonlinear and quantum optics with whispering gallery resonators, J. Opt. **18**, 123002 (2016).

[23] A. A. Savchenkov, A. B. Matsko, V. S. Ilchenko, and L. Maleki, Optical resonators with ten million finesse, Opt. Express **15**, 6768 (2007).

[24] T. J. Kippenberg, S. M. Spillane, and K. J. Vahala, Kerr-Nonlinearity Optical Parametric Oscillation in an Ultrahigh-*q* Toroid Microcavity, Phys. Rev. Lett. **93**, 083904 (2004).

[25] L. Wu, H. Wang, Q. Yang, Q. xin Ji, B. Shen, C. Bao, M. Gao, and K. Vahala, Greater than one billion q factor for on-chip microresonators, Opt. Lett. **45**, 5129 (2020).

[26] N. Imoto, H. A. Haus, and Y. Yamamoto, Quantum nondemolition measurement of the photon number via the optical kerr effect, Phys. Rev. A **32**, 2287 (1985).

[27] J.-F. Roch, K. Vigneron, P. Grelu, A. Sinatra, J.-P. Poizat, and P. Grangier, Quantum Nondemolition Measurements using Cold Trapped Atoms, Phys. Rev. Lett. **78**, 634 (1997).

[28] A. B. Matsko, V. V. Kozlov, and M. O. Scully, Backaction Cancellation in Quantum Nondemolition Measurement of Optical Solitons, Phys. Rev. Lett. **82**, 3244 (1999).

[29] P. D. Drummond, J. Breslin, and R. M. Shelby, Quantum-Nondemolition Measurements with Coherent Soliton Probes, Phys. Rev. Lett. **73**, 2837 (1994).

[30] P. Carruthers, and M. M. Nieto, Phase and angle variables in quantum mechanics, Rev. Mod. Phys. **40**, 411 (1968).

[31] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic, New York, 1976), p. 309.

[32] W. Jin, Q.-F. Yang, L. Chang, B. Shen, H. Wang, M. A. Leal, L. Wu, M. Gao, A. Feshali, M. Paniccia *et al.*, Hertz-linewidth semiconductor lasers using cmos-ready ultra-high-q microresonators, Nat. Photonics **15**, 346 (2021).

[33] W. H. Louisell, *Radiation and Noise in Quantum Electronics* (McGraw-Hill, New York, 1964), p. 398.

[34] W. Schleich, *Quantum Optics in Phase Space* (Wiley-VCH, Berlin, 2001), p. 695.