Bayesian parameter inference from continuously monitored quantum systems

Søren Gammelmark and Klaus Mølmer

Lundbeck Foundation Theoretical Center for Quantum System Research, Department of Physics and Astronomy, University of Aarhus, DK 8000 Aarhus C, Denmark

(Received 27 December 2012; published 25 March 2013)

We review the introduction of likelihood functions and Fisher information in classical estimation theory, and we show how they can be defined in a very similar manner within quantum measurement theory. We show that the stochastic master equations describing the dynamics of a quantum system subject to a definite set of measurements provides likelihood functions for unknown parameters in the system dynamics, and we show that the estimation error, given by the Fisher information, can be identified by stochastic master equation simulations. For large parameter spaces we describe and illustrate the efficient use of Markov chain Monte Carlo sampling of the likelihood function.

DOI: 10.1103/PhysRevA.87.032115 PACS number(s): 03.65.Wj, 03.65.Yz, 02.50.Tt, 42.50.Lc

I. INTRODUCTION

Sensors and measurement devices are affected by the presence or strength of physical effects that influence their dynamics in a detectable way. A proper statistical treatment of measurement data is important when inferring results from complex experiments. With the growing use of quantum systems for high-precision measurements, a whole research domain of quantum metrology has emerged. Limitations to measurement precision from quantum mechanical uncertainties have been investigated and protocols to use measurements to optimally distinguish differently prepared quantum states have been developed; cf. [1–6].

The continuous observation of a quantum system involves leakage of information via coupling of the system to a suitable meter, and an archetypal example is that of measurements of photons emitted from a quantum light source. Laser spectroscopy thus involves the excitation of a quantum system, and detection of the fluorescence signal as function of laser frequency permits a fit, e.g., to a Lorentzian distribution and thus yields information about the resonance frequency and linewidth. The resonance curve, however, represents only a part of the acquired data, as it omits details concerning the temporal dynamics and noise properties of the detection signal. With the emergence of stochastic Schrödinger and master equations, which determine the quantum state conditioned on the full, noisy detection signals, it has been a natural next step to develop strategies to extract information from continuously probed systems. Immediate applications then concern sensing of the magnitude of perturbations acting on the system, such as the magnetic field probed in atomic magnetometers [7,8], and near-field effects, e.g., from nuclear spins, probed by a single nitrogen vacancy center in diamond [9,10]. Theoretical strategies have been proposed to continuously update parameter estimates based on the acquired data in a Bayesian manner on equal footing with the conditioned quantum state of the probe system [11,12]. The latter method is particularly useful, if the system can be approximated by Gaussian states [13,14].

The purpose of the present paper is to provide a formal link between some of the central ideas in classical estimation theory and stochastic master equations and to identify efficient and systematic means to estimate unknown parameters from quantum measurement records. We, in particular, discuss and demonstrate methods applicable in cases where the parameter space is too large to permit a recursive Bayesian update procedure. The methods are general, but for concreteness, we consider light-emitting quantum systems, and we present explicit analyses and results for direct photon detection and for homodyne detection of the emitted radiation.

In photon counting, the measurement signal is a discrete process N_t, characterized by click events at specific times, and the density matrix ρ_t of the emitter conditioned on the detection signal until time t satisfies the nonlinear filter equation

$$d\rho_t = \left[-i\,[H,\rho_t] - \frac{1}{2}\{c^\dagger c,\rho_t\} + \mathrm{Tr}(c^\dagger c\rho_t)\rho_t\right]dt + \left[\frac{c\rho_t c^\dagger}{\mathrm{Tr}(c^\dagger c\rho_t)} - \rho_t\right]dN_t, \tag{1}$$

where the differential measurement result dN_t is a Poisson increment, which is either 0 (no click event) or 1 (detector click event). For the special case of a two-level atom with upper (lower) states $|e(g)\rangle$ and with upper state lifetime $1/\gamma$, the conditional expectation is $\mathbb{E}[dN_t|N_t] = \mathrm{Tr}(c^\dagger c\rho_t)dt$, where $c = \sqrt{\gamma}\,|g\rangle\langle e|$. The time evolution of ρ_t is therefore piecewise continuous (when $dN_t = 0$), but interrupted by jumps $\rho_t \mapsto c\rho_t c^\dagger/\mathrm{Tr}(c^\dagger c\rho_t)$ at discrete times (when $dN_t = 1$).

It is also possible to perform field amplitude measurements by homodyne and heterodyne detection. The measurement signal Y_t is then a continuous function of time, and, e.g., when homodyne detection is performed on the fluorescence emitted by the decaying two-level atom, the conditioned density matrix satisfies an Itô stochastic differential equation,

$$d\rho_t = (-i[H,\rho] - \{c^\dagger c,\rho\}/2 + c\rho c^\dagger)dt + \{\mathcal{M}(\rho_t) - \mathrm{Tr}[\mathcal{M}(\rho_t)]\rho_t\}\{dY_t - \mathrm{Tr}[\mathcal{M}(\rho_t)]dt\}, \tag{2}$$

where $\mathcal{M}(\rho) = c\rho + \rho c^\dagger$. The differential measurement result dY_t satisfies $dY_t = \mathrm{Tr}[\mathcal{M}(\rho_t)]dt + dW_t$, where dW_t is a Wiener increment with zero mean and variance dt.

In a generic experiment, all terms in the stochastic master equation can be parametrized by a vector of classical variables $\theta \in \mathbb{R}^n$, such as laser-atom detunings, which may, in turn, depend on the unknown values of externally applied fields, decay rates, temperature, etc. In order to solve Eqs. (1) and

(2), candidate values for these parameters need to be specified, and the goal of parameter estimation by continuous quantum measurements is to identify the best candidate values for the parameters θ, given the actual measurement record (N_t or Y_t).

In Sec. II, we review general parameter estimation concepts relevant to this paper: Bayesian inference, likelihood functions, and the Fisher information. In Sec. III, we show how likelihood functions and the Fisher information can be efficiently obtained from the solution of the stochastic master equation of continuously monitored quantum systems. In Sec. IV, we introduce the Markov chain Monte Carlo method for efficient sampling of the likelihood function in large search spaces, and we give numerical examples which illustrate the application and the results of our methods. In Sec. V, we present a conclusion and outlook.

II. BAYESIAN INFERENCE

Our theory of estimation is based on Bayes' rule,

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}, \tag{3}$$

where $P(\theta|D)$ is the probability density of the parameters θ, given the observed data D. Informally, this object contains all the information about the system parameters θ contained in the observed data D. From this distribution we can calculate any estimate of interest, including the mean value, the mode, and quantiles. An important advantage of calculating the full probability density $P(\theta|D)$ is that it explicitly contains information about the uncertainty of the estimates.

Bayes' rule relates the conditional probability density to the probability of observing the data given the parameters $P(D|\theta)$ and the existing prior information about the parameters $P(\theta)$. The difficulty in using Eq. (3) stems from the denominator being a weighted integral over all possible parameter values $P(D) = \int d\theta P(D|\theta)P(\theta)$. This integral is high-dimensional when several parameters are estimated, and the integrand can vary many orders of magnitude.

To determine $P(\theta|D)$ in practice, we therefore need a method for calculating $P(D|\theta)$ and a numerically efficient method of calculating $P(D)$.

A. Likelihood functions

Since the data D is a measurement record, i.e., a function of time, its probability density, or *likelihood*, $P(D|\theta)$ is difficult to define. Apart from its use in the Bayesian update rule (3), it is common to maximize the likelihood with respect to the parameters θ and thus to estimate the true value of the parameter by the value for which the likelihood for generating the data is highest.

Instead of maximizing $P(D|\theta)$ with respect to θ, one may maximize the value of any strictly increasing function f of $P(D|\theta)$. The logarithm is commonly used, and the resulting function is then denoted the *log-likelihood* function.

It is also possible to divide $P(D|\theta)$ with any strictly positive function $P_0(D)$, without changing the location of the maximum with respect to θ. Thus, any function $f(P(D|\theta)/P_0(D))$ can be used to determine the maximum likelihood estimate of θ. We use the term *likelihood function* for any function $L(D|\theta) = P(D|\theta)/P_0(D)$ and logarithmic likelihood for $l(D|\theta) = \ln L(D|\theta)$. The likelihood function $L(D|\theta)$ can to a large extent be chosen to have a convenient form.

Since $P(D) = \int d\theta P(D|\theta)P(\theta) = P_0(D)\int d\theta L(D|\theta)P(\theta)$ we can rewrite Eq. (3) as

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P_0(D)\int d\theta L(D|\theta)P(\theta)} = \frac{L(D|\theta)P(\theta)}{\int d\theta L(D|\theta)P(\theta)}. \tag{4}$$

In the following we calculate the likelihood associated with continuously monitored light-emitting quantum systems, where the function $P_0(D)$ is the probability density for either a Poisson or a Wiener process.

B. Fisher information

A reasonable question to ask is, how accurate is the Bayesian estimate on average, and what is the fundamental limit on how accurate it is possible to estimate θ?

The answer to this question is given by the Fisher information matrix. The Fisher information matrix is defined in terms of the probability density for the data given some parameter θ and $P(D|\theta)$ as

$$I = \mathbb{E}\left[\left(\frac{\partial \ln P(D|\theta)}{\partial\theta}\right)^2\right], \tag{5}$$

where the expectation is over all possible realizations of the data D. The Cramér-Rao bound [15] states, that any estimator for θ, $\hat{\theta}(D)$ has a variance larger than $1/I(\theta_0)$, where θ_0 is the true value of θ.

If one uses a uniform prior, the Fisher information of $P(D|\theta)$ is the reciprocal of the width of $P(\theta|D)$ averaged over the possible measurement records. If a nonuniform prior is included, the reciprocal width of $P(\theta|D)$ is then, qualitatively, the sum of the Fisher information for θ and the reciprocal width of the prior.

As described in Sec. II A, for Bayesian inference we can use any likelihood function in place of the conditional probability $P(D|\theta)$. The same result holds for the Fisher information. That is, in Eq. (5) we can use any likelihood function $L(D|\theta)$ instead of the probability density.

For multiple variables, the Fisher information is

$$I_{ij} = \mathbb{E}\left[\frac{\partial \ln L(D|\theta)}{\partial\theta_i}\frac{\partial \ln L(D|\theta)}{\partial\theta_j}\right], \tag{6}$$

$$= \mathbb{E}\left[L(D|\theta)^{-2}\frac{\partial L(D|\theta)}{\partial\theta_i}\frac{\partial L(D|\theta)}{\partial\theta_j}\right], \tag{7}$$

where $L(D|\theta)$ is a likelihood function for observing D given the parameters θ. The Cramér-Rao bound now states (for any unbiased estimator $\hat{\theta}$) that $\mathbb{E}[(\hat{\theta}_i - \theta_i^0)(\hat{\theta}_j - \theta_j^0)] \geqslant [I(\theta)^{-1}]_{ij}$.

III. CONTINUOUSLY MONITORED QUANTUM SYSTEMS

The jump and diffusion quantum filter equations (1) and (2) are special cases of the general transformation of open quantum system density matrices subject to the random backaction of measurements. If, at a given instant of time, a measurement occurs with outcome $m \in M$, there is an effect

operator $\Omega(m)$ associated with each outcome, so that the state, conditioned on the outcome m, reads

$$\rho|m = \frac{\Omega(m)\rho\Omega^\dagger(m)}{\mathrm{Tr}[\Omega^\dagger(m)\Omega(m)\rho]}. \tag{8}$$

The probability (density) for observing the result m is

$$p(m) = \mathrm{Tr}[\Omega^\dagger(m)\Omega(m)\rho], \tag{9}$$

and the effect operators obey the relation

$$\int dm\Omega^\dagger(m)\Omega(m) = \mathbb{1}. \tag{10}$$

The normalization factors in Eq. (8) are exactly the probabilities (9) to obtain the corresponding measurement outcome, and a similar probabilistic interpretation holds for the nonlinear terms including the coefficients $\mathrm{Tr}(c^\dagger c\rho_t)\rho_t$ and $\mathrm{Tr}[\mathcal{M}(\rho_t)]\rho_t$, in Eqs. (1) and (2). This implies that if the stochastic density matrix equation is solved without incorporating the renormalization factors, the decreasing trace of ρ with time yields the likelihood for the actual detection record to occur.

In the quantum jump master equation, the jump probability, and hence the decrease in density matrix norm associated with a single jump is proportional to the duration of the infinitesimal time step dt chosen for the simulation. This causes an undesired and inconvenient dependence of the likelihood function on dt and on the number of click events, which we can, however, eliminate by a simple extension of the theory [2] similar to [16].

We introduce an arbitrary positive function $p_0(m)$ and rescale the effect operators $\Omega(m) \to \Omega(m)/\sqrt{p_0(m)}$ so that they now obey the modified normalization condition,

$$\int_M dmp_0(m)\Omega^\dagger(m)\Omega(m) = \mathbb{1}. \tag{11}$$

Equation (8) still holds, but the probability distribution for the different outcomes factors,

$$p(m) = p_0(m)\,\mathrm{Tr}[\Omega^\dagger(m)\Omega(m)\rho], \tag{12}$$

and we have the freedom to choose the un-normalized conditional states,

$$\tilde{\rho}|m = \Omega(m)\rho\Omega^\dagger(m), \tag{13}$$

whose trace depends now explicitly on the chosen function $p_0(m)$.

The expectation value, denoted by $\mathbb{E}$, of any function $f(m)$ is given by

$$\begin{aligned}\mathbb{E}[f(m)] &= \int_M dmp(m)f(m)\\ &= \int_M dmp_0(m)\,\mathrm{Tr}(\tilde{\rho}|m)f(m) \equiv \mathbb{E}_0[\mathrm{Tr}(\tilde{\rho}|m)f(m)],\end{aligned}$$

where $\mathbb{E}_0$ is to be understood as the expectation with respect to the reference probability p_0. In the following, we suppress the dependence on the measurement outcomes and simply write $\tilde{\rho}$ rather than $\tilde{\rho}|m$ for the conditioned density matrix.

The trace of the conditioned state is renormalized by a factor that depends on the specific detection record and which does not change its relative dependence on different values of the unknown parameters θ. It thus still serves as a likelihood function for the Bayesian determination of their values. Our scaling with the function $p_0(m)$ in Eq. (12) is indeed equivalent to the scaling allowed in the definition of the likelihood function, $L(D|\theta) = P(D|\theta)/P_0(D)$, in Sec. II A. The "ostensible probability" p_0 [2] provides a reference measure $p_0(m)dm$ on the set of measurement outcomes, and for our application it serves as a convenient unit for the effect operators $\Omega(m)$.

We emphasize that $p_0(m)$ and $P_0(D)$ are only requested to be positive functions of the measurement outcomes, independent of the parameters that we want to estimate. In our calculation of the likelihood function, they thus merely serve to provide a convenient normalization.

The above theory can easily be extended to include continuous-time measurements. For a measurement repeated N times, it is natural to consider a general reference probability distribution on M^N, such that the probability of the measurement record factors as in Eq. (12):

$$\begin{aligned}p(m_1,\ldots,m_N) = p_0^{(N)}(m_1,\ldots,m_N)\\ \mathrm{Tr}[\Omega(m_N)\cdots\Omega(m_1)\rho\Omega^\dagger(m_i)\cdots\Omega^\dagger(m_N)].\end{aligned} \tag{14}$$

By taking the limit as $N \to \infty$ for a fixed time span, continuous-time measurements can be described. Convenient reference probabilities $p_0^{(N)}(m_1,\ldots,m_N) = p_0(m_1)\cdots p_0(m_N)$ for the jump-type and diffusionlike measurements are given below.

A. Jump-type equation

For the jump-type measurements there are for each small time interval dt two possible detector outcomes, $dN_t = 0$ and $dN_t = 1$. We use our freedom to choose p_0 as the probability for a Poisson process with rate λ, i.e., $p_0(dN_t = 1) = \lambda dt$ and $p_0(dN_t = 0) = 1 - \lambda dt$, and the correspondingly normalized measurement effect operators,

$$\begin{aligned}\Omega_0 &= \mathbb{1} - iHdt - \frac{1}{2}(c^\dagger c - \lambda)dt,\\ \Omega_1 &= \frac{c}{\sqrt{\lambda}}.\end{aligned} \tag{15}$$

We recall that the value of λ is chosen arbitrarily and does not represent any prior assumptions on the unknown physical variables or the statistics. As we demonstrate explicitly below, apart from normalization the likelihood function is completely independent of its value. The probability for a detector click is $p_0(dN_t = 1)\,\mathrm{Tr}(\Omega_1^\dagger\Omega_1\rho_t) = \mathrm{Tr}(c^\dagger c\rho_t)dt$, as expected from a rate process, and the expected number of events is $\mathbb{E}[dN_t|N_t] = \mathrm{Tr}(c^\dagger c\rho_t)dt$, while the reference expected value is $\mathbb{E}_0[dN_t|N_t] = \lambda dt$.

The un-normalized conditional quantum state can be expressed as

$$d\tilde{\rho}_t = \left[-i\,[H,\tilde{\rho}_t] - \frac{1}{2}\{c^\dagger c,\tilde{\rho}_t\} + \lambda\tilde{\rho}_t\right]dt + dN_t\left[\frac{c\tilde{\rho}_t c^\dagger}{\lambda} - \tilde{\rho}_t\right], \tag{16}$$

while explicit normalization leads to Eq. (1).

The dynamics of $\tilde{\rho}_t$ is governed by the Hamiltonian H and the operator c, which, in turn, depends on the parameters

θ. The likelihood of a specific sequence of detection events at times $t_1, \ldots, t_N < t$ is simply $L(t_1, \ldots, t_N|\theta) = \mathrm{Tr}(\tilde{\rho}_t)$, and Eq. (16) thus provides a differential equation for the likelihood function $L_t = \mathrm{Tr}(\tilde{\rho}_t)$,

$$dL_t = [\lambda L_t - \mathrm{Tr}(c^\dagger c \tilde{\rho}_t)]dt + dN_t \left[\frac{\mathrm{Tr}(c^\dagger c \tilde{\rho}_t)}{\lambda} - L_t \right], \quad (17)$$

where we have suppressed L_t's dependence on θ.

The solutions ρ_t of Eq. (1) and $\tilde{\rho}_t$ of Eq. (16) obey $\tilde{\rho}_t = \mathrm{Tr}(\tilde{\rho}_t)\rho_t = L_t \rho_t$, which can be inserted in (17) to yield

$$dL_t = [\lambda - \mathrm{Tr}(c^\dagger c \rho_t)]L_t dt + dN_t[\lambda^{-1}\,\mathrm{Tr}(c^\dagger c \rho_t) - 1]L_t. \quad (18)$$

This shows that even though the likelihood is formally defined by the trace of the un-normalized conditioned density matrix, L_t can be calculated from the normalized state ρ_t satisfying Eq. (1).

For numerical purposes it is convenient to work with $l_t = \ln L_t$ which satisfies

$$dl_t = [\lambda - \mathrm{Tr}(c^\dagger c \rho_t)]dt + dN_t \ln[\mathrm{Tr}(c^\dagger c \rho_t)/\lambda]. \quad (19)$$

Equation (19) may leave the impression that the log-likelihood function depends on the parameter λ due its different appearance in the no-jump and the jump part of the stochastic equation. The difference in the log likelihood for two different parameters θ_1 and θ_2 with any arbitrary detector signal N_t, however, satisfies $d(l_t^{\theta_1} - l_t^{\theta_2}) = [\mathrm{Tr}(c^\dagger c \rho_t)^{\theta_1} - \mathrm{Tr}(c^\dagger c \rho_t)^{\theta_2}]dt + dN_t \ln[\mathrm{Tr}(c^\dagger c \rho_t)^{\theta_1} / \mathrm{Tr}(c^\dagger c \rho_t)^{\theta_2}]$, which is manifestly independent of λ. Thus, also the ratio $L_t^{\theta_1}/L_t^{\theta_2} = \exp(l_t^{\theta_1} - l_t^{\theta_2})$ is independent of λ.

B. Diffusion equation

For diffusion-type measurements, describing, e.g., homodyne detection of light, the set of outcomes in a small time interval dt is the real numbers. We here use the probability of a Wiener increment dW_t, i.e., a normal distribution with mean zero and variance dt as our reference probability p_0^W. The effect of observing a result dY_t is

$$\Omega(dY_t) = \mathbb{1} - iHdt - \tfrac{1}{2}c^\dagger c dt + c dY_t, \quad (20)$$

and the probability for observing a given value dY_t is

$$p(dY_t) = p_0^W(dY_t)\{1 + \mathrm{Tr}[(c + c^\dagger)\rho_t]dY_t\}. \quad (21)$$

We can calculate $\mathbb{E}[dY_t|Y_t] = \mathbb{E}_0(\{1 + \mathrm{Tr}[(c + c^\dagger)\rho_t]dY_t\}dY_t|Y_t) = \mathrm{Tr}[(c + c^\dagger)\rho_t]dt$ and $\mathbb{E}[dY_t^2|Y_t] = \mathbb{E}_0(\{1 + \mathrm{Tr}[(c + c^\dagger)\rho_t]dY_t\}dY_t^2|Y_t) = dt$, which implies

$$dY_t = \mathrm{Tr}[(c + c^\dagger)\rho]dt + dW_t, \quad (22)$$

where dW_t is a Wiener increment with respect to the full probability distribution p (while dY_t is a Wiener increment with respect to p_0).

The un-normalized stochastic differential equation becomes

$$d\tilde{\rho}_t = \left[-i\,[H, \tilde{\rho}_t] - \tfrac{1}{2}\{c^\dagger c, \tilde{\rho}_t\} + c\tilde{\rho}_t c^\dagger\right]dt + (c\tilde{\rho}_t + \tilde{\rho}_t c^\dagger)dY_t, \quad (23)$$

and it leads to the likelihood $L_t = \mathrm{Tr}(\tilde{\rho}_t)$ satisfying

$$dL_t = \mathrm{Tr}[\mathcal{M}(\tilde{\rho}_t)]dY_t. \quad (24)$$

As above, we can also express L_t in terms of the normalized solution to Eq. (2),

$$dL_t = \mathrm{Tr}[\mathcal{M}(\rho_t)]L_t dY_t, \quad (25)$$

and the log likelihood $l_t = \ln L_t$ satisfies

$$dl_t = \mathrm{Tr}[\mathcal{M}(\rho_t)]\{dY_t - \mathrm{Tr}[\mathcal{M}(\rho_t)]dt\}. \quad (26)$$

C. Fisher information

Using $\mathrm{Tr}(\tilde{\rho}_t)$ as our likelihood function, we can apply Eq. (7) to calculate the Cramér-Rao bound for estimating the unknown parameters in the system dynamics under both types of measurements. Define the matrices

$$\rho_t^i = \frac{1}{\mathrm{Tr}(\tilde{\rho}_t)}\partial_i \tilde{\rho}_t, \quad (27)$$

where the derivative is with respect to the ith component of the vector of parameters θ. The expectation value of $\mathrm{Tr}(\rho_t^i)\,\mathrm{Tr}(\rho_t^j)$ with respect to the probability distribution p (i.e., the actual probability for generating a trajectory) will then be the ij component of the Fisher information matrix for the continuously monitored quantum system. We can therefore evaluate the Fisher information matrix numerically by simulating the stochastic master equation a large number of times and determine the expectation value Eq. (7).

In practice, for the jump-type measurement, this requires the solution of Eq. (1) together with a simultaneous evaluation of the matrices ρ_t^i, which can, in turn, be determined from the inhomogeneous jump-type master equation

$$\begin{aligned} d\rho_t^i = &\left[-i\left[H, \rho_t^i\right] - \tfrac{1}{2}\left\{c^\dagger c, \rho_t^i\right\} + \mathrm{Tr}(c^\dagger c \rho_t^i)\rho_t^i\right]dt \\ &+ \left[-i\left[\partial_i H, \rho_t\right] - \tfrac{1}{2}\left\{\partial_i(c^\dagger c), \rho_t^i\right\}\right]dt \\ &+ dN_t[c\rho_t^i c^\dagger + (\partial_i c)\rho_t c^\dagger + c\rho_t(\partial_i c^\dagger) - \rho_t^i], \end{aligned} \quad (28)$$

where the stochastic term dN_t takes the same value as in Eq. (1), and where the derivative of the Hamiltonian and damping terms with respect to θ_i are assumed known.

Similarly, for the diffusion-type measurement

$$\begin{aligned} d\rho_t^i = &\left[-i\left[H, \rho_t^i\right] - \left\{c^\dagger c, \rho_t^i\right\}/2 + c\rho_t^i c^\dagger\right]dt \\ &+ [-i[\partial_i H, \rho_t] - \{\partial_i(c^\dagger c), \rho_t\}/2 + (\partial_i c)\rho_t c^\dagger \\ &+ c\rho_t(\partial_i c^\dagger)]dt\{\mathcal{M}(\rho_t^i) + (\partial_i \mathcal{M})(\rho_t) \\ &- \mathrm{Tr}[\mathcal{M}(\rho_t)]\rho_t^i\}\{dY_t - \mathrm{Tr}[\mathcal{M}(\rho_t)]dt\}, \end{aligned} \quad (29)$$

where the Wiener increment $dY_t - \mathrm{Tr}[\mathcal{M}(\rho_t)]dt = dW_t$ takes the same value as in Eq. (2).

The Fisher information provides an average quantifier of the asymptotic uncertainty in the estimation problem. With Eqs. (1) and (28) and Eqs. (2) and (29) we have shown how the Fisher information can be calculated by simulating many independent sequences of the stochastic master equation for the two different types of measurement. These simulations have to be carried out for the candidate values of the parameters to yield the precision expected for an estimate based on a typical experimental run. As illustrated by comparison of other such precision measures in [11], different measurement schemes have different resolving power, and in future work, we plan to address these differences in more detail, e.g., by

comparing the Fisher information derived for the jump-type and for different diffusion-type measurements.

We also note that if the field or meter degrees of freedom could be left unmeasured, the full entangled density matrix of the quantum system and the quantized radiation field would depend on the unknown parameters. Thus, the general quantum Cramér-Rao bound derived by Braunstein and Caves [1] to determine a parameter, encoded in a quantum state, yields the ultimate accuracy with which the parameters in our state dynamics can be inferred using any type of measurements. Identifying that accuracy and investigating how closely it is approached by quantum jump and quantum diffusion measurements of the emitted light presents an interesting challenge for further studies.

IV. NUMERICAL INVESTIGATION

In this section we illustrate the theory outlined in the previous sections with a few characteristic examples. One approach for investigating $P(\theta|D)$ is to compute the likelihood function $L(D|\theta)$ on a grid. Using such a calculation, posterior expectation values of θ can be calculated by numerical integration. A numerical maximization routine can also be used to find the maximum of $L(D|\theta)$ and thus provide a maximum likelihood estimate of the parameters. In this case, the uncertainty is given by the Fisher information found by solution of the stochastic master equation with samples of simulated detection records. The posterior probability density may have many local maxima and it can be difficult to find the global maximum of $L(D|\theta)$ using standard maximization techniques.

If the parameter space is very large, more efficient methods for investigating the likelihood function exist [17,18]. To sample a function with an un-normalized probability density $\pi(x)$ [here $\pi(\theta) = L(D|\theta)P(\theta)$], one can apply a random process for the candidate values in the form of a Markov chain, where the values jump in an appropriately chosen manner so that they attain the correct probabilities. The transition probability $t(x_1 \to x_2)$ must hence be chosen such that it asymptotically reproduces the (un-normalized) probability density $\pi(x)$. The requirement for the transition rule t is then that the only function that satisfies $\int dx f(x)t(x \to x') = f(x')$ is proportional to our desired $\pi(x)$. A generic way to construct such a Markov chain is the Metropolis-Hastings algorithm [17,18], which is used in many areas of science, and we provide a brief review of our application of the method.

The basic idea is to compare the relative probability densities of a randomly chosen candidate value x_2 with the one of the current value x_1. The value x_2 may be chosen randomly or, more conveniently, according to a *proposal chain* $q(x_1 \to x_2)$, e.g., in the neighborhood of x_1. A correct sampling of the probability density is obtained by accepting x_2 with the probability

$$\alpha(x_1,x_2) = \min\left(1, \frac{\pi(x_2)q(x_2 \to x_1)}{\pi(x_1)q(x_1 \to x_2)}\right), \tag{30}$$

and otherwise retaining the value x_1. If the proposal chain is able to explore the entire parameter space this Markov chain will have $\pi(x)$ as un-normalized stationary distribution.

A nice feature of the Metropolis-Hastings sampling method is that it uses only ratios between different arguments of the functions π and q. This implies that we can use the un-normalized probabilities $\pi(x)$, and for our purpose we can use the likelihood functions found by solving (18) or (25) with the parameter values $\theta = x_1$ and $\theta = x_2$.

In summary, to sample the posterior density for the estimated parameters $P(\theta|D) \propto L(D|\theta)P(\theta)$ for a continuous quantum measurement using Metropolis-Hastings we select a random θ_1 from the prior distribution $P(\theta)$, let $i = 1$, and generate a sequence of random parameter values $\{\theta_i\}$ as follows.

(i) Determine candidate θ_c according to some proposal distribution $q(\theta_i \to \theta_c)$.

(ii) Calculate the likelihood or, equivalently, the log likelihood l_T^c for the data until the final time T, using the candidate θ_c and Eqs. (18) and (25) or Eqs. (19) and (26) depending on the type of measurement.

(iii) Calculate $\alpha(\theta_i,\theta_c) = \min(1, \exp(l_T^c - l_T^i)P(\theta_c)q(\theta_c \to \theta_i)/P(\theta_i)q(\theta_i \to \theta_c))$, where l_T^i is the log-likelihood for the previous parameter θ_i.

(iv) Accept the candidate with probability $\alpha(\theta_i,\theta_c)$ setting $\theta_{i+1} = \theta_c$; otherwise, keep θ_i and set $\theta_{i+1} = \theta_i$.

(v) Let $i \to i + 1$.

These steps are repeated a large number of times, and by selecting a suitable fraction of the sequence $\{\theta_i\}$ the sampled parameters are then uncorrelated and sampled according to the probability distribution $P(\theta|D)$. The sampled parameters can be used for determination of any property of the distribution $P(\theta|D)$.

In the simulations presented below, the proposal distribution $q(\theta \to \theta_c)$ was chosen as a multivariate normal distribution centered at θ with a variance selected to achieve a reasonable acceptance rate of 10% to 50%.

Many techniques exist for investigating the convergence rate and the correlation length of the Markov chain generated by the above technique [18]. In the simple examples studied in the present paper, the convergence rate and correlation length are readily identified, but a more careful analysis of these issues is necessary when applying the technique to an experimental situation with many parameters and uncertainties.

V. EXAMPLES

A. Two-level atom

Consider a coherently driven two-level atom that decays by spontaneous emission of photons. The atom is described by the Hamiltonian $H = (\Omega/2)\sigma^x + (\Delta/2)\sigma^z$ and by a jump operator $c = \sqrt{\gamma}\sigma^-$, where γ is the effective decay rate, $\boldsymbol{\sigma} = (\sigma^x,\sigma^y,\sigma^z)$ is the vector of Pauli spin-matrices, and σ^- denotes the Pauli lowering operator. The measurement record is the times at which photons are detected with a photodetector.

The top part of Fig. 1 shows an example trajectory, assuming known values $\gamma = 0.55$, $\Omega = 1.3$, and $\Delta = 1.43$ for the atomic and field parameters (in dimensionless units, e.g., relative to the decay rate of another excited state in the same atom). The continuous curves show the components of the Bloch vector $\boldsymbol{r} = \mathrm{Tr}(\rho\boldsymbol{\sigma})$, and they display continuous evolution disrupted at discrete times, where discontinuous quantum jumps of the

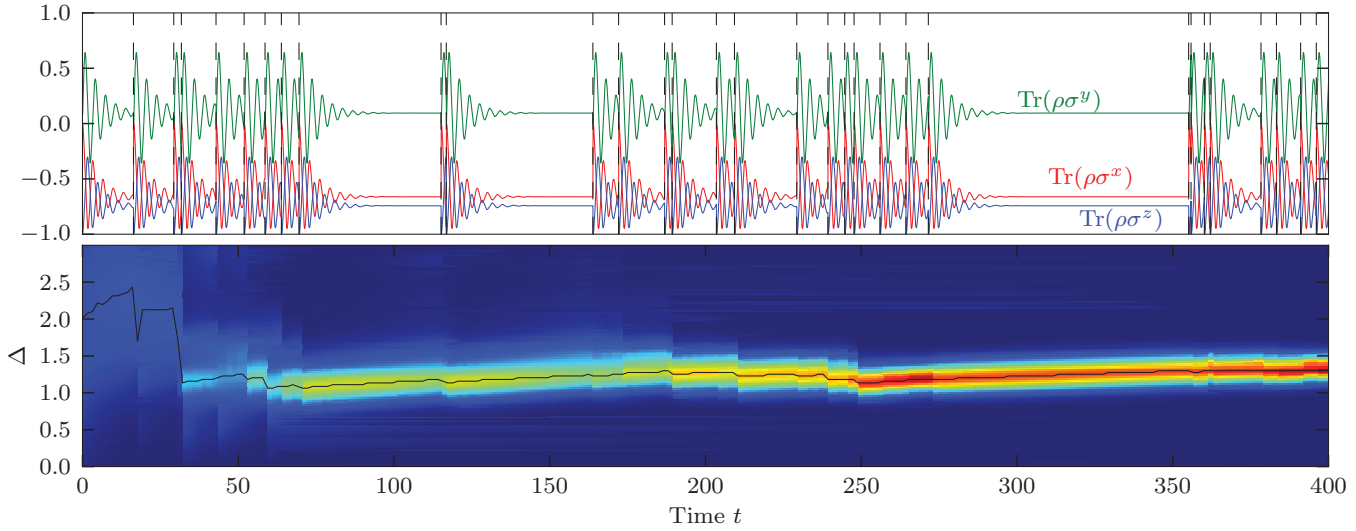

FIG. 1. (Color online) The top panel shows the three components of the Bloch vector for a two-level atom subject to laser excitation with Rabi frequency $\Omega = 1.3$ and detuning $\Delta = 1.43$ (dimensionless units). The atomic inversion is represented by the lower (blue) curve. The atom decays with a rate $\gamma = 0.55$, leading to the observation of quantum jumps at the instants indicated by vertical dashed lines and the transient Bloch vector dynamics. The bottom panel shows the estimation of the detuning Δ, treated as an unknown variable with a probability distribution, which is updated in a Bayesian manner, conditioned on the measurement record. See text.

state occur associated with the detector clicks. In the bottom part of Fig. 1, we have assumed that γ and Ω are known, and we evaluate the probability distribution for the detuning parameter on a grid, assuming a prior normal distribution for Δ with a standard deviation $\sigma_\Delta = 1.0$ and mean value $\mu_\Delta = 2.0$.

The Δ distribution is conditioned on the same detection record as applied in the top part of the figure, and we observe how the no-click periods cause a continuous change of the posterior density, while the discrete jumps are accompanied by more abrupt changes, until the distribution is well converged. The importance of the use of the whole signal and not only the mean photodetection rate is easily understood by the observation that following each quantum jump, the atomic density matrix describes a transient damped Rabi oscillation, and the temporal probability distribution for the subsequent jump event is periodically modulated. Since the period of the transient modulation depends explicitly on Ω and Δ the actual occurrence of the next jump strongly favors (disfavors) certain values of Δ and causes the conditional increase (decrease) in the probability density at those values.

With a single unknown parameter, it is possible to compute the likelihood function on a fine grid, but if we pass to the larger parameter space of more unknown variables, we have recourse to more advanced search methods. In Fig. 2, we show the results of running the Markov chain Monte Carlo algorithm on the trajectory in Fig. 1 with all three parameters Ω, Δ, γ treated as unknown. We assume normal distributed priors, shown with the dashed lines in the left panels of Fig. 2, and the histograms show the values for the three parameters sampled by the Markov chain. Since the trajectory is quite short, there is not sufficient information to perfectly infer the values of the parameters, and the joint densities of pairs of variables in the right panels indicate that two islands of likely values of the set of parameters are not resolved by the measurements.

We have compared the distribution of time differences between click event in the rather short detection record, shown in Fig. 1, with the expected transient Rabi oscillation dynamics and we find that they are, indeed, compatible with the different values for the pair of parameters Ω, Δ, occurring with comparable probabilities in the top right panel in Fig. 2. With a few more "lucky clicks," however, the distribution will favor one choice, and due to the correlations between our estimates for all three parameters, they may then rather rapidly all converge to the correct values.

We have also calculated the Fisher information matrix for a photon counting experiment by applying the simulation methods described above, and we obtain the results shown in Fig. 3. The Fisher information matrix was calculated by simulating the stochastic master equation and the associated equations for the ρ_t^i for different choices of the parameters $(\Omega, \Delta) \in [-3/2, 3/2] \times [-3/2, 3/2]$, while the decay rate was assumed to be known and equal to $\gamma = 0.55$ in our dimensionless units and the initial atomic state was unexcited.

We recall that the Fisher information, evaluated at the estimated values Ω,Δ gives the uncertainties of these two quantities as well as their covariance. It is not surprising that the sensitivity of the photon detection method depends on the actual values of the parameters. The fact that Ω and Δ enter the problem as coefficients on noncommuting spin components in the atomic Hamiltonian suggests that spin uncertainty relations may result in limitations on their joint determination; see also [13]. Such a fundamental limitation may be reflected by the apparent anticorrelation of the occurrence of large and small values of the Fisher information matrix elements $I_{\Omega,\Omega}$ and $I_{\Delta,\Delta}$ in Fig. 3.

The relative entropy between the signal probability p and a Poisson reference distribution p_0, $S(p|p_0) = \mathbb{E}[\ln L_t]$, i.e., the p-expectation value of $\ln L_t$ is shown in Fig. 3(d). The reference distribution p_0 has been chosen as a Poisson

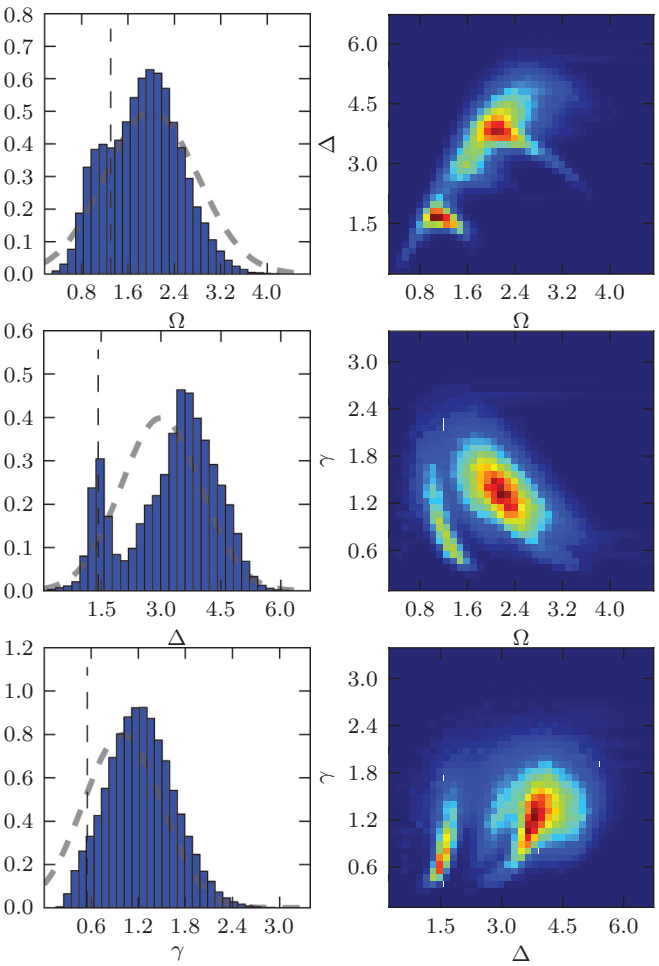

FIG. 2. (Color online) The left panels show histograms of Markov chain Monte Carlo sampled distributions of the parameters Ω, Δ, γ in our two-level atom model. The prior knowledge of the parameters assumes normal distributions, shown by the dashed lines, with mean values $\mu_\Omega = 2.0$, $\mu_\Delta = 3.0$, and $\mu_\kappa = 1.0$ and standard deviations $\sigma_\Omega = 0.8$, $\sigma_\Delta = 1.0$, and $\sigma_\kappa = 0.5$. The right panels display the correlations between the different pairs of sampled parameters.

process with a rate set by the stationary emission rate $\lambda_{st} = \Omega^2\gamma/(\gamma^2 + 4\Delta^2 + 2\Omega^2)$. The relative entropy is close to zero in large regions of the Ω,Δ parameter space, indicating that the emission process is not very different from a Poisson process. In the regions with $|\Omega| \geqslant 0.5$, $\Delta \approx 0$, the dynamics deviate significantly from a Poisson process due to the Rabi oscillations in the photon waiting-time distribution, and the ensemble of trajectories has a higher entropy.

B. Bimodal two-level atom

Imagine now a situation where the two-level atom is not subject to dynamics with a fixed set of unknown parameters, but it may jump randomly between two fixed sets of values. Such jumps may occur due to changes in a binary variable in the surrounding environment. e.g., the quantum states $|a\rangle$ and $|b\rangle$ of a nearby atom, spin, or mesoscopic qubit degree of freedom, or due to the atom moving spatially between two different positions in a laser field configuration. We assume these state changes are purely classical; i.e., we neglect all

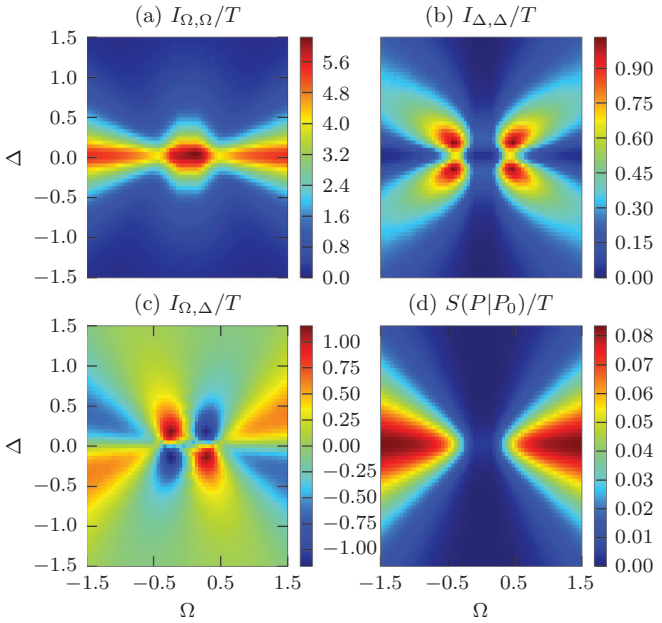

FIG. 3. (Color online) Fisher Information matrix components for photodetection of a decaying two-level atom with decay rate $\gamma = 0.55$ initially in the unexcited state up to $T = 40$. The two top panels show the diagonal elements $I_{\Omega,\Omega}$ and $I_{\Delta,\Delta}$ and the bottom left panel shows the off-diagonal element $I_{\Omega,\Delta}$ of the Fisher information matrix. The bottom right panel shows the relative entropy between the signal probability distribution p and p_0, where p_0 is a Poissson process. Unlike the Fisher information shown in panels (a)–(c), the relative entropy does depend on the parameter λ, chosen for p_0, and it is shown here with the value of the stationary emission rate for the two-level atom $\lambda_{st} = \Omega^2\gamma/(\gamma^2 + 4\Delta^2 + 2\Omega^2)$ and $\gamma = 0.55$. See text.

coherences between the configurations or positions $|a\rangle$ and $|b\rangle$, and we assume that both the Rabi frequency, the detuning, and the decay rate of the two-level atom have different values for the two states.

We describe the system using a conditional master equation where we include the environmental states $|a\rangle$ and $|b\rangle$ of the atoms in a block-diagonal density matrix, $\rho = \rho_a \otimes |a\rangle\langle a| + \rho_b \otimes |b\rangle\langle b|$, where ρ_a (ρ_b) is the density matrix for the atom associated with the environmental state a (b). The system Hamiltonian is $H = [(\Omega_a/2)\sigma^x + (\Delta_a/2)\sigma^z] \otimes |a\rangle\langle a| + [(\Omega_b/2)\sigma^x + (\Delta_b/2)\sigma^z] \otimes |b\rangle\langle b|$ and the effective photodetection jump operator is

$$c = \sigma^- \otimes (\sqrt{\gamma_a}\,|a\rangle\langle a| + \sqrt{\gamma_b}\,|b\rangle\langle b|).$$

The transitions between the two configurations are described by incoherent jumping rates $W(a \to b)$ and $W(b \to a)$ and corresponding jump operators $J_{a\to b} = \sqrt{W(a \to b)}\mathbb{1} \otimes |b\rangle\langle a|$ and $J_{b\to a} = \sqrt{W(b \to a)}\mathbb{1} \otimes |a\rangle\langle b|$. The system is now equivalent to an enlarged quantum system, and it is fully described as a single quantum system by the formalism outlined above.

We have used the parameters $\Omega_a = 1.1$, $\Delta_a = 1.3$, $\gamma_a = 1.6$, $\Omega_b = 2.2$, $\Delta_b = 0.2$, $\gamma_b = 2.4$, and (slow) transition rates $W(a \to b) = 0.03$ and $W(b \to a) = 0.08$ to simulate a typical detection record for the system. In Fig. 4, the black solid line shows the time-binned observed signal for this record as a function of time. As the changes between the two sets

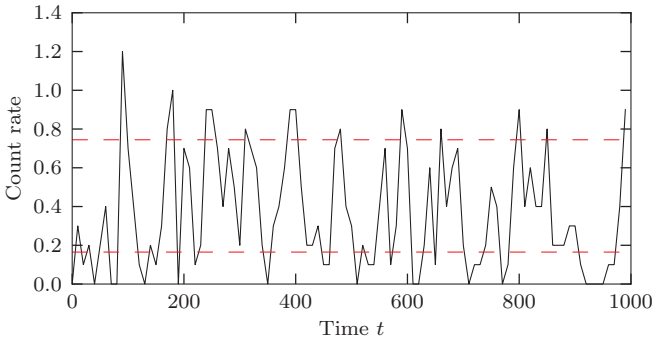

FIG. 4. (Color online) Simulated signal from a bimodal two-level atom, undergoing jumps and coherent evolution with two alternating sets of parameters, a and b. The values used for this trajectory are $\Omega_a = 1.1$, $\Delta_a = 1.3$, $\gamma_a = 1.6$, $\Omega_b = 2.2$, $\Delta_b = 0.2$, $\gamma_b = 2.4$, and transition rates $W(a \to b) = 0.03$ and $W(b \to a) = 0.08$ (dimensionless units). The solid black curve is the (binned) observed signal while the red dashed curves show the mean expected counts for the two sets of parameters.

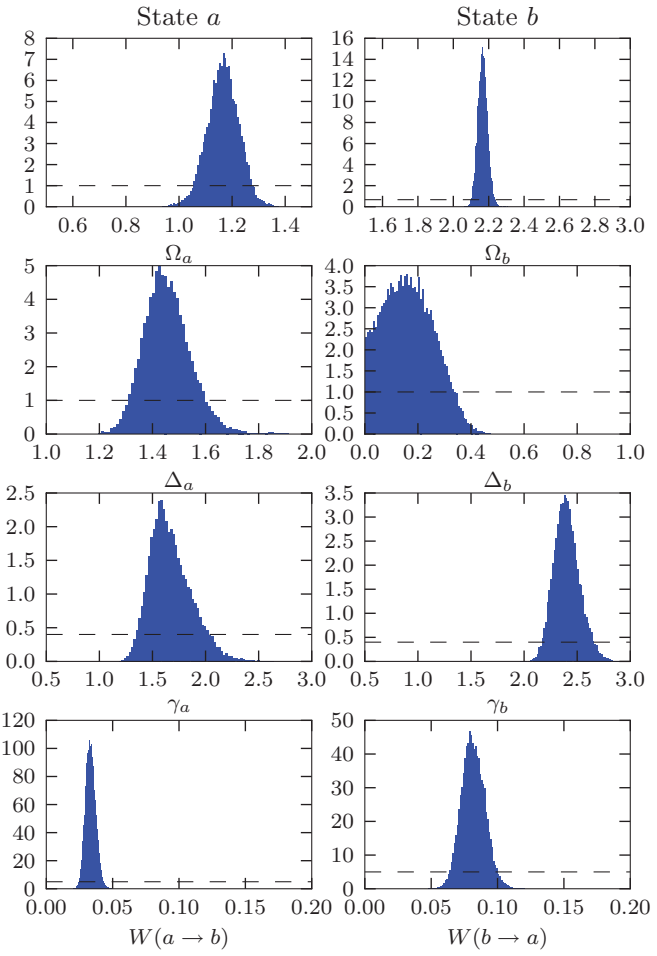

FIG. 5. (Color online) Marginal distributions for the eight unknown parameters in our bimodal two-level atomic system. All prior distributions were taken to be uniform on the shown intervals as indicated by the black dashed line. The estimation was based on the actual click events of the trajectory, represented in Fig. 4.

of parameter occur at low rates, the photon counting signals almost directly reveal the jumps between the classical states a and b as in [19].

Treating all rates and coupling strengths as unknown, the large number of unknown parameters makes a straightforward Bayesian estimation of their values very complicated. In Fig. 5 we show instead the outcome of the Markov chain Monte Carlo sampling of the eight possible parameters over the same measurement sequence represented in Fig. 4. All values were assigned uniform prior probability distributions on the intervals shown (dashed lines in the figures), and the histograms show the concentration of the values sampled on the actual, correct parameters.

VI. DISCUSSION AND OUTLOOK

In this paper we have presented a general method for inferring the values of parameters that govern the time evolution of continuously monitored quantum systems. The systems are described by stochastic master equations, and we have shown that the trace of the un-normalized density matrix can be interpreted and applied as a likelihood function in standard statistical methods for parameter inference. Explicit differential equations for the likelihood function in terms of the normalized density matrix are exemplified for the case of photon counting (19) and homodyne photodetection (26). The differential equations for the likelihood allows us to use numerically stable and efficient stochastic master equation simulations as input to a variety of standard statistical estimation algorithms, e.g., Markov chain Monte Carlo and direct maximum likelihood estimation. Our identification of the conditioned density matrix dynamics with the likelihood function, in addition, leads to an efficient method (28) and (29) to simulate the Fisher information associated with any particular measurement scheme and thus to evaluate the confidence of parameter estimation by continuous measurements.

We presented our formalism for the case of photodetection, and in our examples we assumed that all emitted radiation is detected. If there are unobserved decoherence or loss processes and, e.g., loss of the radiation signal before the detection, averaging over these processes simply contributes further (deterministic) dissipation terms of the Lindblad form $\mathcal{D}[J](\rho) = -\left\{J^\dagger J, \rho\right\}/2 + J\rho J^\dagger$ in the master equations (1) and (2). The likelihood equations (19) and (26), however, remain unchanged.

A technical element in our formulation of the theory involves the introduction of a reference probability p_0, imposing a degree of freedom in the normalization of the effect operators and the density matrix conditioned on the measurement signal. The introduction of the reference probability p_0 is mathematically equivalent to converting the set of measurement outcomes, M, into a classical probability space with a reference probability measure P_0 and the quantum measurement effect operators induce a probability measure on M via the relation $P(A) = \int_A dP_0(m)\,\mathrm{Tr}(\tilde{\rho}|m)$ for subsets $A \subset M$. In mathematical terms the likelihood function $\mathrm{Tr}(\tilde{\rho}|m)$, discussed in Sec. II, can then be identified as the Radon-Nikodym derivative $dP/dP_0(m)$. This points to a further generalization by transforming the probability measure

P to some other measure $\tilde{P}$ such that $dP/d\tilde{P} = Z_t$, where Z_t is a martingale, and where trajectories generated with the transformed probability measure $\tilde{P}$ should be weighted by Z_t to obtain ensemble averages. This may provide a useful technique to control the variance in numerically calculated ensemble averages and to simulate the master equation and the Fisher information matrix more efficiently.

The relative entropy $S(P|P_0)$ between the probability measures P and P_0 is the P-expectation value of $\ln L_t$, $S(P|P_0) = \mathbb{E}[\ln L_t] = \int dP \ln(dP/dP_0)$ and we note that the Fisher information is nothing but the relative entropy between P_θ and $P_{\theta'}$ for infinitesimally close θ and θ'. Apart from their importance in parameter estimation, emphasized in this paper, the entropy $S(P|P_0)$ and the Fisher information I_{ij} provide means to characterize the stochastic dynamics of quantum trajectories in a manner similar to the use of entanglement susceptibility to characterize the correlations in quantum many-body physics [20–23].

[1] S. L. Braunstein and C. M. Caves, Phys. Rev. Lett. **72**, 3439 (1994).

[2] H. M. Wiseman, Quantum Semiclass. Opt. **8**, 205 (1996).

[3] S. Boixo and R. D. Somma, Phys. Rev. A **77**, 052320 (2008).

[4] J. Grond, U. Hohenester, J. Schmiedmayer, and A. Smerzi, Phys. Rev. A **84**, 023619 (2011).

[5] B. Lücke, M. Scherer, J. Kruse, L. Pezzé, F. Deuretzbacher, P. Hyllus, O. Topic, J. Peise, W. Ertmer, J. Arlt, L. Santos, A. Smerzi, and C. Klempt, Science **334**, 773 (2011).

[6] V. Giovannetti, S. Lloyd, and L. Maccone, Science **306**, 1330 (2004).

[7] W. Wasilewski, K. Jensen, H. Krauter, J. J. Renema, M. V. Balabas, and E. S. Polzik, Phys. Rev. Lett. **104**, 133601 (2010).

[8] V. Shah, G. Vasilakis, and M. V. Romalis, Phys. Rev. Lett. **104**, 013601 (2010).

[9] S. Kolkowitz, Q. P. Unterreithmeier, S. D. Bennett, and M. D. Lukin, Phys. Rev. Lett. **109**, 137601 (2012).

[10] N. Zhao, J. Honert, B. Schmid, M. Klas, J. Isoya, M. Markham, D. Twitchen, F. Jelezko, R.-B. Liu, H. Fedder, and J. Wrachtrup, Nat. Nano. **7**, 657 (2012).

[11] J. Gambetta and H. M. Wiseman, Phys. Rev. A **64**, 042105 (2001).

[12] A. Negretti and K. Mølmer, arXiv:1211.2123v2 [quant-ph].

[13] V. Petersen, L. B. Madsen, and K. Mølmer, Phys. Rev. A **71**, 012312 (2005).

[14] J. M. Geremia, J. K. Stockton, A. C. Doherty, and H. Mabuchi, Phys. Rev. Lett. **91**, 250801 (2003).

[15] H. Cramér, *Mathematical methods of statistics*, Princeton Mathematical Series No. 9 (Princeton University Press, Princeton, NJ, 1954).

[16] P. Goetsch and R. Graham, Phys. Rev. A **50**, 5242 (1994).

[17] W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, and S. A. Teukovsky, *Numerical Recipes: The Art of Scientific Computing*, 3rd ed. (Cambridge University Press, Cambridge, UK, 2007).

[18] *Markov Chain Monte Carlo in Practice: Interdisciplinary Statistics*, edited by W. Gilks, S. Richardson, and D. Spiegelhalter (Taylor & Francis, New York, 1995), p. 512.

[19] S. Reick, K. Mølmer, W. Alt, M. Eckstein, T. Kampschulte, L. Kong, R. Reimann, A. Thobe, A. Widera, and D. Meschede, J. Opt. Soc. Am. B **27**, A152 (2010).

[20] P. Zanardi, H. T. Quan, X. Wang, and C. P. Sun, Phys. Rev. A **75**, 032109 (2007).

[21] S.-J. Gu, H.-M. Kwok, W.-Q. Ning, and H.-Q. Lin, Phys. Rev. B **77**, 245109 (2008).

[22] W.-L. You, Y.-W. Li, and S.-J. Gu, Phys. Rev. E **76**, 022101 (2007).

[23] P. Zanardi, P. Giorda, and M. Cozzini, Phys. Rev. Lett. **99**, 100603 (2007).