Supplementary Information to Past quantum states of a monitored system

Søren Gammelmark, Brian Julsgaard, and Klaus Mølmer

Department of Physics and Astronomy, Aarhus University,

Ny Munkegade 120, DK-8000 Aarhus C, Denmark.

(Dated: August 22, 2013)

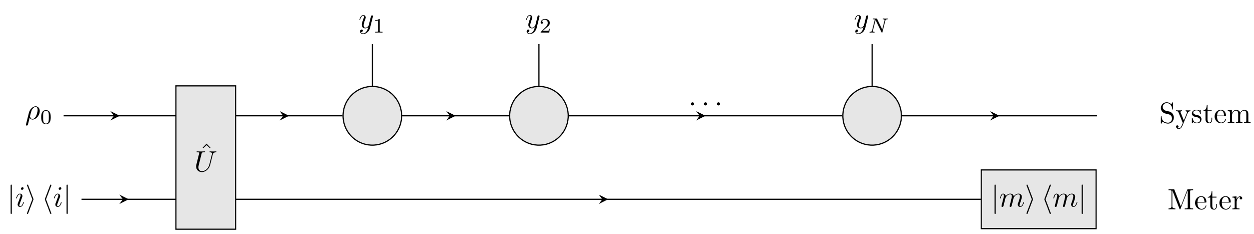

FIG. 1. Graphical illustration of the system dynamics considered in this note. Our open quantum system (top line) and a meter (bottom line) interact as described by the unitary operator $\hat{U}$. The system is probed by subsequent measurements with outcomes $y_1, \ldots y_N$, and later, the meter is subject to a projective measurement in some basis $\{|m\rangle\}$.

In this supplementary note we show that with our definition of the past quantum state we obtain identical results for past measurement outcomes as predicted by the ordinary quantum formalism, when this is applied to deferred measurements on a combined system-meter set-up [1]. Instead of an expression for determining past measurement outcomes propagating forward in time and including the meter system, we derive a backward propagating *effect matrix* E for the system only. We provide the specific stochastic differential equations for diffusion and jump type probing of the system (corresponding to homodyne detection and photon counting schemes). These equations can be chosen linear and non-trace preserving or non-linear and trace-preserving as the usual quantum filtering equations [2]. Finally, we demonstrate that our past quantum state generalizes the so-called *smoothed* state in classical hidden Markov models to the quantum case: The conditional density matrix $\rho(t)$ is a natural generalization of the α-state conditional probability, and the backward propagating effect matrix E is the quantum generalization of the backward propagating β-state in hidden Markov models [3].

A. DEFERRED MEASUREMENTS AND THE PAST QUANTUM STATE

Imagine an observer using a meter to perform any measurement on our system which is initially prepared in a state which is represented by the density matrix ρ_0. The observer correlates the meter and our system by a unitary interaction, $\hat{U}$, between them. After this unitary interaction the system evolves independently of the meter and during this evolution, we measure our system a number of times N as illustrated in Fig. 1. By taking a suitable limit with $N \to \infty$ a description of continuous-time observation can be obtained. The observer who now has the meter in her possession can choose to perform a measurement of her choice at any later time. Our goal is to predict the result of such a measurement.

We have access to the results of the N measurements, where each measurement is in the most general case described by a set of measurement effect operators $\hat{M}_y$ for each measurement outcome y in the set of possible measurement results Y. Using this formalism the effect of the measurement result y is to update a density matrix ρ according to $\rho \overset{y}{\mapsto} \hat{M}_y \rho \hat{M}_y^\dagger / \operatorname{Tr}(\hat{M}_y^\dagger \hat{M}_y \rho)$. The resulting density matrix, which we denote $\rho|y$, is conditioned upon the measurement result y. The probability of obtaining the result y is given by $P(y) = \operatorname{Tr}(\hat{M}_y^\dagger \hat{M}_y \rho)$.

The operators $\hat{M}_y$ should satisfy $\sum_{y \in Y} \hat{M}_y^\dagger \hat{M}_y = \hat{I}$ such that the probability for observing any $y \in Y$ is unity. If the observation is made with less than 100% readout efficiency the resulting conditioned density matrix can be written as a sum

$$\rho \overset{y}{\longmapsto} \frac{\sum_{k=1}^{K} \hat{M}_{k|y} \rho \hat{M}_{k|y}^\dagger}{\operatorname{Tr}\left(\sum_{k=1}^{K} \hat{M}_{k|y}^\dagger \hat{M}_{k|y} \rho\right)}, \tag{A1}$$

where $\hat{M}_{k|y}$ are operators describing different possible effects which are all associated with the measurement result y. The number of terms, K, can depend on y.

Equation (A1) captures the effect of most types of dynamics of open quantum systems, including unitary evolution, dissipation effects described by a master equation on Lindblad form, and measurements. A density matrix subject to unitary evolution with unitary operator $\hat{U}$ is thus updated according to $\rho \mapsto \hat{M} \rho \hat{M}^\dagger$. Where only the single unitary operator $\hat{M}$ is needed clearly satisfies $\hat{M}^\dagger \hat{M} = \hat{I}$.

A projective measurement is described by a set of projectors $\hat{\Pi}_a$ where a are eigenvalues of the observable $\hat{A}$ being measured, and this case is also included by the identification $\hat{M}_a = \hat{\Pi}_a$. The effect on the system density matrix by a projective measurement is then simply the usual projection postulate $\rho \overset{a}{\longmapsto} \hat{\Pi}_a \rho \hat{\Pi}_a / \operatorname{Tr}(\hat{\Pi}_a \rho)$. If we know that a projection measurement is performed, but the result of the measurement is hidden from us, the density matrix is updated according to $\rho \longmapsto \sum_a \hat{\Pi}_a \rho \hat{\Pi}_a$. In this final example, all off-diagonal density matrix elements in the eigen basis of $\hat{A}$ are zeroed, and the unobserved measurement operation is therefore equivalent to a decoherence process.

If a number of measurements are performed in sequence the density matrix ρ is updated repeatedly by the formula Eq. (A1). In the limit of continuous-time measurements we can also describe the conditional time-evolution of the system density matrix. Consider for example a quantum system subject to homodyne detection. In this case the effect of a detection in a small interval of time dt is given by the operators

$$\hat{M}_{dY_t} = (2\pi dt)^{-1/4} \exp(-dY_t^2/4dt)(\hat{I} - i\hat{H}dt - \hat{c}^\dagger\hat{c}/2dt + \hat{c}dY_t), \tag{A2}$$

where $\hat{H}$ is the system Hamiltonian and $\hat{c}$ is the system operator coupling to the real homodyne output signal dY_t. [2, 4] The probability for observing dY_t in this infinitesimal interval of time is a normal distribution with mean value $\mathrm{Tr}((\hat{c}+\hat{c}^\dagger)\rho)dt$ and variance dt. By applying the update formula Eq. (A1) with these operators we get Eq. (3) with $\eta = 1$ and only one Lindblad operator $\hat{L}_1 = \hat{c}$. By including unobserved output channels for Lindblad operator $\hat{L}_i$, $i > 1$ and including limited detector efficiency $\eta < 1$, Eq. (3) turns out to be a special case of Eq. (A1).

Let us return to our main line of inquiry. The meter is assumed to be initialized in a pure state $|i\rangle\langle i|$ when the observer applies the unitary interaction, acting on the combined system-meter state $\rho_0 \otimes |i\rangle\langle i|$. The resulting state is $\hat{U}(\rho_0 \otimes |i\rangle\langle i|)\hat{U}^\dagger$. By inserting complete bases of the meter $\hat{I}_M = \sum_{m\in M} |m\rangle\langle m|$ we obtain

$$\rho = \sum_{m,m'\in M} \hat{\Omega}_m \rho_0 \hat{\Omega}_{m'}^\dagger \otimes |m\rangle\langle m'|, \tag{A3}$$

where we have defined the system operators $\hat{\Omega}_m = (\hat{I} \otimes \langle m|)\hat{U}(\hat{I} \otimes |i\rangle)$. The operators $\hat{\Omega}_m$, defined this way, satisfy the requirement for measurement effect operators, $\sum_{m\in M} \hat{\Omega}_m^\dagger \hat{\Omega}_m = \hat{I}$, and with the chosen form for the operators $\hat{\Omega}_m$, we use the coupling to the meter to formally interrogate the properties of the open quantum system at the intermediate time t. We imagine that the open system dynamics proceeds, including the observations on the system continue with measurements result $y \in Y$, which cause the continued system dynamics described by Eq. (A1). In the general formula Eq. (A1) we can include unitary dynamics, dissipation channels and partially efficient measurements. As noted above, all these effects can be included by a suitable choice of the operators $\hat{M}_{k|y}$.

Since the subsequent dynamics only concern the system, the operators $\hat{M}_{k|y}$ only act on the system degrees of freedom. The normalized system-meter state conditioned on one measurement with the result y is

$$\rho|y = \frac{\sum_{k=1}^{K}\sum_{m,m'\in M}\hat{M}_{k|y}\hat{\Omega}_m\rho_0\hat{\Omega}_{m'}^\dagger\hat{M}_{k|y}^\dagger\otimes|m\rangle\langle m'|}{\sum_{k=1}^{K}\sum_{m\in M}\mathrm{Tr}\left(\hat{M}_{k|y}\hat{\Omega}_m\rho_0\hat{\Omega}_m^\dagger\hat{M}_{k|y}^\dagger\right)}. \tag{A4}$$

We can calculate the expectation value of any meter-observable $\hat{X}$ in the state $\rho|y$ by the usual formalism,

$$\mathbb{E}[\hat{X}|y] = \mathrm{Tr}(\hat{I}\otimes\hat{X}\rho|y) = \frac{\sum_{k=1}^{K}\sum_{m,m'\in M}\mathrm{Tr}(\hat{M}_{k|y}\hat{\Omega}_m\rho_0\hat{\Omega}_{m'}^\dagger\hat{M}_{k|y}^\dagger)\langle m'|\hat{X}|m\rangle}{\sum_{k=1}^{K}\sum_{m'\in M}\mathrm{Tr}\left(\hat{M}_{k|y}\hat{\Omega}_{m'}\rho_0\hat{\Omega}_{m'}^\dagger\hat{M}_{k|y}^\dagger\right)}. \tag{A5}$$

A projective measurement on the meter, which is now conditional on the system measurement result y, and yields the result m with the probability

$$P(m|y) = \frac{\sum_{k=1}^{K}\mathrm{Tr}(\hat{M}_{k|y}\hat{\Omega}_m\rho_0\hat{\Omega}_m^\dagger\hat{M}_{k|y}^\dagger)}{\sum_{k=1}^{K}\sum_{m'\in M}\mathrm{Tr}\left(\hat{M}_{k|y}\hat{\Omega}_{m'}\rho_0\hat{\Omega}_{m'}^\dagger\hat{M}_{k|y}^\dagger\right)}, \tag{A6}$$

where we have calculated the conditional expectation value of the meter projection operator $\hat{X} = |m\rangle\langle m|$ to obtain the conditional probability $P(m|y)$. Note that we can rewrite the numerator in (A6) as

$$\sum_{k=1}^{K}\mathrm{Tr}(\hat{M}_{k|y}\hat{\Omega}_m\rho_0\hat{\Omega}_m^\dagger\hat{M}_{k|y}^\dagger) = \mathrm{Tr}\left(\hat{\Omega}_m\rho_0\hat{\Omega}_m^\dagger\left[\sum_{k=1}^{K}\hat{M}_{k|y}^\dagger\hat{M}_{k|y}\right]\right) \equiv \mathrm{Tr}(M_m\rho_0M_m^\dagger E), \tag{A7}$$

where we have defined the *effect matrix* $E = \sum_{k=1}^{K}\hat{M}_{k|y}^\dagger\hat{M}_{k|y}$, and we obtain

$$P(m|y) = \frac{\mathrm{Tr}(\hat{\Omega}_m\rho_0\hat{\Omega}_m^\dagger E)}{\mathrm{Tr}(\sum_{m'\in M}\hat{\Omega}_{m'}\rho_0\hat{\Omega}_{m'}^\dagger E)}.. \tag{A8}$$

This is the probability for the outcomes, stored "in the safe" as described in the main text. *I.e.*, $P(m|y)$ is the retrodicted probability for the result m of a general measurement on the system with the corresponding measurement operator $\{\hat{\Omega}_m\}$. It notably differs from the usual formula $P(m) = \mathrm{Tr}(\hat{\Omega}_m^\dagger\hat{\Omega}_m\rho_0)$, as it is possible to predict the outcome of measurements on the meter better than before we had access to the later measurement result y.

In the same way as the usual time dependent quantum state of a system, represented by a wave function $\psi(t)$ or a density matrix $\rho(t)$, yields probabilities for general measurements on the system, we have now identified a mathematical structure, composed of ρ_0 and E, which provide the probabilities for past measurements on a quantum system. We thus call the pair of matrices $\Xi = (\rho_0, E)$ *the past quantum state.*

If the system has evolved until time t, and multiple measurements have been performed before the meter interacts with our system, $\Xi(t) = (\rho(t), E(t))$, where $\rho(t)$ is the usual open system density matrix found by a stochastic equation of evolution conditioned on the measurements *before* time t. If multiple measurements are performed in sequence after the meter has interacted with our system at time t, the effect matrix $E(t)$ depends on all measurement results *after* the coupling to the meter. In the following sections, we will derive efficient equations of evolution to determine E for both general probing scenarios and for a few special cases.

1. Dynamical equations for the effect matrix

Assume that the measurements up to time t has been taken into account in the forward state ρ_0 then the generalization of Eq. (A5) when two subsequent measurements with result y_1 and y_2 are performed is

$$\mathbb{E}[\hat{X}|y_1, y_2] = \frac{\sum_{m,m'\in M} \sum_{k_1,k_2=1}^{K} \mathrm{Tr}\left(\hat{M}_{k_2|y_2}\hat{M}_{k_1|y_1}\hat{\Omega}_m\rho_0\hat{\Omega}^\dagger_{m'}\hat{M}^\dagger_{k_1|y_1}\hat{M}^\dagger_{k_2|y_2}\right)\langle m'|\hat{X}|m\rangle}{\sum_{m,\in M} \sum_{k_1,k_2=1}^{K} \mathrm{Tr}\left(\hat{M}_{k_2|y_2}\hat{M}_{k_1|y_1}\hat{\Omega}_m\rho_0\hat{\Omega}^\dagger_{m}\hat{M}^\dagger_{k_1|y_1}\hat{M}^\dagger_{k_2|y_2}\right)}. \quad \text{(A9)}$$

By using the cyclic property of the trace, the numerator can be written as

$$\sum_{m,m'\in M} \mathrm{Tr}\left(\hat{\Omega}_m\rho_0\hat{\Omega}^\dagger_{m'}\sum_{k_1=1}^{K}\left\{\hat{M}^\dagger_{k_1|y_1}\left[\sum_{k_2=1}^{K}\hat{M}^\dagger_{k_2|y_2}\hat{M}_{k_2|y_2}\right]\hat{M}_{k_1|y_1}\right\}\right)\langle m'|\hat{X}|m\rangle\,, \quad \text{(A10)}$$

In this case, the effect matrix E is therefore given by

$$E = \sum_{k_1=1}^{K}\hat{M}^\dagger_{k_1|y_1}\left[\sum_{k_2=1}^{K}\hat{M}^\dagger_{k_2|y_2}\hat{M}_{k_2|y_2}\right]\hat{M}_{k_1|y_1}, \quad \text{(A11)}$$

where E now depends explicitly on the two future measurement results y_1 and y_2.

From this we see that the update formula for E, as a counterpart to Eq. (A1), is given by the adjoint update

$$E \xmapsto{y} \sum_{k=1}^{K}\hat{M}^\dagger_{k|y}E\hat{M}_{k|y} \quad \text{(A12)}$$

where E equals the identity $\hat{I}$ at the final time of measurements T, and is propagated recursively *backward* as indicated in the case of two measurements,

$$\hat{I} \xmapsto{y_2} \sum_{k_2=1}^{K}\hat{M}^\dagger_{k_2|y_2}\hat{M}_{k_2|y_2} \xmapsto{y_1} \sum_{k_1=1}^{K}\hat{M}^\dagger_{k_1|y_1}\left[\sum_{k_2=1}^{K}\hat{M}^\dagger_{k_2|y_2}\hat{M}_{k_2|y_2}\right]\hat{M}_{k_1|y_1}. \quad \text{(A13)}$$

The propagation is readily generalized to the case of N measurements,

$$E_N = \hat{I} \overset{y_N}{\longmapsto} E_{N-1} \overset{y_{N-1}}{\longmapsto} \ldots \overset{y_1}{\longmapsto} E_0. \tag{A14}$$

A hermitian operator E remains hermitian since the right hand side of Eq. (A12) is invariant under hermitian conjugation. Indeed, E has a separate physical interpretation as the positive semi-definite operator which given the state ρ_0 yields the probability for the sequence of future measurement results, $P(y_1, \ldots y_N|\rho_0) = \mathrm{Tr}(E\rho_0)$.

2. The past density matrix

Our definition of the past quantum state necessitates the use of two matrices from which probabilities can be generally determined via (A8). That expression may, however, be simplified in the special, but interesting, case where the system and meter are coupled very weakly.

We choose a two-dimensional quantum system as our meter, represented as a spin-1/2-particle, interacting briefly with our system by an interaction $\hat{V} = ig(\hat{A}\hat{\sigma}^\dagger - \hat{A}^\dagger\hat{\sigma})$ where $\hat{\sigma}$ is the spin lowering operator and $\hat{A}$ is *any*, not necessarily hermitian, system operator. The meter is initially prepared in the spin-down state $|{\downarrow}\rangle$, and we allow the interaction to be active for a duration τ such that $\hat{U} = \exp(\epsilon(\hat{A}\hat{\sigma}^\dagger - \hat{A}^\dagger\hat{\sigma}))$ where $\epsilon = \tau g$. In the (weak) limit $\epsilon \ll 1$

$$\hat{U} = \hat{I} + \epsilon(\hat{A}\hat{\sigma}^\dagger - \hat{A}^\dagger\hat{\sigma}) - \frac{\epsilon^2}{2}\left(\hat{A}\hat{A}^\dagger\hat{\sigma}^\dagger\hat{\sigma} + \hat{A}^\dagger\hat{A}\hat{\sigma}\hat{\sigma}^\dagger\right) + O(\epsilon^3). \tag{A15}$$

If the meter is initialized in spin down in the z-direction $|{\downarrow}\rangle$ then the measurement effect operators in the z-basis are given by $\hat{\Omega}_\mu = (\hat{I} \otimes \langle\mu|)\hat{U}(\hat{I} \otimes |{\downarrow}\rangle)$, $\mu = \downarrow, \uparrow$,

$$\hat{\Omega}_\downarrow = \hat{I} - \frac{\epsilon^2}{2}\hat{A}^\dagger\hat{A} + O(\epsilon^3) \tag{A16}$$

$$\hat{\Omega}_\uparrow = \epsilon\hat{A} + O(\epsilon^3). \tag{A17}$$

The effect of the measurement on the system state ρ_0 when the result is not revealed is

$$\rho_0 \longmapsto \hat{\Omega}_\downarrow\rho_0\hat{\Omega}_\downarrow^\dagger + \hat{\Omega}_\uparrow\rho_0\hat{\Omega}_\uparrow^\dagger = \rho_0 + O(\epsilon^2). \tag{A18}$$

The measurement associated with the subsequent readout of the meter is weak in the sense that it leaves the system undisturbed to first order in ϵ. This type of measurement can thus be performed unnoticed, and our formalism should enable prediction of its outcome in a seemingly "counter-factual" manner ("If $\hat{A}$ had been measured, the outcome would have been ... "). The nature of the weak measurement allows an unknown agent to perform measurements without our knowledge and hide the result from us. At any later time, this person can inform us of the result, and we need the present theory to most accurately predict the outcome of this result. This ability comes with a cost, as to the interpretation of the weak measurement result, which we will return to shortly.

An arbitrary projective measurement of the meter is described by linear combinations of $\hat{\Omega}_{\downarrow}$ and $\hat{\Omega}_{\uparrow}$. For measurements in the $\hat{\sigma}_x$ and $\hat{\sigma}_y$-bases we can express the conditional expectation of $\hat{\sigma}_x = (\hat{\sigma} + \hat{\sigma}^{\dagger})$ and $\hat{\sigma}_y = i(\hat{\sigma} - \hat{\sigma}^{\dagger})$ by the real and imaginary parts of the expectation value of the step down operator $\hat{\sigma}$, respectively.

The expectation value of the meter operator $\hat{\sigma}$ is

$$\mathbb{E}[\hat{\sigma}|y] = \frac{\sum_{k=1}^{K} \mathrm{Tr}(\hat{M}_{k|y} \epsilon \hat{A} \rho_0 \hat{M}_{k|y}^{\dagger})}{\sum_{k=1}^{K} \mathrm{Tr}\left(\hat{M}_{k|y} \rho_0 \hat{M}_{k|y}^{\dagger}\right)}. \tag{A19}$$

In this formula the denominator is independent of $\hat{A}$. This is due to the weak nature of the measurement, and it implies, that the expectation value is *linear* in the system operator $\hat{A}$. Since the meter couples to the system observable $\hat{A}$, it is natural to consider the weak value,

$$\langle A \rangle_{\mathrm{w}} \equiv \lim_{\epsilon \to 0} \frac{1}{\epsilon} \mathbb{E}[\hat{\sigma}|y] = \frac{\sum_{k=1}^{K} \mathrm{Tr}(\hat{M}_{k|y} \hat{A} \rho_0 \hat{M}_{k|y}^{\dagger})}{\sum_{k=1}^{K} \mathrm{Tr}\left(\hat{M}_{k|y} \rho_0 \hat{M}_{k|y}^{\dagger}\right)}. \tag{A20}$$

We observe that to retrieve the weak value, i.e., the average outcome of a weak measurement, the information in the past quantum state $\Xi = (\rho_0, E)$ with the effect matrix $E = \sum_{k=1}^{K} \hat{M}_{k|y}^{\dagger} \hat{M}_{k|y}$ can be deduced from a *past density matrix* ρ_{p}. The expression $\langle A \rangle_{\mathrm{w}} = \mathrm{Tr}(\hat{A} \rho_{\mathrm{p}})$ holds if we identify

$$\rho_{\mathrm{p}} = \frac{\rho_0 E}{\mathrm{Tr}(\rho_0 E)}. \tag{A21}$$

It is worth noting that ρ_{p} is not Hermitian and even if $\hat{A}$ is Hermitian, $\mathrm{Tr}(\hat{A}\rho_{\mathrm{p}})$ may have both real and imaginary parts. This is a well-known property of weak measurements, and it does not require the readout of, non-physical, complex measurement results, since the real part of $\langle A \rangle_{\mathrm{w}}$ refers to the (real) mean value of the $\hat{\sigma}_x$-operator of the meter, while its imaginary part is obtained by measuring the average $\hat{\sigma}_y$-spin component of the meter.

These averages may, in turn, indicate mean values of the system observable that are very different from its spectrum of eigenvalues. But this is well understood as an interference effect, when a measurement outcome is conditioned on different pre- and post-selected state of a physical system. [5]

3. Projective read-out measurements and the past quantum state

Imagine now that we perform a projective read-out measurement of the system observable $\hat{A}$ using our meter. In this case $\hat{\Omega}_m = \hat{\Pi}_{a_m}$ where the different measurement results are the eigenvalues of the observable $\hat{A}$ denoted a_m. Following Eq. (A8) the probability for observing the eigenvalue a conditional on the later measurement result y is

$$P(a|y) = \frac{\mathrm{Tr}(\hat{\Pi}_{a_m}\rho_{0-}\hat{\Pi}_{a_m}E_+)}{\mathrm{Tr}\left(\sum_{m'}\hat{\Pi}_{a_{m'}}\rho_{0-}\hat{\Pi}_{a_{m'}}E_+\right)}$$

where we by the $+$ and $-$ signs emphasize, that E_+ is the effect matrix including measurements from immediately *after* the projective read-out measurement was performed until time T (here exemplified with a single measurement with result y), and ρ_{0-} is the usual quantum state conditioned on the measurements until immediately *before* the projective read-out measurement was performed. The expectation value of the projective measurement of the observable $\hat{A}$ is then

$$\mathbb{E}[\hat{A}|y] = \sum_m a_m P(a_m|y) = \frac{\mathrm{Tr}(\sum_m a_m\hat{\Pi}_{a_m}\rho_{0-}\hat{\Pi}_{a_m}E_+)}{\mathrm{Tr}\left(\sum_{m'}\hat{\Pi}_{a_{m'}}\rho_{0-}\hat{\Pi}_{a_{m'}}E_+\right)}.$$

By inserting a resolution of the identity $\hat{I} = \sum_{m'}\hat{\Pi}_{a_{m'}}$ this expression can be written in two ways

$$\mathbb{E}[\hat{A}|y] = \frac{\mathrm{Tr}\left(\hat{A}\rho_{0-}\sum_m\hat{\Pi}_{a_m}E_+\hat{\Pi}_{a_m}\right)}{\mathrm{Tr}\left(\rho_{0-}\sum_m\hat{\Pi}_{a_m}E_+\hat{\Pi}_{a_m}\right)} = \frac{\mathrm{Tr}\left(\hat{A}\sum_m\hat{\Pi}_{a_m}\rho_{0-}\hat{\Pi}_{a_m}E_+\right)}{\mathrm{Tr}\left(\sum_m\hat{\Pi}_{a_m}\rho_{0-}\hat{\Pi}_{a_m}E_+\right)}.$$

By introducing the effect matrix which takes the unobserved projective measurement into account by the map, $E_- = \sum_m\hat{\Pi}_{a_m}E_+\hat{\Pi}_{a_m}$ we can write this result

$$\mathbb{E}[\hat{A}|y] = \mathrm{Tr}(\hat{A}\rho_{\mathrm{p}-}), \tag{A22}$$

defining the past density matrix as $\rho_{\mathrm{p}-} = \rho_{0-}E_-/\,\mathrm{Tr}(\rho_{0-}E_-)$ immediately prior to the projective measurement.

Alternatively, we get

$$\mathbb{E}[\hat{A}|y] = \frac{\mathrm{Tr}(\hat{A}\rho_{0+}E_+)}{\mathrm{Tr}(\rho_{0+}E_+)}, \tag{A23}$$

by using the past density matrix $\rho_{\mathrm{p}+} = \rho_{0+}E_+/\,\mathrm{Tr}(\rho_{0+}E_+)$, with the unobserved measurement process modifying the state $\rho_{0+} = \sum_m \hat{\Pi}_{a_m}\rho_{0-}\hat{\Pi}_{a_m}$.

The projective measurement disturbs the system, and it must hence be taken into account in one of the components of the past quantum state: In Eq. (A22) it is included in the effect matrix E_- whereas in Eq. (A23) it is included in the forward density matrix ρ_{0+}. The two resulting alternative forms for the expectation value of $\hat{A}$ are equivalent since $\hat{A}$ commutes with the effect of the projective measurement.

While the general formula Eq. (A5) using the past quantum state $\Xi = (\rho_0, E)$ can be applied to yield the probabilities for any past measurement outcome, it is interesting that the formalism may be brought closer to usual mean value expressions by use of the appropriately defined past density matrix.

4. Differential equations for homodyne and counting measurements

A quantum system subject to homodyne or heterodyne detection satisfies an Itô stochastic differential equation of the form given in Eq. (3). As discussed previously such a measurement scenario fits into the present formulation by using the measurement operator for the time interval from t to $t + dt$ is given by Eq. (A2) if we assume 100% efficiency and no unobserved channels, i.e. no dissipation effects. The adjoint update Eq. (A12) is then

$$dE_t \equiv E_{t-dt} - E_t = \left[i\left[\hat{H}, E_t\right] - \frac{1}{2}\left\{\hat{c}^\dagger\hat{c}, E_t\right\} + \hat{c}^\dagger E_t\hat{c}\right]dt + \left[\hat{c}^\dagger E_t + E_t\hat{c}\right]dY_{t-dt}. \tag{A24}$$

where we have chosen a normalization of the effect matrix E such that the front factor $(2\pi dt)^{-1/4}\exp(-dY^2/4dt)$ is discarded and $E(T) = \hat{I}$ where T is the final time of observation.

By including dissipation effects and limited detector efficiency we obtain Eq. (4).

The conditional time evolution of a quantum system state subject to discrete counting signals in N channels can be described by the infinitesimal operators

$$\begin{aligned}\hat{M}_0 &= \hat{I} - i\hat{H}dt - \sum_{m=1}^{N}\frac{1}{2}\hat{L}_m^\dagger\hat{L}_m dt \\ \hat{M}_m &= \hat{L}_m\sqrt{dt} \qquad \text{for } 1 \le m \le N\end{aligned} \tag{A25}$$

where $\hat{L}_m$ are Lindblad operators describing quantum jumps associated with the emission of quanta by the system into the environment. $\hat{M}_0$ yields the measurement effect when no quantum (photon) is detected and $\hat{M}_m$ indicates that a quantum is detected in environment channel number m.

If only the first channel is detected we get the well-established quantum jump filtering equation[2]

$$d\rho_t = \left[-i\left[\hat{H},\rho_t\right] + \sum_{m=2}^{N}\left(\hat{L}_m\rho_t\hat{L}_m^\dagger - \frac{1}{2}\left\{\hat{L}_m^\dagger\hat{L}_m,\rho_t\right\}\right) - \frac{1}{2}\left\{\hat{L}_1^\dagger\hat{L}_1,\rho_t\right\} + \mathrm{Tr}(\hat{L}_1^\dagger\hat{L}_1\rho_t)\rho_t\right]dt + \left[\frac{\hat{L}_1\rho_t\hat{L}_1^\dagger}{\mathrm{Tr}(\hat{L}_1^\dagger\hat{L}_1\rho_t)} - \rho_t\right]dN_t, \quad \text{(A26)}$$

where $dN_t = 0$ in all the intervals where no photon is detected, but $dN_t = 1$ (and $dt = 0$) at the instants of time a photon is detected in channel 1.

The adjoint update of the effect matrix is

$$dE_t \equiv E_{t-dt} - E_t = \left[i\left[\hat{H},E_t\right] + \sum_{m=2}^{N}\left(\hat{L}_m^\dagger E_t\hat{L}_m - \frac{1}{2}\left\{\hat{L}_m^\dagger\hat{L}_m,E_t\right\}\right) - \frac{1}{2}\left\{\hat{L}_1^\dagger\hat{L}_1,E_t\right\}\right]dt + \left[\hat{L}_1^\dagger E_t\hat{L}_1 - E_t\right]dN_t. \quad \text{(A27)}$$

5. Time evolution of mean values and weak values

For a quantum system subject to unit efficiency homodyne detection and no unobserved dissipation channels, the conventional mean value of a system observable $\langle\hat{A}\rangle = \mathrm{Tr}(\hat{A}\rho_t)$ changes with time according to the stochastic differential,

$$d\langle\hat{A}\rangle = -i\,\mathrm{Tr}(\left[\hat{A},\hat{H}\right]\rho_t)dt + \mathrm{Tr}\left(\hat{A}\hat{c}\rho_t\hat{c}^\dagger - \frac{1}{2}\left\{\hat{A},\hat{c}^\dagger\hat{c}\right\}\rho_t\right)dt + \left(\mathrm{Tr}((\hat{A}\hat{c} + \hat{c}^\dagger\hat{A})\rho_t) - \mathrm{Tr}((\hat{c}+\hat{c}^\dagger)\rho_t)\,\mathrm{Tr}(\hat{A}\rho_t)\right)dW_t, \quad \text{(A28)}$$

where $dW_t = dY_t - \mathrm{Tr}((\hat{c}+\hat{c}^\dagger)\rho_t)dt$.

We now address the similar change according to the past quantum state, i.e, the change of the weak value estimate $\langle\hat{A}\rangle_w = \mathrm{Tr}(\hat{A}\rho_\mathrm{p})$. The differential of the un-normalized past density matrix is

$$d(\rho_t E_t) = \rho_{t+dt}E_{t+dt} - \rho_t E_t = d\rho_t E_{t+dt} - \rho_t dE_{t+dt}.$$

By inserting the expressions Eq. (A24) and Eq. (3) we get

$$d(\rho_t E_t) \;=\; -i\left[\hat{H},\rho_t E_{t+dt}\right]dt \,+\, \left[\hat{c},\rho_t\hat{c}^\dagger E_{t+dt}\right]dt \,-\, \frac{1}{2}\left[\hat{c}^\dagger\hat{c},\rho_t E_{t+dt}\right]dt \,+\, \left[\hat{c},\rho_t E_{t+dt}\right]dY_t. \tag{A29}$$

Note that $\mathrm{Tr}(d(\rho_t E_t)) = 0$ as expected since $\mathrm{Tr}(\rho_t E_t) = \mathrm{Tr}(\rho_T)$ is constant. The differential of the weak value $\langle\hat{A}\rangle_{\mathrm{w}}$, $d\langle\hat{A}\rangle_{\mathrm{w}} = d\,\mathrm{Tr}(\hat{A}\rho_{\mathrm{p},t}) = \mathrm{Tr}(\hat{A}d(\rho_t E_t))/\,\mathrm{Tr}(\rho_t E_t)$ thus becomes

$$\begin{aligned} d\langle\hat{A}\rangle_{\mathrm{w}} = \frac{1}{\mathrm{Tr}(\rho_t E_t)}\Big[&-i\,\mathrm{Tr}([\hat{A},\hat{H}]\rho_t E_{t+dt})dt + \mathrm{Tr}\left(\left[\hat{A},\hat{c}\right]\rho_t\hat{c}^\dagger E_{t+dt} - \frac{1}{2}\left[\hat{A},\hat{c}^\dagger\hat{c}\right]\rho_t E_{t+dt}\right)dt \\ &+ \mathrm{Tr}([\hat{A},\hat{c}]\rho_t E_{t+dt})dY_t\Big]. \end{aligned} \tag{A30}$$

This differs from the change in the conventional mean value and, in particular, we observe the suppression of the noise term $\propto dY_t$ for observables which commute with the measurement operator $\hat{c}$. This suppression explains the smooth behavior of the black dashed curves in Figs. 2(a) and 2(b), while the conventional mean values $\langle\hat{A}\rangle$ show fluctuations due to the Wiener noise term, dW, in Eq. (A28).

B. RELATION TO CLASSICAL HIDDEN MARKOV MODELS

The classical theory of hidden Markov models is very well established and an excellent introduction can be found in Ref. 3. Here, a *hidden Markov model* is a discrete-time stochastic process where a hidden system state evolves according to a Markov chain and observations of some output signal depending on the system state are performed. Let X_t be the system state at time t and the output signal at time t be Y_t. The output signal Y_t depends only on the system state at the same time and its probability distribution is therefore completely determined by $P(Y_t|X_t)$. Since the evolution of X_t follows a Markov chain the system dynamics is completely determined by the transition probabilities $P(X_{t+1}|X_t)$. It is a simple matter to generalize these ideas to the continuous time case (such as a system governed by a rate equation), but for simplicity we will consider only the discrete time case here and the times t are integers.

The full joint probability distribution for measurements Y_t and states for a process running from time $t = 0$ to time $t = T$ is then

$$P(X_0,\ldots,X_T,Y_1,\ldots Y_T) = \prod_{t=0}^{T-1} P(X_{t+1}|X_t)\prod_{t=1}^{T} P(Y_t|X_t)P(X_0), \tag{B31}$$

where $P(X_0)$ is the probability for the initial state X_0. The theory of hidden Markov models provides numerically efficient algorithms for calculating the following quantities: (1) the forward filtered state probability distribution $P(X_t|Y_1, \ldots Y_t)$ and (2) the smoothed state probability distribution $P(X_t|Y_1, \ldots Y_T)$ for all times t.

The forward estimate is easily calculated by a standard recursive Bayesian procedure. Following the notation in [3] we define the vectors

$$\alpha_t(i) = P(Y_1, \ldots Y_t, X_t = i) \tag{B32}$$

$$\beta_t(i) = P(Y_{t+1}, \ldots Y_N | X_t = i). \tag{B33}$$

Using the α_t and β_t-vectors we can calculate the filtered and smoothed distributions at time t by the following formulas

$$P(X_t = i|Y_1, \ldots Y_t) = \frac{\alpha_t(i)}{\sum_k \alpha_t(k)}, \tag{B34}$$

$$P(X_t = i|Y_1, \ldots Y_N) = \frac{\alpha_t(i)\beta_t(i)}{\sum_k \alpha_t(k)\beta_t(k)}, \tag{B35}$$

both of which are variations of Bayes' formula. It is not difficult to show that α_t and β_t satisfy the following recursion relations

$$\alpha_{t+1}(i) = \sum_j P(Y_{t+1}|X_{t+1} = i)P(X_{t+1} = i|X_t = j)\alpha_t(j) \tag{B36}$$

$$\beta_t(i) = \sum_j P(Y_{t+1}|X_{t+1} = j)P(X_{t+1} = j|X_t = i)\beta_{t+1}(j), \tag{B37}$$

where $\beta_T(i) = 1$ and $\alpha_0(i) = P(X_0 = i)$. In the following, we show that our effect matrix is equivalent to the β_t-vector and that the past quantum state $\Xi(t)$ is the natural quantum generalization of the hidden Markov model pair (α_t, β_t). In the case of non-disturbing measurements the hidden Markov model smoothed state (B35) is equivalent to both the past quantum state Ξ and the past density matrix ρ_{p}. This is consistent with the weak value assumption of no disturbance of the system due to the measurements since classical measurements may always be thought of as non-disturbing).

Indeed, the above hidden Markov theory can be formulated using diagonal density matrices and updates of the form given in Eq. (A1). Let $|i\rangle$ be an orthogonal basis for a Hilbert space, where i denotes the same internal states as in the hidden Markov model. The Markov chain evolution $P(X_{t+1}|X_t)$ is now given by the update

$$\mathcal{C}\colon \rho \longmapsto \sum_{i,j} |j\rangle \langle j| \, P(X_{t+1} = j|X_t = i) \, \langle i|\rho|i\rangle \, , \tag{B38}$$

which is in fact an evolution of the type given in Eq. (A1) with $\hat{M}_{i,j} = \sqrt{P(X_{t+1}=j|X_t=i)}\,|j\rangle\langle i|$ The observation process is given by the update

$$\mathcal{I}\colon \rho \overset{y}{\longmapsto} \sum_i P(Y_t = y|X_t = i)\,\langle i|\rho|i\rangle\,|i\rangle\langle i| \tag{B39}$$

followed by re-normalization as in Eq. (A1).

The classical hidden Markov model is then reproduced by picking the initial state $\rho_0 = \sum_i P(X_0 = i)\,|i\rangle\langle i|$ and applying the two updates $\mathcal{C}$ and $\mathcal{I}$ in sequence.

The (un-normalized) forward filtered state α_t is simply the (un-normalized) filtered quantum state $\tilde{\rho}_t$ which satisfies the recursion relation

$$\tilde{\rho}_t \overset{\mathcal{C}}{\longmapsto} \sum_{i,j} |i\rangle\langle i|\,P(X_{t+1}=i|X_t=j)\,\langle j|\tilde{\rho}_t|j\rangle \tag{B40}$$

$$\overset{\mathcal{I}}{\longmapsto} \sum_{i,j} |i\rangle\langle i|\,P(Y_{t+1}=y|X_{t+1}=i)P(X_{t+1}=i|X_t=j)\,\langle j|\tilde{\rho}_t|j\rangle \equiv \tilde{\rho}_{t+1} \tag{B41}$$

which for diagonal $\tilde{\rho}_t$ is exactly Eq. (B36).

The effect matrix E_t is initially $E_T = \hat{I}$ which is equivalent to the initial condition $\beta_T(i) = 1$. The effect matrix is propagated according to the adjoint update Eq. (A12) as

$$E_{t+1} \overset{\mathcal{I}^\dagger}{\longmapsto} \sum_j |j\rangle\langle j|\,P(Y_{t+1}=y|X_{t+1}=j)\,\langle j|E_{t+1}|j\rangle \tag{B42}$$

$$\overset{\mathcal{C}^\dagger}{\longmapsto} \sum_{i,j} |i\rangle\langle i|\,P(Y_{t+1}=y|X_{t+1}=j)P(X_{t+1}=j|X_t=i)\,\langle j|E_{t+1}|j\rangle \equiv E_t. \tag{B43}$$

which is exactly the update formula for β_t Eq. (B37). Note here that the $\mathcal{I}$-update is unchanged, whereas the adjoint update for $\mathcal{C}$ is modified.

The theory of hidden Markov models includes numerically efficient algorithms for re-estimating the parameters occurring in the model. The so-called Baum-Welch algorithm is a special case of the Expectation-Maximization algorithm. By using the fully smoothed estimate as encoded in the α- and β-vectors, a simple formula for the re-estimated parameters exists, which leads to a more likely sequence of measurement results $Y_1, \dots Y_T$ given the model. In this way a local maximum of the likelihood can be calculated by iterating parameter re-estimation and the smoothing calculation. We believe that a similar technique can be applied to estimate unknown parameters in quantum processes using the past quantum state.

[1] R. B. Griffiths and C.-S. Niu, Phys. Rev. Lett. **76**, 3228 (1996).

[2] H. M. Wiseman and G. J. Milburn, *Quantum Measurement and Control* (Cambridge University Press, New York, 2010).

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

[4] K. Jacobs and D. A. Steck, Contemp. Phys. **47**, 279 (2006).

[5] Y. Aharonov, D. Z. Albert, and L. Vaidman, Phys. Rev. Lett. **60**, 1351 (1988).