Optimal waveform estimation for classical and quantum systems via time-symmetric smoothing. II. Applications to atomic magnetometry and Hardy's paradox

Optimal waveform estimation for classical and quantum systems via time-symmetric smoothing. II. Applications to atomic magnetometry and Hardy's paradox

Mankei Tsang*

Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

(Received 19 October 2009; published 26 January 2010)

The time-symmetric quantum smoothing theory [Tsang, Phys. Rev. Lett. **102**, 250403 (2009); Phys. Rev. A **80**, 033840 (2009)] is extended to account for discrete jumps in the classical random process to be estimated, discrete variables in the quantum system, such as spin, angular momentum, and photon number, and Poisson measurements, such as photon counting. The extended theory is used to model atomic magnetometers and study Hardy's paradox in phase space.

DOI: 10.1103/PhysRevA.81.013824 PACS number(s): 03.65.Ta, 03.65.Ud, 03.65.Yz

I. INTRODUCTION

In previous papers [1–3], I have proposed a quantum smoothing theory, which is a generalization of the time-symmetric approach to quantum mechanics pioneered by Aharonov *et al.* [4] and Barnett *et al.* [5] and can be used to optimally estimate classical signals coupled to quantum sensors under continuous measurements, such as gravitational wave detectors and atomic magnetometers. Smoothing can be significantly more accurate than current quantum filtering methods [6–14] when the classical signal is a stochastic process and delay is permitted in the estimation. The accuracy improvement due to smoothing for quantum optical phase estimation via phase-locked loops has recently been experimentally demonstrated by Wheatley *et al.* [15].

While Refs. [1–3] focus on diffusive classical random processes, quantum systems with continuous degrees of freedom, and Gaussian measurements, the aim of this paper is to extend the theory to account for discrete variables in the systems and the measurements. In particular, I shall consider discrete jumps in the classical random process, discrete variables in the quantum system, such as spin, angular momentum, and photon number, and Poisson measurements, such as photon counting. Such extensions are especially important for the modeling of atomic magnetometry [11–14,16–18].

In the case of atomic magnetometry, the importance of estimation delay was discovered by Petersen and Mølmer [18], who found that the estimation of a fluctuating magnetic field modeled as an Ornstein-Uhlenbeck process becomes more accurate when the estimation is delayed and observations at later times are taken into account. I shall generalize their results using the quantum smoothing theory, derive the optimal strategy of delayed estimation for atomic magnetometry, and discuss practical methods of implementing the strategy.

A different kind of estimation problem comes up in Hardy's paradox [19], in which one tries to estimate the positions of an electron and a positron in two overlapping interferometers based on the initial conditions and measurement outcomes. While this kind of retrodictive estimation is not allowed in conventional predictive quantum theory, it can be regarded as a smoothing problem from the perspective of estimation theory, and I shall demonstrate that the salient features of the paradox can be reproduced mathematically using the quantum smoothing theory in discrete phase space. It is shown that the negativity of the predictive Wigner distribution can be regarded as the culprit for the disagreement between classical reasoning and quantum mechanics. This phase-space approach is somewhat different from Aharonov *et al.*'s weak value approach [4,20], but more similar to Feynman's attempt to formulate quantum mechanics in terms of negative probabilities [21]. Whether the two can be reconciled remains to be seen.

This paper is organized as follows. Section II reviews the classical filtering and smoothing equations when the system process has jumps and the observations have Poisson statistics, as derived by Snyder [22,23] and Pardoux [24]. Section III generalizes such equations to the quantum regime for smoothing of classical random processes coupled to quantum systems. Sec. IV converts the quantum equations to equivalent phase-space equations for discrete Wigner distributions. Sec. V studies the application of the theory to atomic magnetometry. Sec. VI studies Hardy's paradox using quantum smoothing in discrete phase space.

II. CLASSICAL FILTERING AND SMOOTHING FOR POISSON OBSERVATIONS

Define x_t as the classical system random process, the *a priori* probability density of which satisfies the differential Chapman-Kolmogorov equation [25]

$$\frac{\partial P(x,t)}{\partial t} = \mathcal{L}_C P(x,t), \tag{2.1}$$

$$\begin{aligned}\mathcal{L}_C P(x,t) \equiv & -\sum_{\mu}\frac{\partial}{\partial x_\mu}[A_\mu P(x,t)] \\ & + \frac{1}{2}\sum_{\mu,\nu}\frac{\partial^2}{\partial x_\mu \partial x_\nu}[B_{\mu\nu}P(x,t)] \\ & + \int dx'[J(x|x',t)P(x',t) \\ & - J(x'|x,t)P(x,t)],\end{aligned} \tag{2.2}$$

where $J(x|x',t)$ is the probability density per unit time that x_t will jump from x' to x. For an observation with Poisson noise,

*mankei@mit.edu

the observation probability density is

$$P(\delta\Lambda_{\mu t}|x_t) = \frac{[\lambda_\mu(x_t,t)\delta t]^{\delta\Lambda_\mu}}{\delta\Lambda_\mu!}\exp\left[-\lambda_\mu(x_t,t)\,\delta t\right]. \tag{2.3}$$

The *a posteriori* density given an observation $\delta\Lambda_{\mu t}$ is then determined by the Bayes theorem:

$$P(x_t|\delta\Lambda_{\mu t}) = \frac{P(\delta\Lambda_{\mu t}|x_t)P(x_t)}{\int dx_t(\text{numerator})}. \tag{2.4}$$

In the continuous-time limit,

$$d\Lambda_{\mu t}^2 = d\Lambda_{\mu t}, \tag{2.5}$$
$$P(d\Lambda_{\mu t} = 0|x_t) = 1 - \lambda_\mu(x_t,t)\,dt, \tag{2.6}$$
$$P(d\Lambda_{\mu t} = 1|x_t) = \lambda_\mu(x_t,t)\,dt. \tag{2.7}$$

Defining the observation record in the time interval $t_0 \leqslant t < t$ as

$$d\Lambda_{[t_0,t)} = \{d\Lambda_t, t_0 \leqslant t < t\}, \tag{2.8}$$

and the filtering probability density as the probability density of x_t conditioned upon past observations, given by

$$F(x,t) \equiv P(x_t = x|d\Lambda_{[t_0,t)}), \tag{2.9}$$

the Itō stochastic differential equation for $F(x,t)$ can be derived by evolving $F(x,t)$ according to the Chapman-Kolmogorov equation given by Eq. (2.1) in discrete time, applying the Bayes theorem given by Eq. (2.4), and then taking the continuous limit. The result is called the Snyder equation and is given by [22,23]

$$dF = dt\mathcal{L}_C F + \sum_\mu (d\Lambda_{\mu t} - \langle\lambda_\mu\rangle_F dt)\left(\frac{\lambda_\mu}{\langle\lambda_\mu\rangle_F} - 1\right)F, \tag{2.10}$$

$$\langle\lambda_\mu\rangle_F \equiv \int dx\lambda_\mu(x,t)F(x,t). \tag{2.11}$$

To derive a linear stochastic equation for an unnormalized $F(x,t)$, rewrite the observation density in Eq. (2.3) as

$$\begin{aligned} P(\delta\Lambda_{\mu t}|x_t) &= \tilde{P}(\delta\Lambda_{\mu t})\left(\frac{\lambda_\mu}{\alpha_\mu}\right)^{\delta\Lambda_\mu}\exp\left[-(\lambda_\mu - \alpha_\mu)\,\delta t\right] \\ &= \tilde{P}(\delta\Lambda_{\mu t})\bigg[1 + \delta\Lambda_{\mu t}\left(\frac{\lambda_\mu}{\alpha_\mu} - 1\right) \\ &\quad - \delta t(\lambda_\mu - \alpha_\mu) + o(\delta t)\bigg], \end{aligned} \tag{2.12}$$

$$\tilde{P}(\delta\Lambda_{\mu t}) \equiv \frac{(\alpha_\mu\delta t)^{\delta\Lambda_{\mu t}}}{\delta\Lambda_{\mu t}!}\exp(-\alpha_\mu\delta t), \tag{2.13}$$

where α_μ is an arbitrary positive parameter and $o(\delta t)$ are terms asymptotically smaller than δt. When this form of the observation density is used in the Bayes theorem given by Eq. (2.4), $\tilde{P}(\delta\Lambda_{\mu t})$ appears in both the numerator and the denominator and cancels itself, so it can be neglected in the filtering dynamics. If one does not insist on the normalization of the *a posteriori* density, the denominator of Eq. (2.4), which does not depend on x_t, can also be neglected. The resulting stochastic equation for $f(x,t) \propto P(x_t = x, d\Lambda_{[t_0,t)})$ is given by [24]

$$df = dt\mathcal{L}_C f + \sum_\mu (d\Lambda_{\mu t} - \alpha_\mu dt)\left(\frac{\lambda_\mu}{\alpha_\mu} - 1\right)f. \tag{2.14}$$

$f(x,t)$ is related to the joint probability density $P(x_t, d\Lambda_{[t_0,t)})$ by factors of $\tilde{P}(d\Lambda_{\mu t})$. The filtering density is thus

$$F(x,t) = \frac{f(x,t)}{\int dx f(x,t)}. \tag{2.15}$$

For the optimal estimation of x_τ at time τ, one should also take into account the observations after time τ and perform smoothing. To perform smoothing in the time-symmetric form [24], first solve for an unnormalized retrodictive likelihood function $P(d\Lambda_{[t,T)}|x_t = x) \propto g(x,t)$ using the adjoint of Eq. (2.14),

$$-dg = dt\mathcal{L}_C^* g + \sum_\mu (d\Lambda_{\mu t} - \alpha_\mu dt)\left(\frac{\lambda_\mu}{\alpha_\mu} - 1\right)g, \tag{2.16}$$

to be solved backward in time with final condition $g(x,T) \propto 1$. Similar to $f(x,t)$, $g(x,t)$ is related to $P(d\Lambda_{[t,T)}|x_t)$ by factors of $\tilde{P}(d\Lambda_{\mu t})$. The smoothing probability density at time τ given the observation record $d\Lambda_{[t_0,T)}$ in the time interval $t_0 \leqslant \tau \leqslant T$ is then

$$P(x_\tau = x|d\Lambda_{[t_0,T)}) = \frac{g(x,\tau)f(x,\tau)}{\int dx g(x,\tau)f(x,\tau)}. \tag{2.17}$$

III. HYBRID CLASSICAL-QUANTUM FILTERING AND SMOOTHING FOR POISSON OBSERVATIONS

Using the same approach as Refs. [2,3], it is not difficult to generalize the above classical equations to the quantum regime for hybrid classical-quantum filtering and smoothing. I shall first consider the problem in discrete time before taking the appropriate continuous limit. Define x_t as the classical system process that one wishes to estimate, which is coupled to a quantum system under measurements. As before, the quantum backaction from the quantum system to the classical one is assumed to be negligible. Define the hybrid density operator that describes the joint statistics of the classical and quantum systems [26] as $\hat{\rho}(x,t)$, so that the marginal probability distribution of x_t is $\text{tr}[\hat{\rho}(x,t)]$ and the quantum density operator is $\int dx\hat{\rho}(x,t)$. The *a priori* evolution of $\hat{\rho}(x_t,t)$ is governed by

$$\frac{\partial\hat{\rho}(x,t)}{\partial t} = \mathcal{L}\hat{\rho}(x,t), \tag{3.1}$$
$$\mathcal{L}\hat{\rho}(x,t) \equiv \mathcal{L}_0\hat{\rho}(x,t) + \mathcal{L}_I(x)\hat{\rho}(x,t) + \mathcal{L}_C\hat{\rho}(x,t), \tag{3.2}$$

where $\mathcal{L}_0$ is the superoperator that governs the evolution of the quantum system, $\mathcal{L}_I$ is the superoperator that describes the coupling of the classical system to the quantum system, via an interaction Hamiltonian for example, and $\mathcal{L}_C$ is the Chapman-Kolmogorov operator defined by Eq. (2.2). The measurement, on the other hand, is described by the quantum Bayes theorem

$$\hat{\rho}(x_t|\delta\Lambda_{\mu t}) = \frac{\hat{M}(\delta\Lambda_{\mu t}|x_t)\hat{\rho}(x_t)\hat{M}^\dagger(\delta\Lambda_{\mu t}|x_t)}{\int dx_t\text{tr}(\text{numerator})}, \tag{3.3}$$

where the measurement operator with Poisson statistics is

$$\hat{M} = \frac{[\hat{L}_\mu(x_t,t)\sqrt{\delta t}]^{\delta\Lambda_{\mu t}}}{\sqrt{\delta\Lambda_{\mu t}!}} \exp\left[-\frac{\delta t}{2}\hat{L}_\mu^\dagger(x_t,t)\hat{L}_\mu(x_t,t)\right]. \tag{3.4}$$

$\hat{L}_\mu(x_t,t)$ is a hybrid operator, an annihilation operator for example, and can also depend on x_t. In the continuous-time limit, the stochastic differential equation for the filtering hybrid density operator, defined as

$$\hat{F}(x,t) \equiv \hat{\rho}(x_t = x|d\Lambda_{[t_0,t)}), \tag{3.5}$$

is given by [8,27]

$$d\hat{F} = dt\mathcal{L}\hat{F} + dt\sum_\mu\left(\hat{L}_\mu\hat{F}\hat{L}_\mu^\dagger - \frac{1}{2}\hat{L}_\mu^\dagger\hat{L}_\mu\hat{F} - \frac{1}{2}\hat{F}\hat{L}_\mu^\dagger\hat{L}_\mu\right) + \sum_\mu(d\Lambda_{\mu t} - \langle\hat{L}_\mu^\dagger\hat{L}_\mu\rangle_{\hat{F}}dt)\left(\frac{\hat{L}_\mu\hat{F}\hat{L}_\mu^\dagger}{\langle\hat{L}_\mu^\dagger\hat{L}_\mu\rangle_{\hat{F}}} - \hat{F}\right), \tag{3.6}$$

where

$$\langle\hat{L}_\mu^\dagger\hat{L}_\mu\rangle_{\hat{F}} \equiv \int dx\,\mathrm{tr}\,[\hat{L}_\mu^\dagger(x,t)\hat{L}_\mu(x,t)\hat{F}(x,t)]. \tag{3.7}$$

Equation (3.6) is a quantum generalization of the Snyder equation [Eq. (2.10)]. To derive a linear version of Eq. (3.6) for an unnormalized filtering operator, analogous to Eq. (2.14), rewrite $\hat{M}$ as

$$\begin{aligned}\hat{M} &= \sqrt{\tilde{P}(\delta\Lambda_{\mu t})}\left(\frac{\hat{L}_\mu}{\sqrt{\alpha_\mu}}\right)^{\delta\Lambda_{\mu t}} \exp\left[-\frac{\delta t}{2}(\hat{L}_\mu^\dagger\hat{L}_\mu - \alpha_\mu)\right] \\ &= \sqrt{\tilde{P}}\left[\hat{1} + \delta\Lambda_{\mu t}\left(\frac{\hat{L}_\mu}{\sqrt{\alpha_\mu}} - \hat{1}\right) - \frac{\delta t}{2}(\hat{L}_\mu^\dagger\hat{L}_\mu - \alpha_\mu) + o(\delta t)\right].\end{aligned} \tag{3.8}$$

Similar to the classical case, the *a posteriori* state calculated using Eq. (3.3) does not depend on $\tilde{P}(\delta\Lambda_{\mu t})$, as it appears in both the numerator and denominator of Eq. (3.3) and cancels itself. The denominator of Eq. (3.3) can also be omitted if one does not insist on normalization. The resulting equation in the continuous limit is [27]

$$d\hat{f} = dt\mathcal{L}\hat{f} + dt\sum_\mu\left(\hat{L}_\mu\hat{f}\hat{L}_\mu^\dagger - \frac{1}{2}\hat{L}_\mu^\dagger\hat{L}_\mu\hat{f} - \frac{1}{2}\hat{f}\hat{L}_\mu^\dagger\hat{L}_\mu\right) + \sum_\mu(d\Lambda_{\mu t} - \alpha_\mu dt)\left(\alpha_\mu^{-1}\hat{L}_\mu\hat{f}\hat{L}_\mu^\dagger - \hat{f}\right), \tag{3.9}$$

$$\hat{F}(x,t) = \frac{\hat{f}(x,t)}{\int dx\,\mathrm{tr}\,[\hat{f}(x,t)]}. \tag{3.10}$$

The classical incoherent limit of Eq. (3.9) is obviously Eq. (2.14) and Eq. (3.9) can be verified against Eq. (3.6) by normalizing the former using Itō calculus. The derivation of Eq. (3.9) can be made more rigorous using the reference probability approach [27].

To perform optimal estimation of x_τ at time τ, one also needs to solve for the unnormalized hybrid effect operator $\hat{E}(d\Lambda_{[\tau,T)}|x_\tau = x) \propto \hat{g}(x,\tau)$ using the adjoint of Eq. (3.9) and observations after time τ [2,3]:

$$-d\hat{g} = dt\mathcal{L}^*\hat{g} + dt\sum_\mu\left(\hat{L}_\mu^\dagger\hat{g}\hat{L}_\mu - \frac{1}{2}\hat{g}\hat{L}_\mu^\dagger\hat{L}_\mu - \frac{1}{2}\hat{L}_\mu^\dagger\hat{L}_\mu\hat{g}\right) + \sum_\mu(d\Lambda_{\mu t} - \alpha_\mu dt)\left(\alpha_\mu^{-1}\hat{L}_\mu^\dagger\hat{g}\hat{L}_\mu - \hat{g}\right), \tag{3.11}$$

where the final condition is $\hat{g}(x,T) \propto \hat{1}$ and the adjoint is defined with respect to the hybrid Hilbert-Schmidt inner product

$$\langle\hat{g}(x,t),\hat{f}(x,t)\rangle \equiv \int dx\,\mathrm{tr}\,[\hat{g}(x,t)\hat{f}(x,t)], \tag{3.12}$$

$$\langle\hat{g}(x,t),\mathcal{L}\hat{f}(x,t)\rangle = \langle\mathcal{L}^*\hat{g}(x,t),\hat{f}(x,t)\rangle. \tag{3.13}$$

The smoothing probability density is then

$$h(x,\tau) \equiv P(x_\tau = x|d\Lambda_{[t_0,T)}) = \frac{\mathrm{tr}[\hat{g}(x,\tau)\hat{f}(x,\tau)]}{\int dx\,\mathrm{tr}\,[\hat{g}(x,\tau)\hat{f}(x,\tau)]}. \tag{3.14}$$

Incorporating the Gaussian measurements considered in Refs. [2,3] into the equations above is straightforward. This is useful, for example, when both photon counting and homodyne detection are performed in a quantum optics experiment [28]. With Poisson observations $d\Lambda_t$ and Gaussian observations dy_t, the resulting filtering equation for $\hat{F}(x,t)$ is

$$\begin{aligned} d\hat{F} &= dt\mathcal{L}\hat{F} + dt\sum_\mu\left(\hat{L}_\mu\hat{F}\hat{L}_\mu^\dagger - \frac{1}{2}\hat{L}_\mu^\dagger\hat{L}_\mu\hat{F} - \frac{1}{2}\hat{F}\hat{L}_\mu^\dagger\hat{L}_\mu\right) \\ &\quad + \frac{dt}{8}\left(2\hat{C}^T R^{-1}\hat{F}\hat{C}^\dagger - \hat{C}^{\dagger T}R^{-1}\hat{C}\hat{F} - \hat{F}\hat{C}^{\dagger T}R^{-1}\hat{C}\right) \\ &\quad + \sum_\mu\left(d\Lambda_{\mu t} - \langle\hat{L}_\mu^\dagger\hat{L}_\mu\rangle_{\hat{F}}\,dt\right)\left(\frac{\hat{L}_\mu\hat{F}\hat{L}_\mu^\dagger}{\langle\hat{L}_\mu^\dagger\hat{L}_\mu\rangle_{\hat{F}}} - \hat{F}\right) \\ &\quad + \frac{1}{2}[(\hat{C} - \langle\hat{C}\rangle_{\hat{F}})^T R^{-1}d\eta_t\hat{F} + \mathrm{H.c.}],\end{aligned} \tag{3.15}$$

$$d\eta_t \equiv dy_t - \frac{dt}{2}\langle\hat{C} + \hat{C}^\dagger\rangle_{\hat{F}}, \tag{3.16}$$

where $\hat{C} = \hat{C}(x,t)$ is a vector of hybrid operators, $R = R(t)$ is a positive-definite matrix, $d\eta_t$ is a vectoral Wiener increment with covariance matrix $R(t)dt$, and H.c. denotes Hermitian conjugate.

The equation for $\hat{f}(x,t)$ is

$$\begin{aligned} d\hat{f} &= dt\mathcal{L}\hat{f} + dt\sum_\mu\left(\hat{L}_\mu\hat{f}\hat{L}_\mu^\dagger - \frac{1}{2}\hat{L}_\mu^\dagger\hat{L}_\mu\hat{f} - \frac{1}{2}\hat{f}\hat{L}_\mu^\dagger\hat{L}_\mu\right) \\ &\quad + \frac{dt}{8}(2\hat{C}^T R^{-1}\hat{f}\hat{C}^\dagger - \hat{C}^{\dagger T}R^{-1}\hat{C}\hat{f} - \hat{f}\hat{C}^{\dagger T}R^{-1}\hat{C}) \\ &\quad + \sum_\mu(d\Lambda_{\mu t} - \alpha_\mu dt)\left(\alpha_\mu^{-1}\hat{L}_\mu\hat{f}\hat{L}_\mu^\dagger - \hat{f}\right) \\ &\quad + \frac{1}{2}(\hat{C}^T R^{-1}dy_t\hat{f} + \mathrm{H.c.}),\end{aligned} \tag{3.17}$$

and for $\hat{g}(x,t)$,

$$-d\hat{g} = dt\mathcal{L}^*\hat{g} + dt\sum_\mu\left(\hat{L}_\mu^\dagger\hat{g}\hat{L}_\mu - \frac{1}{2}\hat{g}\hat{L}_\mu^\dagger\hat{L}_\mu - \frac{1}{2}\hat{L}_\mu^\dagger\hat{L}_\mu\hat{g}\right) + \frac{dt}{8}(2\hat{C}^{\dagger T}R^{-1}\hat{g}\hat{C} - \hat{g}\hat{C}^{\dagger T}R^{-1}\hat{C} - \hat{C}^{\dagger T}R^{-1}\hat{C}\hat{g}) + \sum_\mu (d\Lambda_{\mu t} - \alpha_\mu dt)\left(\alpha_\mu^{-1}\hat{L}_\mu^\dagger\hat{g}\hat{L}_\mu - \hat{g}\right) + \frac{1}{2}(\hat{g}\hat{C}^T R^{-1}dy_t + \text{H.c.}). \tag{3.18}$$

IV. QUANTUM SMOOTHING IN PHASE SPACE

One method of solving Eqs. (3.14), (3.17), and (3.18) for hybrid smoothing is to use Wigner distributions [2,3]. For a quantum system with discrete degrees of freedom, such as spin, angular momentum, or an N-level system, one may define the discrete Wigner distribution, according to Feynman [21] and Wootters [29], as

$$f(q,p,x,t) \equiv \frac{1}{N}\text{tr}\,[\hat{f}(x,t)\hat{W}(q,p)], \tag{4.1}$$

$$q \in \{0,1,\ldots,N-1\}, \tag{4.2}$$

$$p \in \{0,1,\ldots,N-1\}. \tag{4.3}$$

The operator $\hat{W}(q,p)$ for prime N is

$$\hat{W}(q,p) \equiv \begin{cases} \frac{1}{2}[(-1)^q\hat{\sigma}_z + (-1)^p\hat{\sigma}_x + (-1)^{q+p}\hat{\sigma}_y + \hat{1}], & N=2; \\ \sum_{q_1,q_2}\delta_{2q,q_1+q_2}\exp\left[\frac{2\pi i}{N}p(q_1-q_2)\right]|q_1\rangle\langle q_2|, & N>2. \end{cases} \tag{4.4}$$

$\hat{\sigma}_x$, $\hat{\sigma}_y$, and $\hat{\sigma}_z$ are Pauli matrices, $|q_1\rangle$ and $|q_2\rangle$ are eigenstates of $\hat{q}$, and modular arithmetic with modulus N is implicitly assumed. For a nonprime N, the system can be decomposed into subsystems with prime N's and the Wigner distribution can be defined using $\hat{W}(q,p)$ for each subsystem [29].

An alternative definition in a $2N\times 2N$ phase space, first suggested by Hannay and Berry [30], is

$$\tilde{f}(q,p,x,t) \equiv \frac{1}{2N}\text{tr}\,[\hat{f}(x,t)\hat{w}(q,p)],$$

$$q \in \left\{0,\frac{1}{2},\ldots,N-\frac{1}{2}\right\},$$

$$p \in \left\{0,\frac{1}{2},\ldots,N-\frac{1}{2}\right\},$$

$$\hat{w}(q,p) \equiv \sum_u \exp\left(\frac{4\pi i p u}{N}\right)|q+u\rangle\langle q-u|,$$

$$u \in \left\{-\frac{N}{2}+\frac{1}{2}, -\frac{N}{2}+1,\ldots,\frac{N}{2}\right\}, \tag{4.5}$$

where the matrix elements with noninteger indices are assumed to be zero. One may also use either Wigner function to describe the energy level n and phase ϕ of a harmonic oscillator by letting $n=q$, $\phi = 2\pi p/N$, and taking the $N\to\infty$ limit at the end of a calculation [31,32].

Both definitions have a particularly desirable property for the purpose of smoothing, namely,

$$\text{tr}\,[\hat{g}(x,t)\hat{f}(x,t)] = N\sum_{q,p} g(q,p,x,t)f(q,p,x,t) = 2N\sum_{q,p}\tilde{g}(q,p,x,t)\tilde{f}(q,p,x,t), \tag{4.6}$$

so the smoothing probability density can be written in terms of the Wigner distributions as

$$h(x,\tau) = \frac{\sum_{q,p} g(q,p,x,\tau)f(q,p,x,\tau)}{\int dx \sum_{q,p} g(q,p,x,\tau)f(q,p,x,\tau)} \tag{4.7}$$

or

$$h(x,\tau) = \frac{\sum_{q,p} \tilde{g}(q,p,x,\tau)\tilde{f}(q,p,x,\tau)}{\int dx \sum_{q,p} \tilde{g}(q,p,x,\tau)\tilde{f}(q,p,x,\tau)}. \tag{4.8}$$

Equations (4.7) and (4.8) become equivalent to the classical smoothing density given by Eq. (2.17), with the quantum degrees of freedom marginalized, if f and g or $\tilde{f}$ and $\tilde{g}$ are non-negative and can be regarded as classical probability densities. The hybrid smoothing problem can then be solved using classical filtering and smoothing techniques.

If one would like to apply smoothing to quantum degrees of freedom, Eqs. (4.7) and (4.8) also motivate the definition of a quantum smoothing quasiprobability distribution as

$$h(q,p,x,\tau) = \frac{g(q,p,x,\tau)f(q,p,x,\tau)}{\int dx \sum_{q,p} g(q,p,x,\tau)f(q,p,x,\tau)} \tag{4.9}$$

or

$$\tilde{h}(q,p,x,\tau) = \frac{\tilde{g}(q,p,x,\tau)\tilde{f}(q,p,x,\tau)}{\int dx \sum_{q,p} \tilde{g}(q,p,x,\tau)\tilde{f}(q,p,x,\tau)}. \tag{4.10}$$

From the perspective of estimation theory, these definitions of quantum smoothing distributions are arguably the most natural, for they both give the correct classical smoothing distribution when the quantum degrees of freedom are marginalized, are equivalent to the smoothing distributions in classical smoothing theory when f and g or $\tilde{f}$ and $\tilde{g}$ are nonnegative, and are explicitly normalized.

There are many other qualified definitions of the Wigner distribution in discrete or periodic phase space [33]. Choosing which definition to use depends on the application. The Feynman-Wootters distribution is defined only on the eigenvalues of $\hat{q}$ and $\hat{p}$, so it appears more physical, but the Hannay-Berry definition is easier to calculate analytically for arbitrary N and, as shown in the Appendix, naturally arises from the statistics of weak measurements.

V. ATOMIC MAGNETOMETRY

A. Optimal smoothing

An important application of quantum estimation theory is atomic magnetometry [11–14,16–18]. Consider the setup described in Refs. [12–14,16] and depicted in Fig. 1. An atomic spin ensemble is initially prepared in a coherent state with the

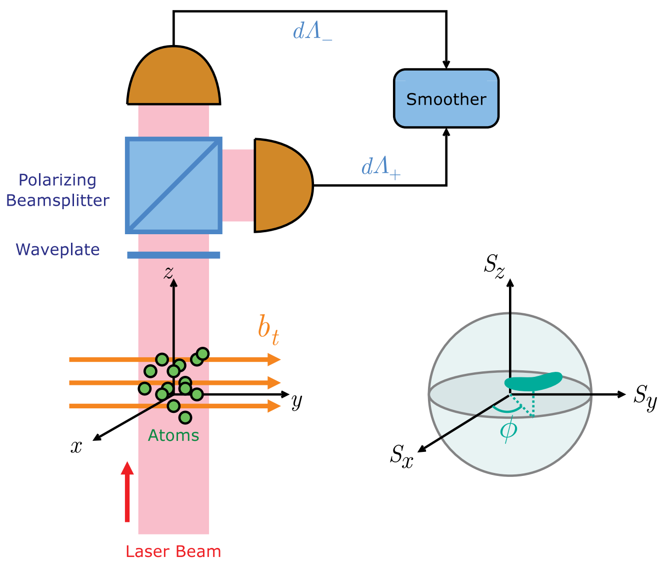

FIG. 1. (Color online) Left: basic setup of atomic magnetometer as considered in Refs. [12–14,16]. Right: the spherical phase space for spin.

mean collective spin vector along the x axis. Let the magnetic field be polarized along y axis and given by

$$b_t \equiv x_{1t}, \tag{5.1}$$

a component of the classical system process to be estimated. The magnetic field introduces Larmor precession to the spin via the interaction Hamiltonian

$$\hat{H}_I(x) = -\gamma b \hat{S}_y, \tag{5.2}$$

$$\mathcal{L}_I(x)\hat{F}(x,t) = -\frac{i}{\hbar}[\hat{H}_I(x), \hat{F}(x,t)] = \frac{i\gamma}{\hbar} b[\hat{S}_y, \hat{F}], \tag{5.3}$$

where $\hat{S}_y$ is the y component of the spin vector operator and γ is the gyromagnetic ratio. Under continuous optical polarimetry measurements, the stochastic equation for the filtering density operator $\hat{F}(x,t)$ has been derived by Bouten *et al.* [14] and is given by

$$\begin{aligned} d\hat{F} = dt\Big\{&\mathcal{L}_C\hat{F} + \frac{i\gamma}{\hbar}b[\hat{S}_y, \hat{F}] + |a|^2[\cos(\kappa\hat{m})\hat{F}\cos(\kappa\hat{m}) \\ &+ \sin(\kappa\hat{m})\hat{F}\sin(\kappa\hat{m}) - \hat{F}]\Big\} \\ &+ \sum_{\mu=+,-}(d\Lambda_{\mu t} - \langle\hat{L}_\mu^\dagger\hat{L}_\mu\rangle_{\hat{F}}dt)\left(\frac{\hat{L}_\mu\hat{F}\hat{L}_\mu^\dagger}{\langle\hat{L}_\mu^\dagger\hat{L}_\mu\rangle_{\hat{F}}} - \hat{F}\right), \end{aligned} \tag{5.4}$$

which is in the form of Eq. (3.6), with

$$\hat{m} \equiv \frac{\hat{S}_z}{\hbar}, \quad \hat{L}_\pm = \frac{a}{\sqrt{2}}[\cos(\kappa\hat{m}) \pm \sin(\kappa\hat{m})], \tag{5.5}$$

κ is the light-spin coupling parameter and a is the normalized optical envelope. The linear predictive and retrodictive equations for $\hat{f}(x,t)$ and $\hat{g}(x,t)$ become

$$\begin{aligned} d\hat{f} = dt\Big\{&\mathcal{L}_C\hat{f} + \frac{i\gamma}{\hbar}b[\hat{S}_y, \hat{f}] + |a|^2[\cos(\kappa\hat{m})\hat{f}\cos(\kappa\hat{m}) \\ &+ \sin(\kappa\hat{m})\hat{f}\sin(\kappa\hat{m}) - \hat{f}]\Big\} \\ &+ \sum_{\mu=+,-}\left(d\Lambda_{\mu t} - \frac{|a|^2}{2}dt\right)\left(\frac{2}{|a|^2}\hat{L}_\mu\hat{f}\hat{L}_\mu^\dagger - \hat{f}\right), \end{aligned} \tag{5.6}$$

$$\begin{aligned} -d\hat{g} = dt\Big\{&\mathcal{L}_C^*\hat{g} - \frac{i\gamma}{\hbar}b[\hat{S}_y, \hat{g}] + |a|^2[\cos(\kappa\hat{m})\hat{g}\cos(\kappa\hat{m}) \\ &+ \sin(\kappa\hat{m})\hat{g}\sin(\kappa\hat{m}) - \hat{g}]\Big\} \\ &+ \sum_{\mu=+,-}\left(d\Lambda_{\mu t} - \frac{|a|^2}{2}dt\right)\left(\frac{2}{|a|^2}\hat{L}_\mu^\dagger\hat{g}\hat{L}_\mu - \hat{g}\right). \end{aligned} \tag{5.7}$$

After solving Eq. (5.6) forward in time for $\hat{f}(x,\tau)$ and Eq. (5.7) backward in time for $\hat{g}(x,\tau)$, the smoothing probability distribution is given by

$$h(x,\tau) = \frac{\mathrm{tr}\,[\hat{g}(x,\tau)\hat{f}(x,\tau)]}{\int dx\,\mathrm{tr}\,[\hat{g}(x,\tau)\hat{f}(x,\tau)]}, \tag{5.8}$$

which can be used to produce the optimal estimate and the associated error of the system process x_τ, including the magnetic field $b_\tau \equiv x_{1\tau}$.

B. Smoothing in phase space

The usual strategy of solving the quantum estimation problem is to take the $S_x \gg S_y, S_x \gg S_z$ limit, assume $\hat{S}_y$ and $\hat{S}_z$ are continuous, and approximate the conditional quantum state as a Gaussian state [11–13,16,18]. This is akin to approximating the spherical phase space with a flat one near $\mathbf{S} = (S_x, 0, 0)$. While the Gaussian approximation is probably the most practical, in order to illustrate the discrete phase-space formalism proposed in Sec. IV, I shall first attempt to convert Eqs. (5.6) and (5.7) to stochastic equations for discrete Wigner distributions in the $2N \times 2N$ phase space before making further approximations.

Let

$$\hat{m} = \hat{q} - s, \quad N = 2s + 1, \tag{5.9}$$

where s is the total spin number. Then

$$\hat{\phi} \equiv \frac{2\pi\hat{p}}{N} \tag{5.10}$$

is the operator for the azimuthal angle of the spin vector. I shall use m and ϕ as the phase-space variables instead of q and p, and let

$$f(m,\phi) = \tilde{f}\left(q = m + s, p = \frac{N\phi}{2\pi}\right). \tag{5.11}$$

First consider the measurement-induced decoherence term in Eq. (5.6), which can be rewritten as

$$\mathcal{L}_M \hat{f} \equiv \tfrac{1}{2}(e^{i\kappa\hat{q}}\hat{f}e^{-i\kappa\hat{q}} + e^{-i\kappa\hat{q}}\hat{f}e^{i\kappa\hat{q}}) - \hat{f}. \tag{5.12}$$

Using the discrete Wigner function in $2N \times 2N$ phase space given by Eqs. (4.5) and $\stackrel{\hat{w}}{\to}$ to denote the transform to the $2N \times 2N$ phase space, it can be shown that

$$\mathcal{L}_M \hat{f} \stackrel{\hat{w}}{\to} \textstyle\sum_{\phi'} J(\phi - \phi') f(m, \phi', x, t) - f(m, \phi, x, t), \tag{5.13}$$

where

$$J(\phi - \phi') \equiv \frac{1}{4N}\left\{\frac{\sin[N(\phi - \phi' - \kappa)]}{\tan[(\phi - \phi' - \kappa)/2]} + \frac{\sin[N(\phi - \phi' + \kappa)]}{\tan[(\phi - \phi' + \kappa)/2]}\right\}. \tag{5.14}$$

While Eq. (5.13) has the appearance of the jump term in the Chapman-Kolmogorov equation [Eq. (2.2)], $J(\phi - \phi')$, which plays the role of a jump probability density, can become negative. In the special case of $\kappa = \pi\mu/N$, where μ is an integer, however, $J(\phi - \phi')$ is simplified to

$$J(\phi - \phi') = \tfrac{1}{2}(\delta_{\phi-\phi',\kappa} + \delta_{\phi-\phi',-\kappa}), \tag{5.15}$$

and the measurement-induced decoherence introduces random azimuthal jumps in steps of κ to the spin vector around the z axis. In the limit of $N \to \infty$, ϕ becomes approximately continuous, $\kappa \approx \pi\mu/N$, and Eq. (5.13) can be rewritten as

$$\mathcal{L}_M \hat{f} \stackrel{\hat{w}}{\to} \tfrac{1}{2}[f(m, \phi + \kappa, x, t) + f(m, \phi - \kappa, x, t)] - f(m, \phi, x, t). \tag{5.16}$$

The $N \to \infty$ limit is akin to approximating the spin system as a harmonic oscillator [31] and the spherical phase space as a cylindrical one. If $\kappa \ll \langle\Delta\hat{\phi}^2\rangle^{1/2}$, we can further make the diffusive approximation:

$$\mathcal{L}_M \hat{f} \stackrel{\hat{w}}{\to} \frac{\kappa^2}{2}\frac{\partial^2}{\partial\phi^2} f(m, \phi, x, t). \tag{5.17}$$

Next, consider the Larmor precession term $(i\gamma/\hbar)b[\hat{S}_y, \hat{f}]$. In terms of $\hat{m}$ and $\hat{\phi}$,

$$\hat{S}_y = \frac{\hbar}{2i}[\exp(-i\hat{\phi})\sqrt{s(s+1) - \hat{m}(\hat{m}+1)} - \exp(i\hat{\phi})\sqrt{s(s+1) - \hat{m}(\hat{m}-1)}]. \tag{5.18}$$

With this form, it is difficult to convert the Larmor precession term to the phase-space picture analytically, so we again make the cylindrical phase-space approximation with $s \gg \langle\hat{m}\rangle, \langle\Delta\hat{m}^2\rangle^{1/2}$, so that the spin vector distribution is concentrated near the equator. This approximation is valid when the magnetic field is small and fluctuating around zero, or a control, such as an applied magnetic field [11–13] or an adjustable direction of the optical beam, is present to realign the spin vector with respect to the optical beam propagation direction. Then

$$\hat{S}_y \approx -\hbar s \sin\hat{\phi}, \tag{5.19}$$

$$\mathcal{L}_I \hat{f} \stackrel{\hat{w}}{\to} -\gamma b s \cos\phi\left[f\left(m+\tfrac{1}{2}, \phi, x, t\right) - f\left(m-\tfrac{1}{2}, \phi, x, t\right)\right]. \tag{5.20}$$

Although this looks like the jump term in Eq. (2.2), the apparent jump probability density is again negative. To make the classical connection, assume that m is continuous and approximate the difference as a derivative:

$$\mathcal{L}_I \hat{f} \stackrel{\hat{w}}{\to} -\gamma b s \cos\phi \frac{\partial}{\partial m} f(m, \phi, x, t), \tag{5.21}$$

which becomes equivalent to the drift term in Eq. (2.2) with $A_m = \gamma b s \cos\phi$.

Finally, let us consider the terms $\hat{L}_\pm \hat{f} \hat{L}_\pm^\dagger$ in Eq. (5.6). It is not difficult to show that, in the continuous ϕ limit,

$$\hat{L}_\pm \hat{f} \hat{L}_\pm^\dagger \stackrel{\hat{w}}{\to} \frac{|a|^2}{2}\left\{\frac{1}{2}[f(m, \phi + \kappa, x, t) + f(m, \phi - \kappa, x, t)] \pm \sin(2\kappa m) f(m, \phi, x, t)\right\}. \tag{5.22}$$

These terms do not have exact analogs in the corresponding classical equation [Eq. (2.14)], unless we make the $\kappa \ll \langle\Delta\hat{\phi}^2\rangle^{1/2}$ approximation, which gives

$$\hat{L}_\pm \hat{f} \hat{L}_\pm^\dagger \stackrel{\hat{w}}{\to} \frac{|a|^2}{2}\left\{f(m, \phi, x, t) + \frac{\kappa^2}{2}\frac{\partial^2}{\partial\phi^2} f(m, \phi, x, t) \pm \sin(2\kappa m) f(m, \phi, x, t)\right\} \tag{5.23}$$

$$\approx \tfrac{|a|^2}{2}\,[1 \pm \sin(2\kappa m)]\, f(m, \phi, x, t). \tag{5.24}$$

Summarizing, a classical model of atomic magnetometry can be obtained if we approximate the spherical phase space as a cylindrical one near the equator, assume m is continuous, and let $\kappa \ll \langle\Delta\hat{\phi}^2\rangle^{1/2}$. The resulting equations for $f(m, \phi, x, t)$ and $g(m, \phi, x, t)$ are

$$df = dt\left(\mathcal{L}_C f - \gamma b s \cos\phi \frac{\partial f}{\partial m} + \frac{|a|^2\kappa^2}{2}\frac{\partial^2 f}{\partial\phi^2}\right) + \sum_{\mu=+,-}\left(d\Lambda_\mu - \frac{|a|^2}{2}dt\right)\left(\frac{2\lambda_\mu}{|a|^2} - 1\right) f, \tag{5.25}$$

$$-dg = dt\left(\mathcal{L}_C^* g + \gamma b s \cos\phi \frac{\partial g}{\partial m} + \frac{|a|^2\kappa^2}{2}\frac{\partial^2 g}{\partial\phi^2}\right) + \sum_{\mu=+,-}\left(d\Lambda_\mu - \frac{|a|^2}{2}dt\right)\left(\frac{2\lambda_\mu}{|a|^2} - 1\right) g, \tag{5.26}$$

$$\lambda_\pm = \frac{|a|^2}{2}[1 \pm \sin(2\kappa m)]. \tag{5.27}$$

The equivalent system equation for m_t is then

$$dm_t = dt\gamma b_t s \cos\phi_t, \tag{5.28}$$

where ϕ_t is a Wiener process with $d\phi_t^2 = |a|^2\kappa dt$. Unlike the Gaussian model [12,13,18], this slightly more general model shows that the z component of the spin is coupled to ϕ_t via Larmor precession, as one would expect from classical dynamics, since $S_x \approx \hbar s \cos\phi$ when $s \gg m$. The diffusion of ϕ would therefore reduce the estimation accuracy in the long run.

To make the Gaussian approximation, let $\langle\hat{\phi}_t\rangle$, $\langle\Delta\hat{\phi}_t^2\rangle^{1/2} \ll 1$, so that $\cos\phi_t \approx 1$, and let x_t be a Gaussian random process, such as the Ornstein-Uhlenbeck process [13,18]. If $\kappa\langle\hat{m}\rangle$, $\kappa\langle\Delta\hat{m}^2\rangle^{1/2} \ll 1$, and the effective noise covariances are $\langle\lambda_\pm\rangle \approx |a|^2/2$, one can use the linear Mayne-Fraser-Potter smoother [1,3,34], which combines the estimates and covariances from a predictive Kalman filter and a retrodictive Kalman filter, to produce the optimal estimate of x_t. Other equivalent linear smoothers may also be used [1].

VI. HARDY'S PARADOX IN PHASE SPACE

In this section, I shall study Hardy's paradox [19] in phase space using the quantum smoothing quasiprobability distribution defined by Eq. (4.9), which allows one to estimate quantum degrees of freedom given past and future observations in a way closely resembling classical estimation theory. The more physical and intuitive Feynman-Wootters distribution is used, since its elements all correspond to actual paths in the setup.

As a brief review of the paradox, consider two Mach-Zehnder interferometers, one for a positron and one for an electron, depicted in Fig. 2. If the interferometers are physically separate, then the setup can be configured so that the particles always arrive at the C^+ and C^- detectors, respectively. Now let us make one arm of an interferometer to overlap with an arm of the other. After the first pair of beam splitters, the two particles may meet in the overlapping arms, in which case they annihilate each other with probability 1. With this overlapping setup, there is a $1/16$ probability that the particles will arrive at D^+ and D^-, respectively, according to quantum theory.

The paradox arises when one tries to use classical reasoning to estimate which arms the particles went through. If D^+ detects a positron, then the electron must have been in the overlapping arm to somehow influence the positron to go to D^+ instead of C^+. The same reasoning can be applied when D^- detects an electron, which should mean that the positron was in the overlapping arm. But if both particles went through the overlapping arms, they should have annihilated each other and would not have been detected.

Denote the position of a particle in one arm as 0 and that in the other arm as 1, as shown in Fig. 2. At the time instant labeled 0,

$$|\Psi\rangle_0 = |0,0\rangle, \tag{6.1}$$

where the first number in the ket denotes the position of the positron, the second number denotes that of the electron, and the subscript is the time label. The corresponding two-particle Wigner distribution using Eqs. (4.1) and (4.4) is

$$f_0(q^+,q^-,p^+,p^-) = \begin{pmatrix} f_0(0,0,0,0) & f_0(0,0,0,1) & f_0(0,0,1,0) & f_0(0,0,1,1) \\ f_0(0,1,0,0) & f_0(0,1,0,1) & f_0(0,1,1,0) & f_0(0,1,1,1) \\ f_0(1,0,0,0) & f_0(1,0,0,1) & f_0(1,0,1,0) & f_0(1,0,1,1) \\ f_0(1,1,0,0) & f_0(1,1,0,1) & f_0(1,1,1,0) & f_0(1,1,1,1) \end{pmatrix} \tag{6.2}$$

$$= \frac{1}{4}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{6.3}$$

After the first pair of beam splitters,

$$|\Psi\rangle_1 = \frac{1}{2}\left(|0,0\rangle + |0,1\rangle + |1,0\rangle + |1,1\rangle\right), \tag{6.4}$$

$$f_1(q^+,q^-,p^+,p^-) = \frac{1}{4}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}. \tag{6.5}$$

There is a $1/4$ chance that an annihilation of the particles occurs. If the annihilation does *not* occur, the *a posteriori* quantum state is

$$|\Psi\rangle_2 = \frac{1}{\sqrt{3}}\left(|0,0\rangle + |0,1\rangle + |1,0\rangle\right), \tag{6.6}$$

$$f_2(q^+,q^-,p^+,p^-) = \frac{1}{12}\begin{pmatrix} 4 & 0 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 2 & 2 & 0 & 0 \\ 1 & -1 & -1 & 1 \end{pmatrix}. \tag{6.7}$$

The Wigner distribution has negative elements and can no longer be regarded as a classical phase-space probability distribution. The negative elements, as one shall see later, can be regarded as the culprits that cause the paradox. The predictive marginal distributions are still non-negative,

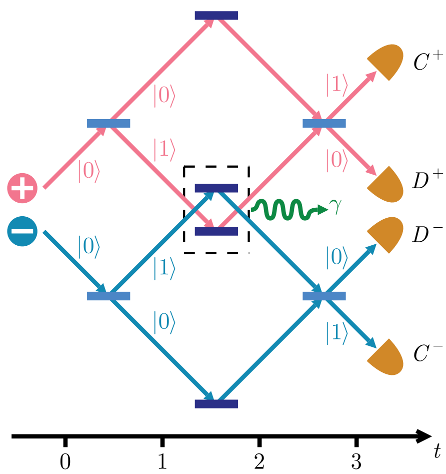

FIG. 2. (Color online) Setup of Hardy's paradox.

however. In particular,

$$f_2(q^+,q^-) \equiv \sum_{p^+,p^-} f_2(q^+,q^-,p^+,p^-) \tag{6.8}$$

$$= \begin{pmatrix} f_2(0,0) \\ f_2(0,1) \\ f_2(1,0) \\ f_2(1,1) \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \tag{6.9}$$

which correctly predicts the *a posteriori* position probability distribution if one measures the positions of the particles at that instant using strong measurements. Most importantly, $f_2(1,1) = 0$, and the probability that one measures both particles in the overlapping arms with strong measurements is zero. After the second pair of beam splitters, the quantum ket is

$$|\Psi\rangle_3 = \frac{1}{2\sqrt{3}}(-|0,0\rangle + |0,1\rangle + |1,0\rangle + 3|1,1\rangle), \tag{6.10}$$

and the Wigner distribution becomes

$$f_3(q^+,q^-,p^+,p^-) = \frac{1}{12}\begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 2 & 0 & -1 \\ 0 & 0 & 2 & -1 \\ 4 & 2 & 2 & 1 \end{pmatrix}, \tag{6.11}$$

$$f_3(q^+,q^-) = \frac{1}{12}\begin{pmatrix} 1 \\ 1 \\ 1 \\ 9 \end{pmatrix}, \tag{6.12}$$

which again correctly predicts the probability distribution of detection outcomes, conditioned upon the observation that annihilation did not occur.

Now let us perform retrodiction and calculate the retrodictive Wigner distribution conditioned upon the detection outcomes. Given that D^+ and D^- click, it can be shown that

$$g_2(q^+,q^-,p^+,p^-) = \frac{1}{4}\begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{6.13}$$

which estimates that the particles had definite momenta $(p^+,p^-) = (1,1)$ at time instant 2. Combining prediction and retrodiction, the smoothing quasiprobability distribution at time instant 2 becomes

$$h_2(q^+,q^-,p^+,p^-) \propto f_2(q^+,q^-,p^+,p^-)g_2(q^+,q^-,p^+,p^-)$$

$$= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{6.14}$$

$$h_2(q^+,q^-) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}. \tag{6.15}$$

Hence, given that the annihilation did not occur and D^+ and D^- click, the particles must both have been in the overlapping arms according to quantum smoothing. This result contradicts with the fact that annihilation did not occur and the particles could not have met, but is consistent with the classical reasoning that leads one to the same paradoxical conclusion. Mathematically, the paradox arises because the predictive estimation according to $f_2(q^+,q^-)$ contradicts the smoothing estimation according to $h_2(q^+,q^-)$, with the former ascertaining that the particles cannot both be in the overlapping arms, while the latter insisting the opposite.

To see why this cannot happen in classical estimation theory, assume for the time being that $f_2(q^+,q^-,p^+,p^-)$ is nonnegative and therefore a qualified joint probability distribution for (q^+,q^-,p^+,p^-). Then

$$f_2(1,1) = \sum_{p^+,p^-} f_2(1,1,p^+,p^-) = 0 \tag{6.16}$$

if and only if

$$f_2(1,1,p^+,p^-) = 0 \quad \text{for all} \quad (p^+,p^-). \tag{6.17}$$

If $f_2(1,1,p^+,p^-)$ is zero, the smoothing probability $h_2(1,1,p^+,p^-)$ must also be zero,

$$h_2(1,1,p^+,p^-) \propto f_2(1,1,p^+,p^-)g_2(1,1,p^+,p^-) = 0, \tag{6.18}$$

$$h_2(1,1) = \sum_{p^+,p^-} h_2(1,1,p^+,p^-) = 0, \tag{6.19}$$

and smoothing would also conclude that the particles could not have both been in the overlapping arms. In other words, in classical estimation, if the predictive theory estimates with certainty that the two particles cannot both be in the overlapping arms, then no amount of measurements afterwards can alter the certainty of this fact.

Quantum smoothing, on the other hand, contradicts with quantum prediction because some elements of $f_2(1,1,p^+,p^-)$ are negative. This way $f_2(1,1)$ can still be zero with nonzero $f_2(1,1,p^+,p^-)$ elements, and $h_2(1,1,p^-,p^+)$ and $h_2(1,1)$, conditioned upon the detection outcomes, can become nonzero. The negative elements of $f_2(q^+,q^-,p^+,p^-)$ thus cause prediction and smoothing to produce contradictory trajectories.

If the detection outcomes are different, say, C^+ and D^- click, then

$$g_2(q^+, q^-, p^+, p^-) = \frac{1}{4}\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \tag{6.20}$$

$$h_2(q^+, q^-, p^+, p^-) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}, \tag{6.21}$$

$$h_2(q^+, q^-) = \begin{pmatrix} 0 \\ 0 \\ 2 \\ -1 \end{pmatrix}, \tag{6.22}$$

and we have a negative "probability." Leaving aside the question of interpreting a negative probability [21], $h_2(q^+, q^-)$ still suggests that the most likely positions are $(q^+, q^-) = (1, 0)$, which are consistent with classical reasoning. Similarly, when C^+ and C^- click, the most likely (q^+, q^-) according to $h_2(q^-, q^+)$ is $(0, 0)$, which is again what one would expect from a classical argument. In this example at least, the most likely positions suggested by quantum smoothing therefore coincide with the ones obtained by qualitative classical reasoning, as depicted in Fig. 3.

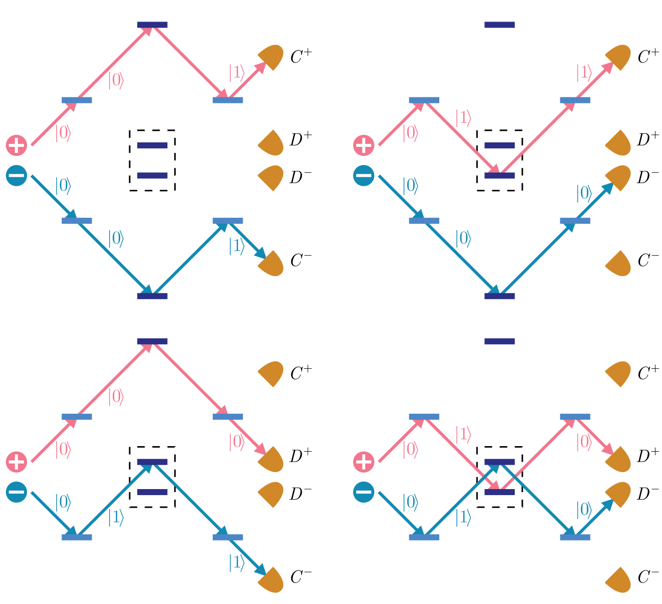

FIG. 3. (Color online) The most likely paths undertaken by the particles indicated by quantum smoothing given the detection outcomes, provided that annihilation did not occur. These paths coincide with those suggested by qualitative classical reasoning. When D^+ and D^- click, the estimated paths, as shown in the bottom-right figure, contradict with the fact that annihilation did not occur and the two particles could not have both been in the overlapping arms.

To summarize, the quantum phase-space smoothing approach is able to reproduce the salient features of Hardy's paradox and identify the negativity of $f_2(q^+, q^-, p^+, p^-)$ as the culprit that makes the classical phase-space picture and quantum theory incompatible. $f_2(q^+, q^-, p^+, p^-)$ should be measurable experimentally, as it is simply the expected value of the operator $\hat{W}(q^+, p^+)\hat{W}(q^-, p^-)/4$ for the quantum state $|\Psi\rangle_2$. In light of the paradox, one might be tempted to dismiss the quantum smoothing distribution as meaningless, but its usefulness for sensing applications and correspondence with classical reasoning suggests that it is still a valuable computational and conceptual tool for quantum information processing applications and offers an alternative view of the quantum reality.

VII. CONCLUSION

In conclusion, the time-symmetric smoothing theory is extended to account for discrete variables in classical systems, quantum systems, and observations. To illustrate the extended theory, atomic magnetometry and Hardy's paradox are studied using quantum phase-space smoothing. The generalized smoothing theory outlined in this paper is expected to be useful in future quantum sensing and information processing applications.

ACKNOWLEDGMENTS

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

APPENDIX: OBTAINING THE QUANTUM SMOOTHING DISTRIBUTION BY WEAK MEASUREMENTS

In the case of continuous variables, the quantum smoothing distribution can be obtained from the statistics of weak position and momentum measurements, conditioned upon past and future observations [3]. One may also apply a similar method to the discrete-variable case. Interestingly, the statistics of weak measurements naturally lead to a $2N \times 2N$ phase space.

Consider consecutive q and p measurements of a quantum system. Let the measurement operators be

$$\hat{M}(y_q) = \sqrt{\mathcal{C}_q}\sum_{q=0}^{N-1}\exp\left[\frac{\epsilon_q}{2}\cos\frac{2\pi}{N}(y_q - q)\right]|q\rangle\langle q|, \tag{A1}$$

$$\hat{M}(y_p) = \sqrt{\mathcal{C}_p}\sum_{p=0}^{N-1}\exp\left[\frac{\epsilon_p}{2}\cos\frac{2\pi}{N}(y_p - p)\right]|p\rangle\langle p|, \tag{A2}$$

where $\mathcal{C}_{q,p}$ are normalization constants and $\epsilon_{q,p}$ parametrize the measurement strengths and accuracies. The probability distribution of y_q and y_p, conditioned upon past and future observations, is

$$P(y_q, y_p) = \frac{\mathrm{tr}\,[\hat{g}\hat{M}(y_p)\hat{M}(y_q)\hat{f}\hat{M}^\dagger(y_q)\hat{M}^\dagger(y_p)]}{\mathrm{tr}(\hat{g}\hat{f})}$$

$$= \frac{\mathcal{C}_q\mathcal{C}_p}{N\mathrm{tr}(\hat{g}\hat{f})} \sum_{q,q',p,p'} \exp\left[\frac{2\pi i(p'q'-pq)}{N} + \frac{\epsilon_q}{2}\cos\frac{2\pi(y_q-q)}{N} + \frac{\epsilon_q}{2}\cos\frac{2\pi(y_q-q')}{N} + \frac{\epsilon_p}{2}\cos\frac{2\pi}{N}(y_p-p) + \frac{\epsilon_p}{2}\cos\frac{2\pi}{N}(y_p-p')\right] \times \langle p'|\hat{g}|p\rangle\langle q|\hat{f}|q'\rangle. \quad (A3)$$

Let

$$\bar{q} = \frac{q+q'}{2}, \quad u = \frac{q'-q}{2}, \quad \bar{p} = \frac{p+p'}{2}, \quad v = \frac{p'-p}{2}. \quad (A4)$$

Applying trigonometric identities, one obtains

$$P(y_q, y_p) = \frac{\mathcal{C}_q\mathcal{C}_p}{N\mathrm{tr}(\hat{g}\hat{f})} \sum_{q,q',p,p'} \exp\left[\frac{4\pi i(v\bar{q}+\bar{p}u)}{N} + \epsilon_q\cos\frac{2\pi(y_q-\bar{q})}{N} + \epsilon_p\cos\frac{2\pi(y_p-\bar{p})}{N} - 2\epsilon_q\cos\frac{2\pi(y_q-\bar{q})}{N}\sin^2\frac{\pi u}{N} - 2\epsilon_p\cos\frac{2\pi(y_p-\bar{p})}{N}\sin^2\frac{\pi v}{N}\right] \times\langle\bar{p}+v|\hat{g}|\bar{p}-v\rangle\langle\bar{q}-u|\hat{f}|\bar{q}+u\rangle. \quad (A5)$$

Utilizing the periodic nature of the above expression, one can change the sum in terms of (q, q') to a sum in terms of $(\bar{q}, u)$,

$$\sum_{q=0}^{N-1}\sum_{q'=0}^{N-1} \to \frac{1}{2}\sum_{\bar{q},u}, \quad (A6)$$

$$\bar{q} \in \left\{0, \frac{1}{2}, \ldots, N-\frac{1}{2}\right\}, \quad (A7)$$

$$u \in \left\{-\frac{N}{2}+\frac{1}{2}, \frac{N}{2}+1, \ldots, \frac{N}{2}\right\}, \quad (A8)$$

likewise for (p, p') and $(\bar{p}, v)$, and the matrix elements $\langle\bar{p}+v|\hat{f}|\bar{p}-v\rangle$ and $\langle\bar{q}-u|\hat{g}|\bar{q}+u\rangle$ are assumed to be zero whenever $\bar{p}+v$, $\bar{p}-v$, $\bar{q}-u$, or $\bar{q}+u$ is not an integer. Thus,

$$P(y_q, y_p) = \frac{N\mathcal{C}_q\mathcal{C}_p}{\mathrm{tr}(\hat{g}\hat{f})} \sum_{\bar{q},\bar{p}} \exp\left[\epsilon_q\cos\frac{2\pi(y_q-\bar{q})}{N} + \epsilon_p\cos\frac{2\pi(y_p-\bar{p})}{N}\right]\tilde{g}(\bar{q},\bar{p})\tilde{f}(\bar{q},\bar{p}), \quad (A9)$$

where

$$\tilde{f}(\bar{q},\bar{p}) = \frac{1}{2N}\sum_v \exp\left[-2\epsilon_q\cos\frac{2\pi(y_q-\bar{q})}{N}\sin^2\frac{\pi u}{N}\right] \times\exp\left(\frac{4\pi i\bar{p}u}{N}\right)\langle\bar{q}-u|\hat{f}|\bar{q}+u\rangle, \quad (A10)$$

$$\tilde{g}(\bar{q},\bar{p}) = \frac{1}{2N}\sum_u \exp\left[-2\epsilon_p\cos\frac{2\pi(y_p-\bar{p})}{N}\sin^2\frac{\pi v}{N}\right] \times\exp\left(\frac{4\pi iv\bar{q}}{N}\right)\langle\bar{p}+v|\hat{g}|\bar{p}-v\rangle. \quad (A11)$$

In the limit of infinitesimally weak measurements and $\epsilon_{q,p} \ll 1$,

$$\tilde{f}(\bar{q},\bar{p}) \approx \frac{1}{2N}\sum_v \exp\left(\frac{4\pi i\bar{p}u}{N}\right)\langle\bar{q}-u|\hat{f}|\bar{q}+u\rangle, \quad (A12)$$

$$\tilde{g}(\bar{q},\bar{p}) \approx \frac{1}{2N}\sum_u \exp\left(\frac{4\pi iv\bar{q}}{N}\right)\langle\bar{p}+v|\hat{g}|\bar{p}-v\rangle, \quad (A13)$$

which are precisely the discrete Wigner distributions in the $2N \times 2N$ phase space. Equation (A9) becomes

$$P(y_q, y_p) = \mathcal{C}_q\mathcal{C}_p\sum_{\bar{q},\bar{p}}\exp\left[\epsilon_q\cos\frac{2\pi(y_q-\bar{q})}{N} + \epsilon_p\cos\frac{2\pi(y_p-\bar{p})}{N}\right]\tilde{h}(\bar{q},\bar{p}), \quad (A14)$$

and can be regarded, from the perspective of classical probability theory, as the probability distribution for noisy q and p measurements, when the system has a phase-space distribution given by the quantum smoothing distribution $\tilde{h}(\bar{q},\bar{p})$. $\tilde{h}(\bar{q},\bar{p})$ can therefore be obtained in an experiment with small $\epsilon_{q,p}$ by measuring $P(y_q, y_p)$ for the same $\hat{g}$ and $\hat{f}$ and deconvolving Eq. (A14).

[1] M. Tsang, J. H. Shapiro, and S. Lloyd, Phys. Rev. A **78**, 053820 (2008); **79**, 053843 (2009).
[2] M. Tsang, Phys. Rev. Lett. **102**, 250403 (2009).
[3] M. Tsang, Phys. Rev. A **80**, 033840 (2009).
[4] Y. Aharonov, P. G. Bergmann, and J. L. Lebowitz, Phys. Rev. **134**, B1410 (1964); Y. Aharonov, D. Z. Albert, and L. Vaidman, Phys. Rev. Lett. **60**, 1351 (1988); Y. Aharonov and L. Vaidman, Phys. Rev. A **41**, 11 (1990); J. Phys. A **24**, 2315 (1991).
[5] 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).
[6] V. P. Belavkin, Radiotech. Elektron. **25**, 1445 (1980); in *Information Complexity and Control in Quantum Physics*, edited by A. Blaquière, S. Diner, and G. Lochak (Springer, Vienna, 1987), p. 311; in *Stochastic Methods in Mathematics and Physics*, edited by R. Gielerak and W. Karwowski (World Scientific, Singapore, 1989), p. 310; in *Modeling and Control of Systems in Engineering, Quantum Mechanics, Economics, and Biosciences*, edited by A. Blaquière (Springer, Berlin, 1989), p. 245; J. Multivariate Anal. **42**, 171 (1992); Rep. Math. Phys. **43**, 405 (1999).

[7] A. Barchielli, L. Lanz, and G. M. Prosperi, Nuovo Cimento B **72**, 79 (1982); Found. Phys. **13**, 779 (1983); H. Carmichael, *An Open Systems Approach to Quantum Optics* (Springer-Verlag, Berlin, 1993).
[8] C. W. Gardiner and P. Zoller, *Quantum Noise* (Springer-Verlag, Berlin, 2000).
[9] H. Mabuchi, Phys. Rev. A **58**, 123 (1998); F. Verstraete, A. C. Doherty, and H. Mabuchi, *ibid.* **64**, 032111 (2001).
[10] D. W. Berry and H. M. Wiseman, Phys. Rev. A **65**, 043803 (2002); **73**, 063824 (2006).
[11] L. K. Thomsen, S. Mancini, and H. M. Wiseman, Phys. Rev. A **65**, 061801(R) (2002); J. Phys. B: At. Mol. Opt. Phys. **35**, 4937 (2002).
[12] J. M. Geremia, J. K. Stockton, A. C. Doherty, and H. Mabuchi, Phys. Rev. Lett. **91**, 250801 (2003); B. A. Chase and J. M. Geremia, Phys. Rev. A **79**, 022314 (2009); B. A. Chase, B. Q. Baragiola, H. L. Partner, B. D. Black, and J. M. Geremia, *ibid.* **79**, 062107 (2009).
[13] J. K. Stockton, J. M. Geremia, A. C. Doherty, and H. Mabuchi, Phys. Rev. A **69**, 032109 (2004).
[14] L. Bouten, J. Stockton, G. Sarma, and H. Mabuchi, Phys. Rev. A **75**, 052111 (2007).
[15] T. A. Wheatley, D. W. Berry, H. Yonezawa, D. Nakane, H. Arao, D. T. Pope, T. C. Ralph, H. M. Wiseman, A. Furusawa, and E. H. Huntington, e-print arXiv:0912.1162.
[16] A. Kuzmich, L. Mandel, and N. P. Bigelow, Phys. Rev. Lett. **85**, 1594 (2000).
[17] D. Budker, W. Gawlik, D. F. Kimball, S. M. Rochester, V. V. Yashchuk, and A. Weis, Rev. Mod. Phys. **74**, 1153 (2002).
[18] V. Petersen and K. Mølmer, Phys. Rev. A **74**, 043802 (2006).
[19] L. Hardy, Phys. Rev. Lett. **68**, 2981 (1992).
[20] Y. Aharonov *et al.*, Phys. Lett. **A301**, 130 (2002); J. S. Lundeen and A. M. Steinberg, Phys. Rev. Lett. **102**, 020404 (2009); K. Yokota, T. Yamamoto, M. Koashi, and N. Imoto, New J. Phys. **11**, 033011 (2009).
[21] R. Feynman, in *Quantum Implications: Essays in Honour of David Bohm*, edited by B. Hiley and D. Peat (Routledge, London, 1987), p. 235.
[22] D. L. Snyder, IEEE Trans. Inf. Theory **18**, 91 (1972).
[23] D. L. Snyder, *Random Point Processes* (Wiley, New York, 1975).
[24] E. Pardoux, in *Stochastic Systems: The Mathematics of Filtering and Identification and Applications*, edited by M. Hazewinkel and J. C. Willems (Reidel, Dordrecht, 1981), p. 529; see also B. D. O. Anderson and I. B. Rhodes, Stochastics **9**, 139 (1983).
[25] C. W. Gardiner, *Handbook of Stochastic Methods* (Springer-Verlag, Berlin, 1985).
[26] I. V. Aleksandrov, Z. Naturforsch. A **36**, 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).
[27] A. S. Holevo, *Statistical Structure of Quantum Theory* (Springer-Verlag, Berlin, 2001); L. Bouten, R. van Handel, and M. R. James, SIAM J. Control Optim. **46**, 2199 (2007).
[28] G. T. Foster, L. A. Orozco, H. M. Castro-Beltran, and H. J. Carmichael, Phys. Rev. Lett. **85**, 3149 (2000); H. Nha and H. J. Carmichael, *ibid.* **93**, 020401 (2004); R. García-Patrón, J. Fiurášek, N. J. Cerf, J. Wenger, R. Tualle-Brouri, and Ph. Grangier, *ibid.* **93**, 130409 (2004).
[29] W. K. Wootters, Ann. Phys. (NY) **176**, 1 (1987).
[30] J. H. Hannay and M. V. Berry, Physica D **1**, 267 (1980); see also U. Leonhardt, Phys. Rev. A **53**, 2998 (1996); P. Bianucci, C. Miquel, J. P. Paz, and M. Saraceno, Phys. Lett. **A297**, 353 (2002).
[31] J. A. Vaccaro and D. T. Pegg, Phys. Rev. A **41**, 5156 (1990).
[32] A. Lukš and V. Peřinová, Phys. Scr. **T48**, 94 (1993).
[33] L. Cohen and M. Scully, Found. Phys. **16**, 295 (1986); K. S. Gibbons, M. J. Hoffman, and W. K. Wootters, Phys. Rev. A **70**, 062101 (2004); G. S. Agarwal, *ibid.* **24**, 2889 (1981); J. P. Dowling, G. S. Agarwal, and W. P. Schleich, *ibid.* **49**, 4101 (1994).
[34] D. Q. Mayne, Automatica **4**, 73 (1966); D. C. Fraser and J. E. Potter, IEEE Trans. Autom. Control **14**, 387 (1969).