Optimal waveform estimation for classical and quantum systems via time-symmetric smoothing

Mankei Tsang*
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
(Received 29 June 2009; published 28 September 2009; publisher error corrected 30 September 2009)

Classical and quantum theories of time-symmetric smoothing, which can be used to optimally estimate wave forms in classical and quantum systems, are derived using a discrete-time approach, and the similarities between the two theories are emphasized. Application of the quantum theory to homodyne phase-locked loop design for phase estimation with narrowband squeezed optical beams is studied. The relation between the proposed theory and weak value theory of Aharonov *et al.* is also explored.

DOI: 10.1103/PhysRevA.80.033840 PACS number(s): 03.65.Ta, 03.65.Yz, 42.50.Dv

I. INTRODUCTION

Estimation theory is the science of determining the state of a system, such as a dice, an aircraft, or the weather in Boston, from noisy observations [1–4]. As shown in Fig. 1, estimation problems can be classified into four classes, namely, prediction, filtering, retrodiction, and smoothing. For applications that do not require real-time data, such as sensing and communication, smoothing is the most accurate estimation technique.

I have recently proposed a time-symmetric quantum theory of smoothing, which allows one to optimally estimate classical diffusive Markov random processes, such as gravitational waves or magnetic fields, coupled to a quantum system, such as a quantum mechanical oscillator or an atomic spin ensemble, under continuous measurements [5]. In this paper, I shall demonstrate in more detail the derivation of this theory using a discrete-time approach and how it closely parallels the classical time-symmetric smoothing theory proposed by Pardoux [6]. I shall apply the theory to the design of homodyne phase-locked loops (PLLs) for narrowband squeezed optical beams, as previously considered by Berry and Wiseman [7]. I shall show that their approach can be regarded as a special case of my theory and discuss how their results can be generalized and improved. I shall also discuss the weak value theory proposed by Aharonov *et al.* [8] in relation with the smoothing theory and how their theory may be regarded as a smoothing theory for quantum degrees of freedom. In particular, the smoothing quasiprobability distribution proposed in Ref. [5] is shown to naturally arise from the statistics of weak position and momentum measurements.

This paper is organized as follows. In Sec. II, Pardoux's classical time-symmetric smoothing theory is derived using a discrete-time approach, which is then generalized to the quantum regime for hybrid classical-quantum smoothing in Sec. III. Application of the hybrid classical-quantum smoothing theory to PLL design is studied in Sec. IV. The relation between the smoothing theory and weak value theory of Aharonov *et al.* is then discussed in Sec. V. Section VI concludes the paper and points out some possible extensions of the proposed theory.

*mankei@mit.edu

II. CLASSICAL SMOOTHING

A. Problem statement

Consider the classical smoothing problem depicted in Fig. 2. Let

$$x_t \equiv \begin{bmatrix} x_{1t} \\ x_{2t} \\ \vdots \\ x_{nt} \end{bmatrix} \tag{2.1}$$

be a vectoral diffusive Markov random process that satisfies the system Itō differential equation [1],

$$dx_t = A(x_t,t)dt + B(x_t,t)dW_t, \tag{2.2}$$

where dW_t is a vectoral Wiener increment with mean and covariance matrix given by

$$\langle dW_t \rangle = 0, \tag{2.3}$$

$$\langle dW_t dW_t^T \rangle = Q(t)dt. \tag{2.4}$$

The superscript T denotes the transpose. The vectoral observation process dy_t satisfies the observation Itō equation,

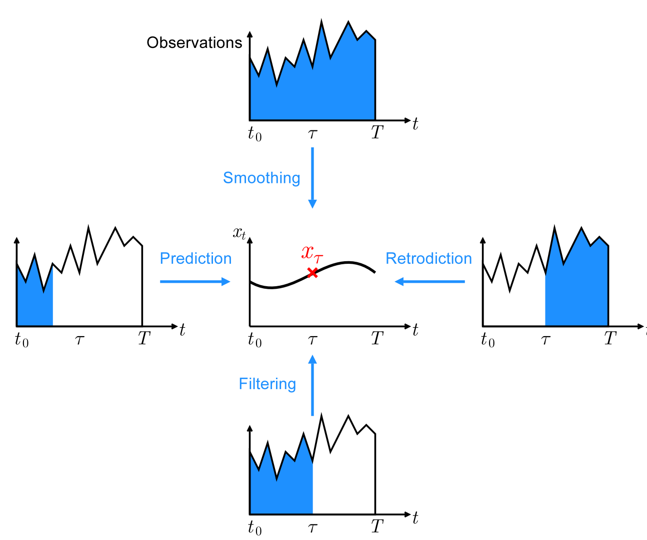

FIG. 1. (Color online) Four classes of estimation problems, depending on the observation time interval relative to τ, the time at which the signal is to be estimated.

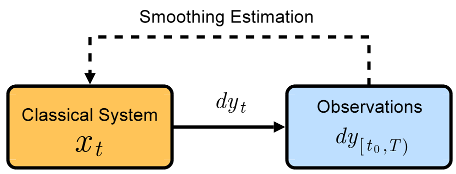

FIG. 2. (Color online) The classical smoothing problem.

$$dy_t = C(x_t,t)dt + dV_t, \tag{2.5}$$

where dV_t is another vectoral Wiener increment with mean and covariance matrix given by

$$\langle dV_t \rangle = 0, \tag{2.6}$$

$$\langle dV_t dV_t^T \rangle = R(t)dt. \tag{2.7}$$

For generality and later purpose, dW_t and dV_t are assumed to be correlated, with covariance

$$\langle dW_t dV_t^T \rangle = S(t)dt. \tag{2.8}$$

Define the observation record in the time interval $[t_1,t_2)$ as

$$dy_{[t_1,t_2)} \equiv \{dy_t, t_1 \le t < t_2\}. \tag{2.9}$$

The goal of smoothing is to calculate the conditional probability density of x_τ, given the observation record $dy_{[t_0,T)}$ in the time interval $t_0 \le \tau \le T$.

It is more intuitive to consider the problem in discrete time first. The discrete-time system and observation equations [Eqs. (2.2) and (2.5)] are

$$\delta x_t = A(x_t,t)\delta t + B(x_t,t)\delta W_t, \tag{2.10}$$

$$\delta y_t = C(x_t,t)\delta t + \delta V_t. \tag{2.11}$$

The observation record

$$\delta y_{[t_0,T-\delta t]} \equiv \{\delta y_{t_0}, \delta y_{t_0+\delta t}, \ldots, \delta y_{T-\delta t}\} \tag{2.12}$$

also becomes discrete. The covariance matrices for the increments are

$$\langle \delta W_t \delta W_t^T \rangle = Q(t)\delta t, \tag{2.13}$$

$$\langle \delta V_t \delta V_t^T \rangle = R(t)\delta t, \tag{2.14}$$

$$\langle \delta W_t \delta V_t^T \rangle = S(t)\delta t, \tag{2.15}$$

and the increments at different times are independent of one another. Because δW_t and δV_t are proportional to $\sqrt{\delta t}$, one should keep all linear *and* quadratic terms of the Wiener increments in an equation according to Itō calculus when taking the continuous time limit.

With correlated δW_t and δV_t, it is preferable, for technical reasons, to rewrite the system equation [Eq. (2.10)] as [2]

$$\delta x_t = A(x_t,t)\delta t + B(x_t,t)\delta W_t + D(x_t,t)[\delta y_t - C(x_t,t)\delta t - \delta V_t], \tag{2.16}$$

where $D(x_t,t)$ can be arbitrarily set because the expression in square brackets is zero. The system equation becomes

$$\delta x_t = [A(x_t,t) - D(x_t,t)C(x_t,t)]\delta t + D(x_t,t)\delta y_t + B(x_t,t)\delta W_t - D(x_t,t)\delta V_t. \tag{2.17}$$

The new system noise is

$$\delta Z_t \equiv B(x_t,t)\delta W_t - D(x_t,t)\delta V_t, \tag{2.18}$$

$$\langle \delta Z_t \delta Z_t^T \rangle = [B(x_t,t)Q(t)B^T(x_t,t) + D(x_t,t)R(t)D^T(x_t,t) - B(x_t,t)S(t)D^T(x_t,t) - D(x_t,t)S^T(t)B^T(x_t,t)]\delta t. \tag{2.19}$$

The covariance between the new system noise δZ_t and the observation noise δV_t is

$$\langle \delta Z_t \delta V_t^T \rangle = [B(x_t,t)S(t) - D(x_t,t)R(t)]\delta t \tag{2.20}$$

and can be made to vanish if one lets

$$D(x_t,t) = B(x_t,t)S(t)R^{-1}(t). \tag{2.21}$$

The new equivalent system and observation model is then

$$\delta x_t = A(x_t,t)\delta t + B(x_t,t)S(t)R^{-1}(t)[\delta y_t - C(x_t,t)\delta t] + B(x_t,t)\delta U_t, \tag{2.22}$$

$$\delta y_t = C(x_t,t)\delta t + \delta V_t, \tag{2.23}$$

with covariances

$$\langle \delta U_t \delta U_t^T \rangle = [Q(t) - S(t)R^{-1}S^T(t)]\delta t, \tag{2.24}$$

$$\langle \delta V_t \delta V_t^T \rangle = R(t)\delta t, \tag{2.25}$$

$$\langle \delta U_t \delta V_t^T \rangle = 0. \tag{2.26}$$

The new system and observation noises are now independent, but note that δx_t becomes dependent on δy_t.

B. Time-symmetric approach

According to the Bayes theorem, the smoothing probability density for x_τ can be expressed as

$$P(x_\tau | \delta y_{[t_0,T-\delta t]}) = \frac{P(\delta y_{[t_0,T-\delta t]} | x_\tau) P(x_\tau)}{P(\delta y_{[t_0,T-\delta t]})}, \tag{2.27}$$

$$P(\delta y_{[t_0,T-\delta t]}) = \int dx_\tau P(\delta y_{[t_0,T-\delta t]} | x_\tau) P(x_\tau), \tag{2.28}$$

where

$$\int dx_\tau \equiv \int dx_{1\tau} \ldots \int dx_{n\tau} \tag{2.29}$$

and $P(x_\tau)$ is the *a priori* probability density, which represents one's knowledge of x_τ absent any observation. Functions of x_τ are assumed to also depend implicitly on τ. Splitting $\delta y_{[t_0,T-\delta t]}$ into the past record

$$\delta y_{\text{past}} \equiv \delta y_{[t_0,\tau-\delta t]} \tag{2.30}$$

and the future record

$$\delta y_{\text{future}} \equiv \delta y_{[\tau, T-\delta t]} \tag{2.31}$$

relative to time τ, $P(\delta y_{[t_0,T)}|x_\tau)$ in Eq. (2.27) can be rewritten as

$$\begin{aligned} P(\delta y_{[t_0,T-\delta t]}|x_\tau) &= P(\delta y_{\text{past}}, \delta y_{\text{future}}|x_\tau) \\ &= P(\delta y_{\text{future}}|\delta y_{\text{past}}, x_\tau)P(\delta y_{\text{past}}|x_\tau). \end{aligned} \tag{2.32}$$

Because δV_t are independent increments, the future record is independent of the past record given x_τ and

$$P(\delta y_{[t_0,T-\delta t]}|x_\tau) = P(\delta y_{\text{future}}|x_\tau)P(\delta y_{\text{past}}|x_\tau). \tag{2.33}$$

Equation (2.27) becomes

$$\begin{aligned} P(x_\tau|\delta y_{[t_0,T-\delta t]}) &= \frac{P(\delta y_{\text{future}}|x_\tau)P(\delta y_{\text{past}}|x_\tau)P(x_\tau)}{\int dx_\tau(\text{numerator})} \\ &= \frac{P(\delta y_{\text{future}}|x_\tau)P(x_\tau|\delta y_{\text{past}})}{\int dx_\tau(\text{numerator})}. \end{aligned} \tag{2.34}$$

Thus, the smoothing density can be obtained by combining the filtering probability density $P(x_\tau|\delta y_{\text{past}})$ and a retrodictive likelihood function $P(\delta y_{\text{future}}|x_\tau)$.

C. Filtering

To derive an equation for the filtering probability density $P(x_\tau|\delta y_{\text{past}})$, first express $P(x_{t+\delta t}|\delta y_{[t_0,t]})$ in terms of $P(x_t|\delta y_{[t_0,t]})$ as

$$\begin{aligned} P(x_{t+\delta t}|\delta y_{[t_0,t]}) &= \int dx_t P(x_{t+\delta t}, x_t|\delta y_{[t_0,t]}) \\ &= \int dx_t P(x_{t+\delta t}|x_t, \delta y_{[t_0,t]})P(x_t|\delta y_{[t_0,t]}). \end{aligned} \tag{2.35}$$

$P(x_{t+\delta t}|x_t, \delta y_{[t_0,t]}) = P(x_{t+\delta t}|x_t, \delta y_t, \delta y_{[t_0,t-\delta t]})$ can be determined from the system [Eq. (2.22)] and is equal to $P(x_{t+\delta t}|x_t, \delta y_t)$ due to the Markovian nature of the system process. So

$$P(x_{t+\delta t}|\delta y_{[t_0,t]}) = \int dx_t P(x_{t+\delta t}|x_t, \delta y_t)P(x_t|\delta y_{[t_0,t]}), \tag{2.36}$$

which is a generalized Chapman-Kolmogorov equation [9]. $P(x_{t+\delta t}|x_t, \delta y_t)$ is

$$P(x_{t+\delta t}|x_t, \delta y_t) \propto \exp\Big\{-\tfrac{1}{2}\delta Z_t^T[B(x_t,t)Q(t)B^T(x_t,t)\delta t]^{-1}\delta Z_t\Big\}, \tag{2.37}$$

where

$$\delta Z_t \equiv x_{t+\delta t} - x_t - A(x_t,t)\delta t + B(x_t,t)S(t)R^{-1}(t)[\delta y_t - C(x_t,t)\delta t]. \tag{2.38}$$

Next, write $P(x_t|\delta y_{[t_0,t]})$ in terms of $P(x_t|\delta y_{[t_0,t-\delta t]})$ using the Bayes theorem as

$$\begin{aligned} P(x_t|\delta y_{[t_0,t]}) &= P(x_t|\delta y_{[t_0,t-\delta t]}, \delta y_t) \\ &= \frac{P(\delta y_t|x_t, \delta y_{[t_0,t-\delta t]})P(x_t|\delta y_{[t_0,t-\delta t]})}{\int dx_t(\text{numerator})} \\ &= \frac{P(\delta y_t|x_t)P(x_t|\delta y_{[t_0,t-\delta t]})}{\int dx_t(\text{numerator})}, \end{aligned} \tag{2.39}$$

where $P(\delta y_t|x_t, \delta y_{[t_0,t-\delta t]}) = P(\delta y_t|x_t)$ due to the Markovian property of the observation process. $P(\delta y_t|x_t)$ is determined by the observation equation [Eq. (2.23)] and given by

$$\begin{aligned} P(\delta y_t|x_t) \propto \exp\Big\{&-\tfrac{1}{2}[\delta y_t - C(x_t,t)\delta t]^T[R(t)\delta t]^{-1} \\ &\times[\delta y_t - C(x_t,t)\delta t]\Big\}. \end{aligned} \tag{2.40}$$

Hence, starting with the *a priori* probability density $P(x_{t_0})$, one can solve for $P(x_\tau|\delta y_{\text{past}})$ by iterating the formula

$$\begin{aligned} P(x_{t+\delta t}|\delta y_{[t_0,t]}) = \int & dx_t P(x_{t+\delta t}|x_t, \delta y_t) \\ &\times \frac{P(\delta y_t|x_t)P(x_t|\delta y_{[t_0,t-\delta t]})}{\int dx_t(\text{numerator})}. \end{aligned} \tag{2.41}$$

To obtain a stochastic differential equation for the filtering probability density, defined as

$$F(x,t) \equiv P(x_t = x|dy_{[t_0,t)}) \tag{2.42}$$

in the continuous time limit, one should expand Eq. (2.41) to first order with respect to δt and second order with respect to δy_t in a Taylor series, then apply the rules of Itō calculus. The result is the Kushner-Stratonovich (KS) equation [1,10], generalized for correlated system and observation noises by Fujisaki *et al.* [11], given by

$$\begin{aligned} dF = &- dt\sum_\mu \frac{\partial}{\partial x_\mu}(A_\mu F) + \frac{dt}{2}\sum_{\mu,\nu}\frac{\partial^2}{\partial x_\mu \partial x_\nu}[(BQB^T)_{\mu\nu}F] \\ &+ (C - \langle C\rangle_F)^T R^{-1} d\eta_t F - \sum_\mu \frac{\partial}{\partial x_\mu}[(BSR^{-1}d\eta_t)_\mu F], \end{aligned} \tag{2.43}$$

where

$$dF \equiv F(x,t+dt) - F(x,t), \tag{2.44}$$

$$\langle C\rangle_F \equiv \int dx C(x,t)F(x,t), \tag{2.45}$$

$$d\eta_t \equiv dy_t - dt\langle C\rangle_F. \tag{2.46}$$

The initial condition is

$$F(x,t_0) = P(x_{t_0}). \tag{2.47}$$

$d\eta_t$ is called the innovation process and is also a Wiener increment with covariance matrix $R(t)dt$ [11,12].

A linear stochastic equation for an unnormalized F is called the Duncan-Mortensen-Zakai (DMZ) equation [6,13], given by

$$df = -dt\sum_\mu \frac{\partial}{\partial x_\mu}(A_\mu f) + \frac{dt}{2}\sum_{\mu,\nu}\frac{\partial^2}{\partial x_\mu \partial x_\nu}[(BQB^T)_{\mu\nu}f] + C^T R^{-1}dy_t f - \sum_\mu \frac{\partial}{\partial x_\mu}[(BSR^{-1}dy_t)_\mu f], \tag{2.48}$$

where the normalization is

$$F(x,t) = \frac{f(x,t)}{\int dx f(x,t)}. \tag{2.49}$$

D. Retrodiction and smoothing

To solve for the retrodictive likelihood function $P(\delta y_{\text{future}}|x_\tau)$, note that

$$P(\delta y_{\text{future}}) = \int dx_\tau P(\delta y_{\text{future}}|x_\tau)P(x_\tau), \tag{2.50}$$

but $P(\delta y_{\text{future}})$ can also be expressed in terms of the multitime probability density as

$$P(\delta y_{[\tau,T-\delta t]}) = \int Dx_{[\tau,T]}P(x_{[\tau,T]},\delta y_{[\tau,T-\delta t]}), \tag{2.51}$$

where

$$x_{[\tau,T]} \equiv \{x_\tau, x_{\tau+\delta t}, \ldots, x_T\}, \tag{2.52}$$

$$\int Dx_{[\tau,T]} \equiv \int dx_\tau \int dx_{\tau+\delta t}\ldots\int dx_T. \tag{2.53}$$

The multitime density can be rewritten as

$$P(x_{[\tau,T]},\delta y_{[\tau,T-\delta t]}) = P(x_T|x_{[\tau,T-\delta t]},\delta y_{[\tau,T-\delta t]}) \times P(x_{[\tau,T-\delta t]},\delta y_{[\tau,T-\delta t]}). \tag{2.54}$$

Again using the Markovian property of the system process,

$$P(x_T|x_{[\tau,T-\delta t]},\delta y_{[\tau,T-\delta t]}) = P(x_T|x_{T-\delta t},\delta y_{T-\delta t}), \tag{2.55}$$

which can be determined from the system [Eq. (2.22)] and is given by Eq. (2.37). Furthermore, $P(x_{[\tau,T-\delta t]},\delta y_{[\tau,T-\delta t]})$ in Eq. (2.54) can be expressed as

$$P(x_{[\tau,T-\delta t]},\delta y_{[\tau,T-\delta t]}) = P(\delta y_{T-\delta t}|x_{[\tau,T-\delta t]},\delta y_{[\tau,T-2\delta t]}) \times P(x_{[\tau,T-\delta t]},\delta y_{[\tau,T-2\delta t]}). \tag{2.56}$$

Using the Markovian property of the observation process,

$$P(\delta y_{T-\delta t}|x_{[\tau,T-\delta t]},\delta y_{[\tau,T-2\delta t]}) = P(\delta y_{T-\delta t}|x_{T-\delta t}), \tag{2.57}$$

which can be determined from the observation equation [Eq. (2.23)] and is given by Eq. (2.40). Applying Eqs. (2.54)–(2.57) repeatedly, one obtains

$$P(\delta y_{[\tau,T-\delta t]}) = \int dx_T \int dx_{T-\delta t}P(x_T|x_{T-\delta t},\delta y_{T-\delta t})P(\delta y_{T-\delta t}|x_{T-\delta t}) \times \int dx_{T-2\delta t}P(x_{T-\delta t}|x_{T-2\delta t},\delta y_{T-2\delta t})P(\delta y_{T-2\delta t}|x_{T-2\delta t})\ldots\int dx_\tau P(x_{\tau+\delta t}|x_\tau,\delta y_\tau)P(\delta y_\tau|x_\tau)P(x_\tau). \tag{2.58}$$

Comparing this equation with Eq. (2.50), $P(\delta y_{\text{future}}|x_\tau)$ can be expressed as

$$P(\delta y_{\text{future}}|x_\tau) = P(\delta y_\tau|x_\tau)\int dx_{\tau+\delta t}P(x_{\tau+\delta t}|x_\tau,\delta y_\tau)\cdots P(\delta y_{T-2\delta t}|x_{T-2\delta t}) \times \int dx_{T-\delta t}P(x_{T-\delta t}|x_{T-2\delta t},\delta y_{T-2\delta t})P(\delta y_{T-\delta t}|x_{T-\delta t})\int dx_T P(x_T|x_{T-\delta t},\delta y_{T-\delta t}). \tag{2.59}$$

Defining the unnormalized retrodictive likelihood function at time t as

$$g(x,t) \propto P(dy_{[t,T)}|x_t = x), \tag{2.60}$$

one can derive a linear backward stochastic differential equation for $g(x,t)$ by applying Itō calculus backward in time to Eq. (2.59). The result is [6]

$$-dg = dt\sum_\mu A_\mu \frac{\partial g}{\partial x_\mu} + \frac{dt}{2}\sum_{\mu,\nu}(BQB^T)_{\mu\nu}\frac{\partial^2 g}{\partial x_\mu \partial x_\nu} + C^T R^{-1}dy_t g + \sum_\mu (BSR^{-1}dy_t)_\mu \frac{\partial g}{\partial x_\mu}, \tag{2.61}$$

which is the adjoint equation of the forward DMZ equation

[Eq. (2.48)], to be solved backward in time in the backward Itō sense, defined by

$$-dg \equiv g(x,t-dt) - g(x,t), \tag{2.62}$$

with the final condition

$$g(x,T) \propto 1. \tag{2.63}$$

The adjoint equation with respect to a linear differential equation

$$df(x,t) = \hat{L}f(x,t) \tag{2.64}$$

is defined as

$$-dg(x,t) = \hat{L}^{\dagger}g(x,t), \tag{2.65}$$

where $\hat{L}$ is a linear operator and $\hat{L}^{\dagger}$ is the adjoint of $\hat{L}$, defined by

$$\langle g(x),\hat{L}f(x)\rangle = \langle \hat{L}^{\dagger}g(x),f(x)\rangle \tag{2.66}$$

with respect to the inner product

$$\langle g(x),f(x)\rangle \equiv \int dxg(x)f(x). \tag{2.67}$$

After solving Eq. (2.48) for $f(x,\tau)$ and Eq. (2.61) for $g(x,\tau)$, the smoothing probability density is

$$h(x,\tau) \equiv P(x_\tau = x|dy_{[t_0,T)}) = \frac{g(x,\tau)f(x,\tau)}{\int dxg(x,\tau)f(x,\tau)}. \tag{2.68}$$

Since $f(x,\tau)$ and $g(x,\tau)$ are solutions of adjoint equations, their inner product, which appears as the denominator of Eq. (2.68), is constant in time [6]. The denominator also ensures that $h(x,\tau)$ is normalized and $f(x,\tau)$ and $g(x,\tau)$ need not be normalized separately.

The estimation errors depend crucially on the statistics of x_t. If any component of x_t, say $x_{\mu t}$, is constant in time, then filtering of that particular component is as accurate as smoothing for the simple reason that $P(x_{\mu\tau}|dy_{[t_0,T)})$ must be the same for any τ, and one can simply estimate $x_{\mu\tau}$ at the end of the observation interval ($\tau=T$) using filtering alone. This also means that smoothing is not needed when one only needs to detect the presence of a signal in detection problems [3] since the presence can be regarded as a constant binary parameter within a certain time interval. In general, however, smoothing can be significantly more accurate than filtering for the estimation of a fluctuating random process in the middle of the observation interval. Another reason for modeling unknown signals as random processes is robustness, as introducing fictitious system noise can improve the estimation accuracy when there are modeling errors [1,4].

E. Linear time-symmetric smoothing

If f, g, and h are Gaussian, one can just solve for their means and covariance matrices, which completely determine the probability densities. This is the case when the *a priori* probability density $P(x_{t_0})$ is Gaussian, and

$$A(x_t,t) = J(t)x_t, \tag{2.69}$$

$$B(x_t,t) = B(t), \tag{2.70}$$

$$C(x_t,t) = K(t)x_t. \tag{2.71}$$

The means and covariance matrices of f, g, and h can then be solved using the linear Mayne-Fraser-Potter (MFP) smoother [14]. The smoother first solves for the mean x' and covariance matrix Σ of f using the Kalman filter [1], given by

$$dx' = Jx'dt + \Gamma(dy - Kx'dt), \tag{2.72}$$

$$\Gamma \equiv (\Sigma K^T + BS)R^{-1}, \tag{2.73}$$

$$d\Sigma = (J\Sigma + \Sigma J^T - \Gamma R \Gamma^T + BQB^T)dt, \tag{2.74}$$

with the initial conditions at t_0 determined from $P(x_{t_0})$. The mean x'' and covariance matrix Ξ of g are then solved using a backward Kalman filter,

$$-dx'' = -Jx''dt + \Upsilon(dy - Kx''dt), \tag{2.75}$$

$$\Upsilon \equiv (\Xi K^T + BS)R^{-1}, \tag{2.76}$$

$$-d\Xi = (-J\Xi - \Xi J^T - \Upsilon R \Upsilon^T + BQB^T)dt, \tag{2.77}$$

with the final condition $\Xi_T^{-1}x_T''=0$ and $\Xi_T^{-1}=0$. In practice, the information filter formalism should be used to solve the backward filter in order to avoid dealing with the infinite covariance matrix at T [2,14]. Finally, the smoothing mean $\tilde{x}_\tau$ and covariance matrix Π_τ are

$$\tilde{x}_\tau = \Pi_\tau(\Sigma_\tau^{-1}x_\tau' + \Xi_\tau^{-1}x_\tau''), \tag{2.78}$$

$$\Pi_\tau = (\Sigma_\tau^{-1} + \Xi_\tau^{-1})^{-1}. \tag{2.79}$$

Note that x'' and Ξ are the mean and covariance matrix of a likelihood function $P(dy_{[t,T)}|x_t)$ and not those of a conditional probability density $P(x_t|dy_{[t,T)})$, so to perform optimal retrodiction ($\tau=t_0$) one should still combine x'' and Ξ with the *a priori* values [15].

III. HYBRID CLASSICAL-QUANTUM SMOOTHING

A. Problem statement

Consider the problem of waveform estimation in a hybrid classical-quantum system depicted in Fig. 3. The classical system produces a vectoral classical diffusive Markov random process x_t, which obeys Eq. (2.2) and is coupled to the quantum system. The goal is to estimate x_τ via continuous measurements of both systems. This setup is slightly more general than that considered in [5]; here the observations can also depend on x_t. This allows one to apply the theory to PLL design for squeezed beams, as considered by Berry and Wiseman [7] and potentially to other quantum estimation problems as well [16]. The statistics of x_t are assumed to be unperturbed by the coupling to the quantum system in order to avoid the nontrivial issue of quantum backaction on classical systems [17]. For simplicity, in this section we neglect

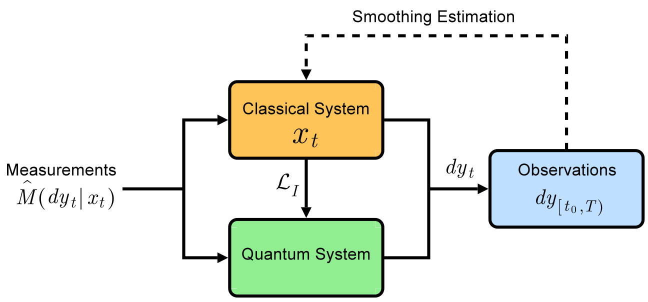

FIG. 3. (Color online) Schematic of the hybrid classical-quantum smoothing problem.

the possibility that the system noise driving the classical system is correlated with the observation noise, although the noise driving the quantum system can still be correlated with the observation noise due to quantum measurement backaction. Just as in the classical smoothing problem, the hybrid smoothing problem is solved by calculating the smoothing probability density $P(x_\tau|dy_{[t_0,T)})$.

B. Time-symmetric approach

Because a quantum system is involved, one may be tempted to use a hybrid density operator [5,7,16,17] to represent one's knowledge about the hybrid classical-quantum system. The hybrid density operator $\hat{\rho}(x_\tau)$ describes the joint classical and quantum statistics of a hybrid system, with the marginal classical probability density for x_τ and the marginal density operator for the quantum system given by

$$P(x_\tau) = \mathrm{tr}[\hat{\rho}(x_\tau)], \tag{3.1}$$

$$\hat{\rho}(\tau) = \int dx_\tau \hat{\rho}(x_\tau), \tag{3.2}$$

respectively. The hybrid operator can also be regarded as a special case of the quantum density operator when certain degrees of freedom are approximated as classical. Unfortunately, the density operator in conventional predictive quantum theory can only be conditioned upon past observations and not future ones, so it cannot be used as a quantum version of the smoothing probability density.

The classical time-symmetric smoothing theory, as a combination of prediction and retrodiction, offers an important clue to how one can circumvent the difficulty of defining the smoothing quantum state. Again casting the problem in discrete time and defining a hybrid effect operator as $\hat{E}(\delta y_{\text{future}}|x_\tau)$, which can be used to determine the statistics of future observations given a density operator at τ,

$$P(\delta y_{\text{future}}) = \int dx_\tau\, \mathrm{tr}[\hat{E}(\delta y_{\text{future}}|x_\tau)\hat{\rho}(x_\tau)], \tag{3.3}$$

one may write, in analogy with Eq. (2.34) [5],

$$P(x_\tau|\delta y_{[t_0,T-\delta t]}) = \frac{P(x_\tau, \delta y_{\text{future}}|\delta y_{\text{past}})}{P(\delta y_{\text{future}}|\delta y_{\text{past}})} = \frac{\mathrm{tr}[\hat{E}(\delta y_{\text{future}}|x_\tau)\hat{\rho}(x_\tau|\delta y_{\text{past}})]}{\int dx_\tau\, \mathrm{tr}[\hat{E}(\delta y_{\text{future}}|x_\tau)\hat{\rho}(x_\tau|\delta y_{\text{past}})]}, \tag{3.4}$$

where $\hat{\rho}(x_\tau|\delta y_{\text{past}})$ is the analog of the filtering probability density $P(x_\tau|\delta y_{\text{past}})$ and $\hat{E}(\delta y_{\text{future}}|x_\tau)$ is the analog of the retrodictive likelihood function $P(\delta y_{\text{future}}|x_\tau)$. One can then solve for the density and effect operators separately before combining them to form the classical smoothing probability density.

C. Filtering

Since the hybrid density operator can be regarded as a special case of the density operator, the same tools in quantum measurement theory can be used to derive a filtering equation for the hybrid operator. First, write $\hat{\rho}(x_{t+\delta t}|\delta y_{[t_0,t]})$ in terms of $\hat{\rho}(x_t|\delta y_{[t_0,t]})$ as

$$\hat{\rho}(x_{t+\delta t}|\delta y_{[t_0,t]}) = \int dx_t \mathcal{K}(x_{t+\delta t}|x_t)\hat{\rho}(x_t|\delta y_{[t_0,t]}), \tag{3.5}$$

where $\mathcal{K}$ is a completely positive map that governs the Markovian evolution of the hybrid state independent of the measurement process. Equation (3.5) may be regarded as a quantum version of the classical Chapman-Kolmogorov equation. For infinitesimal δt,

$$\int dx_t \mathcal{K}(x_{t+\delta t}|x_t)\hat{\rho}(x_t) \approx [(\hat{1} + \delta t\mathcal{L})\hat{\rho}(x_t = x)]_{x=x_{t+\delta t}}. \tag{3.6}$$

The hybrid superoperator $\mathcal{L}$ can be expressed as

$$\mathcal{L}\hat{\rho}(x) = \mathcal{L}_0\hat{\rho}(x) + \mathcal{L}_I(x)\hat{\rho}(x) - \sum_\mu \frac{\partial}{\partial x_\mu}[A_\mu\hat{\rho}(x)] + \frac{1}{2}\sum_{\mu,\nu} \frac{\partial^2}{\partial x_\mu\, \partial x_\nu}[(BQB^T)_{\mu\nu}\hat{\rho}(x)], \tag{3.7}$$

where $\mathcal{L}_0$ governs the evolution of the quantum system, $\mathcal{L}_I$ governs the coupling of x_t to the quantum system, via an interaction Hamiltonian, for example, and the last two terms govern the classical evolution of x_t.

Next, write $\hat{\rho}(x_t|\delta y_{[t_0,t]})$ in terms of $\hat{\rho}(x_t|\delta y_{[t_0,t-\delta t]})$ using the quantum Bayes theorem [18] as

$$\hat{\rho}(x_t|\delta y_{[t_0,t]}) = \hat{\rho}(x_t|\delta y_{[t_0,t-\delta t]}, \delta y_t) = \frac{\mathcal{J}(\delta y_t|x_t)\hat{\rho}(x_t|\delta y_{[t_0,t-\delta t]})}{\int dx_t\, \mathrm{tr}(\text{numerator})}. \tag{3.8}$$

The measurement superoperator $\mathcal{J}(\delta y_t|x_t)$, a quantum version of $P(\delta y_t|x_t)$, is defined as

TABLE I. Important quantities in classical smoothing and their generalizations in hybrid smoothing.

Classical	Description	Hybrid	Description
$P(x_{t+\delta t}\|x_t,\delta y_t)$	Transition probability density, appears in Chapman-Kolmogorov equation [Eq. (2.36)]	$\mathcal{K}(x_{t+\delta t}\|x_t)$	Transition superoperator, appears in quantum Chapman-Kolmogorov equation [Eq. (3.5)]
$P(\delta y_t\|x_t)$	Observation probability density, appears in Bayes theorem [Eq. (2.39)]	$\mathcal{J}(\delta y_t\|x_t)$	Measurement superoperator, appears in quantum Bayes theorem [Eq. (3.8)]
$P(x_t\|dy_{[t_0,t)})$, $F(x,t)$	Filtering probability density, obeys Kushner-Stratonovich equation [Eq. (2.43)]	$\hat{\rho}(x_t\|dy_{[t_0,t)})$, $\hat{F}(x,t)$	Filtering hybrid density operator, obeys Belavkin equation [Eq. (3.17)]
$f(x,t)$	Unnormalized $F(x,t)$, obeys Duncan-Mortensen-Zakai (DMZ) equation [Eq. (2.48))]	$\hat{f}(x,t)$	Unnormalized $f(x,t)$, obeys quantum DMZ equation [Eq. (3.20)]
$P(dy_{[t,T)}\|x_t)$	Retrodictive likelihood function	$\hat{E}(dy_{[t,T)}\|x_t)$	Hybrid effect operator
$g(x,t)$	Unnormalized $P(dy_{[t,T)}\|x_t)$, obeys backward DMZ equation [Eq. (2.61)]	$\hat{g}(x,t)$	Unnormalized $\hat{E}(dy_{[t,T)}\|x_t)$, obeys backward quantum DMZ equation [Eq. (3.30)]
$P(x_\tau\|dy_{[t_0,T)})$, $h(x,\tau)$	Smoothing probability density, obeys Eq. (2.68)	$P(x_\tau\|dy_{[t_0,T)})$, $h(x,\tau)$	Smoothing probability density, obeys Eq. (3.32)

$$\mathcal{J}(\delta y_t|x_t)\hat{\rho}(x_t) \equiv \hat{M}(\delta y_t|x_t)\hat{\rho}(x_t)\hat{M}^\dagger(\delta y_t|x_t). \tag{3.9}$$

For infinitesimal δt and measurements with Gaussian noise, the measurement operator $\hat{M}$ can be approximated as [19]

$$\hat{M}(\delta z_t|x_t) \propto \hat{1} + \sum_\mu \gamma_\mu(t)\left[\frac{1}{2}\hat{c}_\mu(x_t,t)\,\delta z_{\mu t} - \frac{\delta t}{8}\hat{c}^\dagger_\mu(x_t,t)\hat{c}_\mu(x_t,t)\right], \tag{3.10}$$

where δz_t is a vectoral observation process, $\hat{c}(x_t,t)$ is a vector of hybrid operators, generalized from the purely quantum $\hat{c}$ operators in Ref. [5] so that the observations may also depend directly on the classical degrees of freedom, and $\gamma_\mu(t)$ is assumed to be positive. To cast the theory in a form similar to the classical one, perform unitary transformations on δz_t and $\hat{c}$,

$$\delta y_t = U\delta z_t, \tag{3.11}$$

$$\hat{C}(x_t,t) = U\hat{c}(x_t,t), \tag{3.12}$$

where U is a unitary matrix, and rewrite the measurement operator as

$$\hat{M}(\delta y_t|x_t) \propto \hat{1} + \frac{1}{2}\hat{C}^T(x_t,t)R^{-1}(t)\,\delta y_t - \frac{\delta t}{8}\hat{C}^{\dagger T}(x_t,t)R^{-1}(t)\hat{C}(x_t,t). \tag{3.13}$$

$\hat{C}(x_t,t)$ is a generalization of $C(x_t,t)$ in the classical case and $R(t)$ is again a positive-definite matrix that characterizes the observation uncertainties and is real and symmetric with eigenvalues $1/\gamma_\mu$. Note that superscript $\dagger$ is defined as the adjoint of each vector element and superscript T is defined as the matrix transpose of the vector. For example,

$$\hat{C}^{\dagger T}R^{-1}\hat{C} \equiv \sum_{\mu,\nu}\hat{C}^\dagger_\mu(R^{-1})_{\mu\nu}\hat{C}_\nu. \tag{3.14}$$

The evolution of $\hat{\rho}(x_t|\,\delta y_{[t_0,t-\delta t]})$ can thus be calculated by iterating the formula

$$\hat{\rho}(x_{t+\delta t}|\,\delta y_{[t_0,t]}) = \int dx_t \mathcal{K}(x_{t+\delta t}|x_t)\frac{\mathcal{J}(\delta y_t|x_t)\hat{\rho}(x_t|\,\delta y_{[t_0,t-\delta t]})}{\int dx_t\ \mathrm{tr}(\mathrm{numerator})}. \tag{3.15}$$

Taking the continuous time limit via Itō calculus and defining the conditional hybrid density operator at time t as

$$\hat{F}(x,t) \equiv \hat{\rho}(x_t = x|dy_{[t_0,t)}), \tag{3.16}$$

one obtains [5]

$$d\hat{F} = dt\mathcal{L}\hat{F} + \frac{dt}{8}(2\hat{C}^TR^{-1}\hat{F}\hat{C}^\dagger - \hat{C}^{\dagger T}R^{-1}\hat{C}\hat{F} - \hat{F}\hat{C}^{\dagger T}R^{-1}\hat{C}) + \frac{1}{2}[(\hat{C}-\langle\hat{C}\rangle_{\hat{F}})^TR^{-1}d\eta_t\hat{F} + \mathrm{H.c.}], \tag{3.17}$$

where

$$\langle\hat{C}\rangle_{\hat{F}} \equiv \int dx\ \mathrm{tr}[\hat{C}(x,t)\hat{F}(x,t)], \tag{3.18}$$

$$d\eta_t \equiv dy_t - \frac{dt}{2}\langle\hat{C}+\hat{C}^\dagger\rangle_{\hat{F}} \tag{3.19}$$

is a Wiener increment with covariance matrix $R(t)dt$ [19], H.c. denotes the Hermitian conjugate, and the initial condition is the *a priori* hybrid density operator $\hat{\rho}(x_{t_0})$. Equation (3.17) is a quantum version of the KS equation [Eq. (2.43)] and can be regarded as a special case of the Belavkin quantum filtering equation [20].

A linear version of the KS equation for an unnormalized $\hat{F}(x,t)$ is

$$d\hat{f} = dt\mathcal{L}\hat{f} + \frac{dt}{8}(2\hat{C}^T R^{-1}\hat{f}\hat{C}^\dagger - \hat{C}^{\dagger T}R^{-1}\hat{C}\hat{f} - \hat{f}\hat{C}^{\dagger T}R^{-1}\hat{C}) + \frac{1}{2}(\hat{C}^T R^{-1} dy_t \hat{f} + \text{H.c.}) \tag{3.20}$$

and the normalization is

$$\hat{F}(x,t) = \frac{\hat{f}(x,t)}{\int dx\, \mathrm{tr}[\hat{f}(x,t)]}. \tag{3.21}$$

Equation (3.20) is a quantum generalization of the DMZ equation [Eq. (2.48)].

D. Retrodiction and smoothing

Taking a similar approach to the one in Sec. II D and using the quantum regression theorem, one can express the future observation statistics as [21]

$$P(\delta y_{\text{future}}) = \int dx_\tau\, \mathrm{tr}[\hat{E}(\delta y_{\text{future}}|x_\tau)\hat{\rho}(x_\tau)] \tag{3.22}$$

$$= \int dx_T\, \mathrm{tr}\Bigg[\int dx_{T-\delta t}\mathcal{K}(x_T|x_{T-\delta t}) \cdot \mathcal{J}(\delta y_{T-\delta t}|x_{T-\delta t}) \times \int dx_{T-2\delta t}\mathcal{K}(x_{T-\delta t}|x_{T-2\delta t}) \cdot \mathcal{J}(\delta y_{T-2\delta t}|x_{T-2\delta t}) \ldots \cdot \int dx_\tau \mathcal{K}(x_{\tau+\delta t}|x_\tau)\mathcal{J}(\delta y_\tau|x_\tau)\hat{\rho}(x_\tau)\Bigg], \tag{3.23}$$

which are analogous to Eqs. (2.50) and (2.58), respectively. Comparing Eq. (3.22) with Eq. (3.23) and defining the adjoint of a superoperator $\mathcal{O}$ as $\mathcal{O}^*$, such that

$$\mathrm{tr}[\hat{E}(x)\mathcal{O}\hat{\rho}(x)] = \mathrm{tr}\{[\mathcal{O}^*\hat{E}(x)]\hat{\rho}(x)\}, \tag{3.24}$$

the hybrid effect operator can be written as

$$\hat{E}(\delta y_{\text{future}}|x_\tau) = \mathcal{J}^*(\delta y_\tau|x_\tau)\int dx_{\tau+\delta t}\mathcal{K}^*(x_{\tau+\delta t}|x_\tau)\ldots \cdot \mathcal{J}^*(\delta y_{T-2\delta t}|x_{T-2\delta t})\int dx_{T-\delta t}\mathcal{K}^*(x_{T-\delta t}|x_{T-2\delta t}) \cdot \mathcal{J}^*(\delta y_{T-\delta t}|x_{T-\delta t})\int dx_T\mathcal{K}^*(x_T|x_{T-\delta t})\hat{1}. \tag{3.25}$$

The operation $\mathcal{K}^* \equiv \int dx' \mathcal{K}^*(x'|x)\cdot$ may also be regarded as a hybrid superoperator on a hybrid operator and is the adjoint of $\mathcal{K} \equiv \int dx' \mathcal{K}(x|x')\cdot$, defined by

$$\langle \hat{E}(x), \mathcal{K}\hat{\rho}(x)\rangle = \langle \mathcal{K}^*\hat{E}(x), \hat{\rho}(x)\rangle, \tag{3.26}$$

with respect to the Hilbert-Schmidt inner product

$$\langle \hat{E}(x), \hat{\rho}(x)\rangle \equiv \int dx\, \mathrm{tr}[\hat{E}(x)\hat{\rho}(x)]. \tag{3.27}$$

One can then rewrite Eqs. (3.22), (3.23), and (3.25) more elegantly as

$$\langle \hat{E}(x), \hat{\rho}(x)\rangle = \langle \hat{1}, \mathcal{K}\mathcal{J}\ldots\mathcal{K}\mathcal{J}\hat{\rho}(x)\rangle, \tag{3.28}$$

$$\hat{E}(x) = \mathcal{J}^*\mathcal{K}^*\ldots\mathcal{J}^*\mathcal{K}^*\hat{1}. \tag{3.29}$$

In the continuous time limit, a linear stochastic differential equation for the unnormalized effect operator $\hat{g}(x,t) \propto \hat{E}(dy_{[t,T)}|x_t = x)$ can be derived. The result is [5]

$$-d\hat{g} = dt\mathcal{L}^*\hat{g} + \frac{dt}{8}(2\hat{C}^{\dagger T}R^{-1}\hat{g}\hat{C} - \hat{g}\hat{C}^{\dagger T}R^{-1}\hat{C} - \hat{C}^{\dagger T}R^{-1}\hat{C}\hat{g}) + \frac{1}{2}(\hat{g}\hat{C}^T R^{-1}dy_t + \text{H.c.}), \tag{3.30}$$

to be solved backward in time in the backward Itō sense, with the final condition

$$\hat{g}(x,t) \propto \hat{1}. \tag{3.31}$$

Equation (3.30) is the adjoint equation of the forward quantum DMZ equation [Eq. (3.20)] with respect to the inner product defined by Eq. (3.27). It is a generalization of the classical backward DMZ equation [Eq. (2.61)].

Finally, after solving Eq. (3.20) for $\hat{f}(x,\tau)$ and Eq. (3.30) for $\hat{g}(x,\tau)$, the smoothing probability density is

$$h(x,\tau) \equiv P(x_\tau = x|dy_{[t_0,T)}) = \frac{\mathrm{tr}[\hat{g}(x,\tau)\hat{f}(x,\tau)]}{\int dx\, \mathrm{tr}[\hat{g}(x,\tau)\hat{f}(x,\tau)]}. \tag{3.32}$$

The denominator of Eq. (3.32) ensures that $h(x,\tau)$ is normalized, so $\hat{f}(x,\tau)$ and $\hat{g}(x,\tau)$ need not be normalized separately. Table I lists some important quantities in classical smoothing

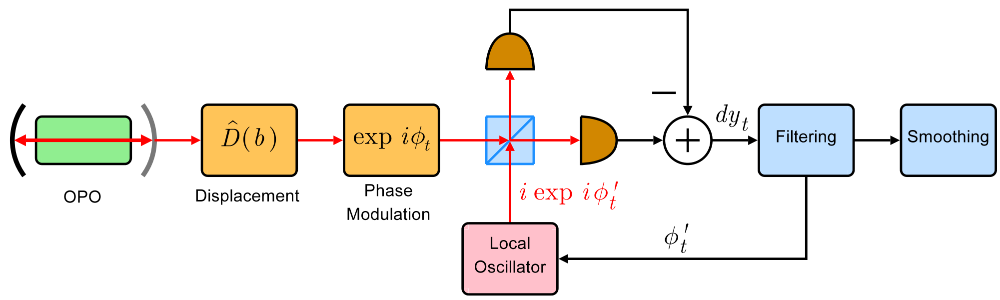

FIG. 4. (Color online) Homodyne phase-locked loop (PLL) for phase estimation with a narrowband squeezed optical beam produced from an optical parametric oscillator (OPO).

with their generalizations in hybrid smoothing for comparison.

E. Smoothing in terms of Wigner distributions

To solve Eqs. (3.20), (3.30), and (3.32), one way is to convert them to equations for quasiprobability distributions [22]. The Wigner distribution is especially useful for quantum systems with continuous degrees of freedom. It is defined as [22,23]

$$f(q,p) \equiv \frac{1}{2\pi}\int du \left\langle q-\frac{u}{2}\middle|\hat{f}\middle|q+\frac{u}{2}\right\rangle \exp(ip^T u), \tag{3.33}$$

where q and p are normalized position and momentum vectors. It has the desirable property

$$\int dqdp g(q,p)f(q,p) = \frac{1}{2\pi}\mathrm{tr}(\hat{g}\hat{f}), \tag{3.34}$$

which is unique among generalized quasiprobability distributions [23]. The smoothing probability density given by Eq. (3.32) can then be rewritten as

$$h(x,\tau) = \frac{\int dqdp g(q,p,x,\tau)f(q,p,x,\tau)}{\int dqdpdx g(q,p,x,\tau)f(q,p,x,\tau)}, \tag{3.35}$$

where $f(q,p,x,\tau)$ and $g(q,p,x,\tau)$ are the Wigner distributions of $\hat{f}$ and $\hat{g}$, respectively. Equation (3.35) resembles classical expression (2.68) with the quantum degrees of freedom q and p marginalized. If $f(q,p,x,t_0)$ is nonnegative and the stochastic equations for $f(q,p,x,t)$ and $g(q,p,x,t)$ converted from Eqs. (3.20) and (3.30) have the same form as the classical DMZ equations given by Eqs. (2.48) and (2.61), the hybrid smoothing problem becomes equivalent to a classical one and can be solved using well known classical smoothers. For example, if $f(q,p,x,t)$ and $g(q,p,x,t)$ are Gaussian, $h(x,\tau)$ is also Gaussian, and their means and covariances can be solved using the linear MFP smoother described in Sec. II E.

IV. PHASE-LOCKED LOOP DESIGN FOR NARROWBAND SQUEEZED BEAMS

Consider the PLL setup depicted in Fig. 4. The OPO produces a squeezed vacuum with a squeezed p quadrature and an antisqueezed q quadrature. The squeezed vacuum is then displaced by a real constant b to produce a phase-squeezed beam, the phase of which is modulated by $\phi_t = x_{1t}$, an element of the vectoral random process x_t described by the system Itō equation [Eq. (2.2)]. The output beam is measured continuously by a homodyne PLL, and the local-oscillator phase ϕ_t' is continuously updated according to the real-time measurement record.

The use of PLL for phase estimation in the presence of quantum noise has been mentioned as far back as 1971 by Personick [24]. Wiseman suggested an adaptive homodyne scheme to measure a constant phase [25], which was then experimentally demonstrated by Armen *et al.* for the optical coherent state [26]. Berry and Wiseman [27] and Pope *et al.* [28] studied the problem with ϕ_t being a Wiener process. Berry and Wiseman later generalized the theory to account for narrowband squeezed beams [7]. Tsang *et al.* also studied the problem for the case of x_t being a Gaussian process [29,30], but the squeezing model considered in Refs. [29,30] is not realistic. Using the hybrid smoothing theory developed in Sec. III, one can now generalize these earlier results to the case of an arbitrary diffusive Markov process and a realistic squeezing model.

Let $\hat{\rho}(x_t)$ be the hybrid density operator for the combined quantum-OPO-classical-modulator system. The evolution of the OPO below threshold in the interaction picture is governed by

$$\mathcal{L}_0\hat{\rho}(x) = -\frac{i}{\hbar}[\hat{H}_0, \hat{\rho}(x)], \tag{4.1}$$

$$\hat{H}_0 = -i\frac{\hbar\chi}{2}(\hat{a}\hat{a} - \hat{a}^\dagger\hat{a}^\dagger) \tag{4.2}$$

$$= \frac{\hbar\chi}{2}(\hat{q}\hat{p} + \hat{p}\hat{q}), \tag{4.3}$$

where $\hat{a}$ is the annihilation operator for the cavity optical mode and $\hat{q}$ and $\hat{p}$ are the antisqueezed and squeezed quadrature operators, respectively, defined as

$$\hat{q} \equiv \frac{\hat{a} + \hat{a}^\dagger}{\sqrt{2}}, \tag{4.4}$$

$$\hat{p} \equiv \frac{\hat{a}-\hat{a}^\dagger}{\sqrt{2}i}, \tag{4.5}$$

with the commutation relation

$$[\hat{q},\hat{p}] = i. \tag{4.6}$$

The classical phase modulator does not influence the evolution of the OPO, so

$$\mathcal{L}_I = 0, \tag{4.7}$$

but it modulates the OPO output. $\hat{C}(x_t,t)$ in this case is

$$\hat{C}(x_t,t) = -2i(b+\sqrt{\gamma}\hat{a})\exp(i\phi_t - i\phi_t'), \tag{4.8}$$

where γ is the transmission coefficient of the partially reflecting OPO output mirror, $R=1$, and the symbol and sign conventions here roughly follow those in Refs. [29,30]. To ensure the correct unconditional quantum dynamics, the Hamiltonian should be changed to (Sec. 11.4.3 in Ref. [18])

$$\hat{H}_0' = \hat{H}_0 - i\frac{i\hbar b\sqrt{\gamma}}{2}(\hat{a}-\hat{a}^\dagger), \quad \mathcal{L}_0\hat{\rho}(x) = -\frac{i}{\hbar}[\hat{H}_0',\hat{\rho}(x)] \tag{4.9}$$

in order to eliminate the spurious effect of the displacement term in $\hat{C}$ on the OPO. After some algebra, the forward stochastic equation for the Wigner distribution $f(q,p,x,t)$ becomes

$$\begin{aligned} df = dt\Bigg\{ &-\sum_\mu \frac{\partial}{\partial x_\mu}(A_\mu f) + \frac{1}{2}\sum_{\mu,\nu}\frac{\partial^2}{\partial x_\mu \partial x_\nu}[(BQB^T)_{\mu\nu}f] \\ &- \left[\left(\chi - \frac{\gamma}{2}\right)\frac{\partial}{\partial q}(qf) + \left(-\chi - \frac{\gamma}{2}\right)\frac{\partial}{\partial p}(pf)\right] + \frac{\gamma}{4}\left(\frac{\partial^2 f}{\partial q^2} \right.\\ &\left.+ \frac{\partial^2 f}{\partial p^2}\right)\Bigg\} + dy_t\Bigg[\sin(\phi - \phi_t')\left(2b + \sqrt{2\gamma}q + \sqrt{\frac{\gamma}{2}}\frac{\partial}{\partial q}\right) \\ &+ \cos(\phi - \phi_t')\left(\sqrt{2\gamma}p + \sqrt{\frac{\gamma}{2}}\frac{\partial}{\partial p}\right)\Bigg]f. \end{aligned} \tag{4.10}$$

This is precisely the classical DMZ equation [Eq. (2.48)] with correlated system and observation noises. The equivalent classical system equations are then

$$dq_t = \left(\chi - \frac{\gamma}{2}\right)q_t dt + \sqrt{\frac{\gamma}{2}}d\alpha_t,$$

$$dp_t = \left(-\chi - \frac{\gamma}{2}\right)p_t dt + \sqrt{\frac{\gamma}{2}}d\beta_t,$$

$$dx_t = A(x_t,t)dt + B(x_t,t)dW_t, \tag{4.11}$$

and the equivalent observation equation is

$$dy_t = 2b\sin(\phi_t - \phi_t')dt + d\zeta_t,$$

$$\begin{aligned} d\zeta_t \equiv &\sin(\phi_t - \phi_t')(\sqrt{2\gamma}q_t dt - d\alpha_t) \\ &+ \cos(\phi_t - \phi_t')(\sqrt{2\gamma}p_t dt - d\beta_t), \end{aligned} \tag{4.12}$$

where $d\alpha_t$ and $d\beta_t$ are independent Wiener increments with covariance dt. $d\alpha_t$ and $d\beta_t$, which appear in both the system equation and the observation equation, are simply quadratures of the vacuum field, coupled to both the cavity mode and the output field via the OPO output mirror. Equations (4.11) and (4.12) coincide with the model of Berry and Wiseman in Ref. [7] when x_t is a Wiener process, and Eq. (4.10) is the continuous limit of their approach to phase estimation. This approach can also be regarded as an example of the general method of accounting for colored observation noise by modeling the noise as part of the system [2–4].

If $\chi=0$, $d\zeta_t/dt$ is an additive white Gaussian noise, and the model is reduced to that studied in Refs. [27–30]. In that case, it is desirable to make ϕ_t' follow ϕ_t as closely as possible, so that dy_t can be approximated as

$$dy_t \approx 2b(\phi_t - \phi_t')dt + d\zeta_t, \tag{4.13}$$

and the Kalman filter can be used if x_t is Gaussian [30]. Provided that Eq. (4.13) is valid, one should make ϕ_t' the conditional expectation of $\phi_t = x_{1t}$, given by

$$\phi_t' = \langle\phi_t\rangle_{\hat{f}} = \int dq dp dx x_1 f(q,p,x,t). \tag{4.14}$$

For phase-squeezed beams, it also seems desirable to make ϕ_t' close to ϕ_t in order to minimize the magnitude of $d\zeta_t$. Equation (4.14) may not provide the optimal ϕ_t' in general, however, as it does not necessarily minimize the magnitude of $d\zeta_t$ or the estimation errors. The optimal control law for ϕ_t' should be studied in the context of control theory.

While ϕ_t' needs to be updated in real time and must be calculated via filtering, the estimation accuracy can be improved by smoothing. The backward DMZ equation for $g(q,p,x,t)$ is the adjoint equation with respect to Eq. (4.10), given by

$$\begin{aligned} -dg = dt\Bigg\{ &\sum_\mu A_\mu \frac{\partial g}{\partial x_\mu} + \frac{1}{2}\sum_{\mu,\nu}(BQB^T)_{\mu\nu}\frac{\partial^2 g}{\partial x_\mu \partial x_\nu} \\ &+ \left[\left(\chi - \frac{\gamma}{2}\right)q\frac{\partial g}{\partial q} + \left(-\chi - \frac{\gamma}{2}\right)p\frac{\partial g}{\partial p}\right] + \frac{\gamma}{4}\left(\frac{\partial^2 g}{\partial q^2}\right.\\ &\left.+ \frac{\partial^2 g}{\partial p^2}\right)\Bigg\} + dy_t\Bigg[\sin(\phi - \phi_t')\left(2b + \sqrt{2\gamma}q - \sqrt{\frac{\gamma}{2}}\frac{\partial}{\partial q}\right) \\ &+ \cos(\phi - \phi_t')\left(\sqrt{2\gamma}p - \sqrt{\frac{\gamma}{2}}\frac{\partial}{\partial p}\right)\Bigg]g, \end{aligned} \tag{4.15}$$

and the smoothing probability density $h(x,\tau)$ is given by Eq. (3.35). The use of linear smoothing for the case of x_t being a Gaussian process and $d\zeta_t/dt$ being a white Gaussian noise has been studied in Refs. [29,30]. Practical strategies of solving Eqs. (4.10) and (4.15) in general are beyond the scope of this paper, but classical nonlinear filtering and smoothing techniques should help [1–4].

One can also use the hybrid smoothing theory to study the general problem of force estimation via a squeezed probe

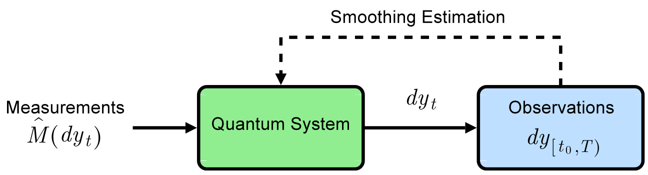

FIG. 5. (Color online) The quantum smoothing problem.

beam and a homodyne PLL by modeling the phase modulator as a quantum mechanical oscillator instead and combining the problem studied in this section with the force estimation problem studied in Ref. [5].

V. WEAK VALUES AS QUANTUM SMOOTHING ESTIMATES

Previous sections focus on the estimation of classical signals, but there is no reason why one cannot apply smoothing to quantum degrees of freedom as well, as shown in Fig. 5. First consider the predicted density operator at time τ conditioned upon past observations, given by

$$\hat{\rho}(\tau) \equiv \frac{\hat{f}(\tau)}{\mathrm{tr}[\hat{f}(\tau)]}, \tag{5.1}$$

where the classical degrees of freedom are neglected for simplicity. The predicted expectation of an observable, such as the position of a quantum mechanical oscillator, is

$$\langle \hat{O} \rangle_{\hat{f}} \equiv \mathrm{tr}[\hat{O}\hat{\rho}(\tau)] = \frac{\mathrm{tr}[\hat{O}\hat{f}(\tau)]}{\mathrm{tr}[\hat{f}(\tau)]}. \tag{5.2}$$

One may also use retrodiction, after some measurements of a quantum system have been made, to estimate its initial quantum state before the measurements [31,32], using the retrodictive density operator defined as

$$\hat{\rho}_{\mathrm{ret}}(\tau) \equiv \frac{\hat{g}(\tau)}{\mathrm{tr}[\hat{g}(\tau)]}. \tag{5.3}$$

The retrodicted expectation of an observable is

$${}_{\hat{g}}\langle \hat{O} \rangle \equiv \mathrm{tr}[\hat{\rho}_{\mathrm{ret}}(\tau)\hat{O}] = \frac{\mathrm{tr}[\hat{g}(\tau)\hat{O}]}{\mathrm{tr}[\hat{g}(\tau)]}. \tag{5.4}$$

Causality prevents one from going back in time to verify the retrodicted expectation, but if the degree of freedom with respect to $\hat{O}$ at time τ is entangled with another "probe" system, then one can verify the retrodicted expectation by measuring the probe and inferring $\hat{O}$ [32].

The idea of verifying retrodiction by entangling the system at time τ with a probe can also be extended to the case of smoothing, as proposed by Aharonov *et al.* [8]. In the middle of a sequence of measurements, if one weakly couples the system to a probe for a short time, so that the system is weakly entangled with the probe and the probe is subsequently measured, the measurement outcome on average can be characterized by the so-called weak value of an observable, defined as [8,33]

$${}_{\hat{g}}\langle \hat{O} \rangle_{\hat{f}} \equiv \frac{\mathrm{tr}[\hat{g}(\tau)\hat{O}\hat{f}(\tau)]}{\mathrm{tr}[\hat{g}(\tau)\hat{f}(\tau)]}. \tag{5.5}$$

The weak value becomes a prediction given by Eq. (5.2) when future observations are neglected, such that $\hat{g}(\tau)=\hat{1}$, and becomes a retrodiction given by Eq. (5.4) when past observations are neglected and there is no *a priori* information about the quantum system at time τ, such that $\hat{f}(\tau)=\hat{1}$. When $\hat{f}(\tau)$ and $\hat{g}(\tau)$ are incoherent mixtures of $\hat{O}$ eigenstates,

$$\hat{f}(\tau) = \sum_O f(O,\tau)|O\rangle\langle O|, \tag{5.6}$$

$$\hat{g}(\tau) = \sum_O g(O,\tau)|O\rangle\langle O|, \tag{5.7}$$

the weak value becomes

$${}_{\hat{g}}\langle \hat{O} \rangle_{\hat{f}} = \frac{\sum_O O g(O,\tau) f(O,\tau)}{\sum_O g(O,\tau) f(O,\tau)} \tag{5.8}$$

and is consistent with the classical time-symmetric smoothing theory described in Sec. II. Hence, the weak value can be regarded as a quantum generalization of the smoothing estimate, conditioned upon past and future observations.

One can also establish a correspondence between a classical theory and a quantum theory via quasiprobability distributions. Given the smoothing probability density in terms of the Wigner distributions in Eq. (3.35), one may be tempted to undo the marginalizations over the quantum degrees of freedom and define a smoothing quasiprobability distribution as

$$h(q,p,\tau) = \frac{g(q,p,\tau)f(q,p,\tau)}{\int dq dp g(q,p,\tau)f(q,p,\tau)}, \tag{5.9}$$

where $f(q,p,\tau)$ and $g(q,p,\tau)$ are the Wigner distributions of $\hat{f}(\tau)$ and $\hat{g}(\tau)$, respectively. Intriguingly, $h(q,p,\tau)$, being the product of two Wigner distributions, can exhibit quantum position and momentum uncertainties that violate the Heisenberg uncertainty principle. This has been shown in Ref. [30] when the position of a quantum mechanical oscillator is monitored via continuous measurements and smoothing is applied to the observations. From the perspective of classical estimation theory, it is perhaps not surprising that smoothing can improve upon an uncertainty relation based on a predictive theory. The important question is whether the sub-Heisenberg uncertainties can be verified experimentally. Tsang *et al.* [30] argued that it can be done only by Bayesian estimation, but in the following I shall propose another method based on weak measurements.

It can be shown that the expectation of q using $h(q,p,\tau)$ is

$$\langle q \rangle_h \equiv \int dq dp q h(q,p,\tau) \tag{5.10}$$

$$=\mathrm{Re}\frac{\mathrm{tr}[\hat{g}(\tau)\hat{q}\hat{f}(\tau)]}{\mathrm{tr}[\hat{g}(\tau)\hat{f}(\tau)]}=\mathrm{Re}\ {}_{\hat{g}}\langle\hat{q}\rangle_{\hat{f}}, \tag{5.11}$$

which is the real part of the weak value, and likewise for $\langle p\rangle_h$, so the smoothing position and momentum estimates are closely related to their weak values. More generally, consider the joint probability density for a quantum position measurement followed by a quantum momentum measurement, conditioned upon past and future observations,

$$P(y_q,y_p)=\frac{1}{\mathcal{C}}\mathrm{tr}[\hat{g}(\tau)\hat{M}_p(y_p)\hat{M}_q(y_q)\hat{f}(\tau)\hat{M}_q^{\dagger}(y_q)\hat{M}_p^{\dagger}(y_p)], \tag{5.12}$$

$$\mathcal{C}\equiv\int dy_q dy_p\ \mathrm{tr}[\hat{g}(\tau)\hat{M}_p(y_p)\hat{M}_q(y_q)\hat{f}(\tau)\hat{M}_q^{\dagger}(y_q)\hat{M}_p^{\dagger}(y_p)], \tag{5.13}$$

where the measurement operators

$$\hat{M}_q(y_q)=\int dq\left(\frac{\epsilon_q}{2\pi}\right)^{1/4}\exp\left[-\frac{\epsilon_q}{4}(y_q-q)^2\right]|q\rangle\langle q|, \tag{5.14}$$

$$\hat{M}_p(y_p)=\int dp\left(\frac{\epsilon_p}{2\pi}\right)^{1/4}\exp\left[-\frac{\epsilon_p}{4}(y_p-p)^2\right]|p\rangle\langle p| \tag{5.15}$$

are assumed to be Gaussian and backaction evading. After some algebra,

$$P(y_q,y_p)=\int dqdp\left(\frac{\epsilon_q}{2\pi}\right)^{1/2}\left(\frac{\epsilon_p}{2\pi}\right)^{1/2}\times\exp\left[-\frac{\epsilon_q}{2}(y_q-q)^2-\frac{\epsilon_p}{2}(y_p-p)^2\right]\tilde{P}(q,p), \tag{5.16}$$

$$\tilde{P}(q,p)\equiv\frac{1}{2\pi\mathcal{C}}\int dudv\ \exp\left(-\frac{\epsilon_q u^2+\epsilon_p v^2}{8}\right)\left\langle p+\frac{v}{2}\middle|\hat{g}(\tau)\middle|p-\frac{v}{2}\right\rangle\exp(ivq)\left\langle q-\frac{u}{2}\middle|\hat{f}(\tau)\middle|q+\frac{u}{2}\right\rangle\exp(ipu). \tag{5.17}$$

From the perspective of classical probability theory, Eq. (5.16) can be interpreted as the probability density of noisy position and momentum measurements with noise variances $1/\epsilon_q$ and $1/\epsilon_p$ when the measured object has a classical phase-space density given by $\tilde{P}(q,p)$. In the limit of infinitesimally weak measurements, $\epsilon_q,\epsilon_p\to 0$ and

$$\lim_{\epsilon_q,\epsilon_p\to 0}\tilde{P}(q,p)=h(q,p,\tau). \tag{5.18}$$

Thus, $h(q,p,\tau)$ can be obtained approximately from an experiment with small ϵ_q and ϵ_p by measuring $P(y_q,y_p)$ for the same $\hat{g}$ and $\hat{f}$ and deconvolving Eq. (5.16). In practice, ϵ_q and ϵ_p only need to be small enough such that $\tilde{P}(q,p)\approx h(q,p,\tau)$. This allows one, at least in principle, to experimentally demonstrate the sub-Heisenberg uncertainties predicted in Ref. [30] in a frequentist way, not just by Bayesian estimation as described in Ref. [30]. Note, however, that $h(q,p,\tau)$ can still go negative, so it cannot always be regarded as a classical probability density. This underlines the wave nature of a quantum object and may be related to the negative probabilities encountered in the use of weak values to explain Hardy's paradox [34].

VI. CONCLUSION

In conclusion, I have used a discrete-time approach to derive the classical and quantum theories of time-symmetric smoothing. The hybrid smoothing theory is applied to the design of PLL, and the relation between the proposed theory and weak value theory of Aharonov *et al.* is discussed. Possible generalizations of the theory include taking jumps into account for the classical random process [9] and adding quantum measurements with Poisson statistics, such as photon counting [18,21–23]. Potential applications not discussed in this paper include cavity quantum electrodynamics [18,21–23], photodetection theory [16,18,23], atomic magnetometry [35], and quantum information processing in general. On a more fundamental level, it might also be interesting to generalize the weak value theory and the smoothing quasiprobability distribution to other kinds of quantum degrees of freedom in addition to position and momentum, such as spin, photon number, and phase. A general quantum smoothing theory would complete the correspondence between classical and quantum estimation theories.

ACKNOWLEDGMENTS

Discussions with Seth Lloyd and Jeffrey Shapiro are gratefully acknowledged. This work was financially supported by the Keck Foundation Center for Extreme Quantum Information Theory.

[1] A. H. Jazwinski, *Stochastic Processes and Filtering Theory* (Academic Press, New York, 1970).
[2] J. C. Crassidis and J. L. Junkins, *Optimal Estimation of Dynamic Systems* (Chapman and Hall, Boca Raton, 2004).
[3] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I* (Wiley, New York, 2001); *Detection, Estimation, and Modulation Theory, Part II: Nonlinear Modulation Theory* (Wiley, New York, 2002); *Detection, Estimation, and Modulation Theory, Part III: Radar-Sonar Processing and Gaussian Signals in Noise* (Wiley, New York, 2001).
[4] D. Simon, *Optimal State Estimation* (Wiley, Hoboken, 2006).
[5] M. Tsang, Phys. Rev. Lett. **102**, 250403 (2009).
[6] E. Pardoux, Stochastics **6**, 193 (1982); see also B. D. O. Anderson and I. B. Rhodes, *ibid.* **9**, 139 (1983).
[7] D. W. Berry and H. M. Wiseman, Phys. Rev. A **73**, 063824 (2006).
[8] Y. Aharonov, D. Z. Albert, and L. Vaidman, Phys. Rev. Lett. **60**, 1351 (1988); Y. Aharonov and L. Vaidman, J. Phys. A **24**, 2315 (1991).
[9] C. W. Gardiner, *Handbook of Stochastic Methods* (Springer-Verlag, Berlin, 1985).
[10] R. L. Stratonovich, Theor. Probab. Appl. **5**, 156 (1960); H. J. Kushner, J. Math. Anal. Appl. **8**, 332 (1964); SIAM J. Control **2**, 106 (1964).
[11] M. Fujisaki, G. Kallianpur, and H. Kunita, Osaka J. Math. **9**, 19 (1972).
[12] P. A. Frost and T. Kailath, IEEE Trans. Autom. Control **16**, 217 (1971).
[13] R. E. Mortensen, Ph.D. thesis, University of California–Berkeley, 1966; T. E. Duncan, Ph.D. thesis, Stanford University, 1967; M. Zakai, Z. Wahrscheinlichkeitstheor. Verwandte Geb. **11**, 230 (1969).
[14] D. Q. Mayne, Automatica **4**, 73 (1966); D. C. Fraser and J. E. Potter, IEEE Trans. Autom. Control **14**, 387 (1969).
[15] J. E. Wall, Jr., A. S. Willsky, and N. R. Sandell, Jr., Stochastics **5**, 1 (1981).
[16] P. Warszawski, H. M. Wiseman, and H. Mabuchi, Phys. Rev. A **65**, 023802 (2002); P. Warszawski and H. M. Wiseman, J. Opt. B: Quantum Semiclassical Opt. **5**, 1 (2003); **5**, 15 (2003); N. P. Oxtoby, P. Warszawski, H. M. Wiseman, He-Bi Sun, and R. E. S. Polkinghorne, Phys. Rev. B **71**, 165317 (2005).
[17] I. V. Aleksandrov, Z. Naturforsch. C **36A**, 902 (1981); W. Boucher and J. Traschen, Phys. Rev. D **37**, 3522 (1988); L. Diósi, N. Gisin, and W. T. Strunz, Phys. Rev. A **61**, 022108 (2000).
[18] C. W. Gardiner and P. Zoller, *Quantum Noise* (Springer-Verlag, Berlin, 2000).
[19] H. M. Wiseman and L. Diósi, Chem. Phys. **268**, 91 (2001).
[20] V. P. Belavkin, Radiotekh. Elektron. (Moscow) **25**, 1445 (1980); *Information Complexity and Control in Quantum Physics*, edited by A. Blaquière, S. Diner, and G. Lochak (Springer, Vienna, 1987), p. 311; *Stochastic Methods in Mathematics and Physics*, edited by R. Gielerak and W. Karwowski (World Scientific, Singapore, 1989), p. 310; *Modeling and Control of Systems in Engineering, Quantum Mechanics, Economics, and Biosciences*, edited by A. Blaquière (Springer, Berlin, 1989), p. 245.
[21] A. Barchielli, L. Lanz, and G. M. Prosperi, Nuovo Cimento Soc. Ital. Fis., B **72**, 79 (1982) ; Found. Phys. **13**, 779 (1983); H. Carmichael, *An Open Systems Approach to Quantum Optics* (Springer-Verlag, Berlin, 1993).
[22] D. F. Walls and G. J. Milburn, *Quantum Optics* (Springer-Verlag, Berlin, 2008).
[23] L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics* (Cambridge University Press, Cambridge, 1995).
[24] S. D. Personick, IEEE Trans. Inf. Theory **17**, 240 (1971).
[25] H. M. Wiseman, Phys. Rev. Lett. **75**, 4587 (1995).
[26] M. A. Armen, J. K. Au, J. K. Stockton, A. C. Doherty, and H. Mabuchi, Phys. Rev. Lett. **89**, 133602 (2002).
[27] D. W. Berry and H. M. Wiseman, Phys. Rev. A **65**, 043803 (2002).
[28] D. T. Pope, H. M. Wiseman, and N. K. Langford, Phys. Rev. A **70**, 043812 (2004).
[29] M. Tsang, J. H. Shapiro, and S. Lloyd, Phys. Rev. A **78**, 053820 (2008).
[30] M. Tsang, J. H. Shapiro, and S. Lloyd, Phys. Rev. A **79**, 053843 (2009).
[31] S. M. Barnett, D. T. Pegg, J. Jeffers, O. Jedrkiewicz, and R. Loudon, Phys. Rev. A **62**, 022313 (2000); S. M. Barnett, D. T. Pegg, J. Jeffers, and O. Jedrkiewicz, Phys. Rev. Lett. **86**, 2455 (2001); D. T. Pegg, S. M. Barnett, and J. Jeffers, Phys. Rev. A **66**, 022106 (2002).
[32] M. Yanagisawa, e-print arXiv:0711.3885.
[33] H. M. Wiseman, Phys. Rev. A **65**, 032111 (2002).
[34] Y. Aharonov, A. Botero, S. Popescu, B. Reznik, and J. Tollaksen, Phys. Lett. A **301**, 130 (2002).
[35] D. Budker, W. Gawlik, D. F. Kimball, S. M. Rochester, V. V. Yashchuk, and A. Weis, Rev. Mod. Phys. **74**, 1153 (2002); J. M. Geremia, J. K. Stockton, A. C. Doherty, and H. Mabuchi, Phys. Rev. Lett. **91**, 250801 (2003).