Past observable dynamics of a continuously monitored qubit

Luis Pedro García-Pintos¹ and Justin Dressel¹,²

¹Institute for Quantum Studies, Chapman University, Orange, California 92866, USA
²Schmid College of Science and Technology, Chapman University, Orange, California 92866, USA

(Received 17 August 2017; published 8 December 2017)

Monitoring a quantum observable continuously in time produces a stochastic measurement record that noisily tracks the observable. For a classical process, such noise may be reduced to recover an average signal by minimizing the mean squared error between the noisy record and a smooth dynamical estimate. We show that for a monitored qubit, this usual procedure returns unusual results. While the record seems centered on the expectation value of the observable during causal generation, examining the collected past record reveals that it better approximates a moving-mean Gaussian stochastic process centered at a distinct (smoothed) observable estimate. We show that this shifted mean converges to the real part of a generalized weak value in the time-continuous limit without additional postselection. We verify that this smoothed estimate minimizes the mean squared error even for individual measurement realizations. We go on to show that if a second observable is weakly monitored concurrently, then that second record is consistent with the smoothed estimate of the second observable based solely on the information contained in the first observable record. Moreover, we show that such a smoothed estimate made from incomplete information can still outperform estimates made using full knowledge of the causal quantum state.

DOI: 10.1103/PhysRevA.96.062110

I. INTRODUCTION

Over the past decade, time-continuous quantum measurements [1–9] of superconducting qubits (such as transmons [10]) have become an important and increasingly well-controlled component of emerging quantum computing technology [11–32]. Indeed, the primary method for extracting information from a superconducting transmon is to dispersively couple it to a pumped microwave resonator, then amplify and mix the leaked microwave field with a local oscillator to perform a homodyne measurement of the traveling field, which produces a stochastic time-dependent voltage that encodes information about the transmon energy basis [33–35]. Understanding what information is contained in the resulting stochastic readout is thus an essential theoretical issue.

In simple terms, a continuous measurement can be understood as a sequence of weak measurements [36,37] on the qubit. In the superconducting case, each temporal segment of the steady-state traveling coherent microwave field acts as an independent and approximately Gaussian meter that becomes entangled with the qubit and later measured [35]. During the measurement of the field, the finite bandwidth of the circuitry typically discretizes the field into time bins of size $dt$. Provided that $dt$ is longer than the correlation time scale of the traveling field, the statistics of the averaged homodyne voltage collected in each independent time bin are approximately Gaussian, producing a discrete temporal sequence of Gaussian-distributed measurement results $r_j$ with a wide variance that inversely depends upon the time step size $dt$. All information about the qubit must be extracted by processing this stochastic time series.

For convenience, this time series is traditionally interpolated to construct a time-continuous stochastic process $r(t)$ that preserves the physically correct averages over the time bins $dt$. Such an interpolation then has the structure of a moving-mean stochastic white noise process. The mean of this process is widely recognized to be the expectation value of the monitored observable [38,39], following straightforward arguments about the increasing width of the Gaussian distributions in the time-continuous limit. This understanding of the mean as an expectation value raises the natural question of whether the white noise may be reduced by classical signal processing techniques to recover that expectation value via temporal averaging, rather than ensemble averaging. Indeed, such temporal averaging exposes quantum jumps between measurement eigenstates in the quantum Zeno regime [24,32,40]. More dramatically, simultaneously monitoring multiple orthogonal observables and processing the collected readouts with simple exponential filtering has been used to estimate an evolving qubit state with surprisingly high fidelity [41,42]. One might therefore suspect that optimizing the classical signal processing of the readout could allow quantum state tomography for a single realization of an evolving qubit state to high accuracy with minimal prior information, and thus challenge the operational interpretation of the quantum state as describing only ensemble statistics.

In this paper, we carefully revisit the derivation of the collected readout as a stochastic process and show the counterintuitive result that the moving mean is not in fact the expectation value of the observable, as is usually assumed. Instead, the mean is altered by the measurement backaction away from the expectation value. As such, optimally removing the noise from the readout will not recover the causal quantum state, as might be suspected, which will bound the achievable fidelity of classical filtering state tomography schemes. Instead, the moving mean tracks a smoothed estimate that we show converges to the real part of a generalized quantum weak value [37,43–45] in the time-continuous limit. Notably, this weak value naturally appears without additional postselection for each individual measurement realization. We derive this result analytically, then verify numerically that for individual trajectories the mean squared error of this smoothed estimate is consistently smaller than that of the expectation value. This minimization of the mean squared error is consistent with the usual metric of classical signal processing for determining...
the optimal estimate of a time-dependent noisy signal. We independently verify the result numerically using a hypothesis testing approach, confirming that the smoothed estimate is indeed a better fit to the collected data than the expectation value. We go on to show that in the presence of a simultaneous second observer, the smoothed estimate retains its objective character. That is, a smoothed estimate made from incomplete data taken only by the first observer can be a better fit to the unknown data of the second observer than even the pure causal qubit state that uses all available data. Notably, this last result improves upon a recent proposal [46] that constructs a “smoothed quantum state” to estimate the observations made by an unknown second observer, since that method can never outperform the most pure causal state that uses all collected data. The conclusions of our study are consistent with prior work concerned with time-symmetric quantum state estimates, such as the two-state-vector formalism [37,47,48], quantum smoothing [49–52], bidirectional quantum states [53], and past quantum states [54–57]. However, we emphasize here the practical consequence for ongoing research into continuous past quantum states [54–57]. However, we emphasize here the objective even in the presence of a second simultaneous observer, thereby reinforcing their operational relevance. We conclude in Sec. VI.

II. OBSERVABLE DYNAMICS
FROM ANTERIOR MEASUREMENTS

We focus our discussion on what can be inferred from a collected measurement record about its associated observable dynamics. As such, in what follows we consider a simplified model of time-continuous measurements that is adequate for isolating the relevant features. To keep this paper self-contained, we briefly review the essential details of how a temporal sequence of independent Gaussian measurements models continuous-in-time measurements from a quantum information perspective. This model is a slight idealization of those that describe recent experimental work on quantum state trajectories well [16,18,25–27,32], but deliberately neglects relevant experimental details—such as measurement inefficiency, environmental decoherence, energy relaxation, phase backaction, and non-Markovian effects from finite detector bandwidth [34,35]—in order to isolate the essential effect of the measurement backaction. (For an alternative recent derivation of a simple continuous measurement model that includes some of these nonidealities in the context of feedback, see also Ref. [58].)

Consider a system, such as a qubit, that is assigned a quantum state represented by a density operator $\rho$. To measure an observable $A = \sum a |a⟩⟨a|$ of the system, such as the Pauli operator $σ_z$, we couple the system to a measurement device, which reports classical results $r \in \mathbb{R}$ that are correlated with the distinct values of $A$. That is, each distinct value $a$ of $A$ corresponds to a probability measure $P(r|a)dr$ for obtaining $r$ on the detector given that particular $a$, such that each measure is normalized over the possible results of the measurement, $\int_{\mathbb{R}} P(r|a)dr = 1$, and the total probability for obtaining a measurable subset $R \subseteq \mathbb{R}$ of $r$ is $\int_{R} P(r|\rho)dr = \int_{R} \sum_{a} P(r|a)P(a|\rho)dr$ with $P(a|\rho) = |a⟩⟨a|\rho|a⟩$. As a convenient way to formally encapsulate these detector properties, the map $a \rightarrow P(r|a)dr$ from observable values to probability measures generates a map $A \rightarrow E_r dr = \int_{\mathbb{R}} P(r|a)P(a|\rho)dr$ from the observable operator $A$ to a probability operator-valued measure (POVM) $E_r dr$. The positive Hermitian operators $E_r$ of the POVM then partition unity $\int_{\mathbb{R}} E_r dr = 1$, and obey Lüder’s probability rule [59], $P(r|\rho)dr = Tr[E_r|\rho⟩⟨\rho|]dr$.

Such a generalized measurement of $r$ on the detector produces measurement backaction on the state $\rho$ described by a quantum instrument, $\rho \rightarrow E_r(\rho)dr$, which is a completely positive map-valued measure satisfying $P(r|\rho)dr = Tr[E_r(\rho)dr]$ [53]. In the case of no classical mixing from information loss, this instrument can be represented by a single $r$-dependent Kraus operator $M_r$, according to $E_r(\rho)dr = M_rρM_r^†dr$, which relates $M_r$ to the POVM $E_r$ according to $E_r dr = M_r^†M_r dr$. As such, the Kraus operator factors into a polar form $M_r = Ut/Er$, with $\sqrt{Er} = \sum a \sqrt{P(r|a)}|a⟩⟨a|$ corresponding to partial state collapse (informational backaction) and $Ut$ corresponding to additional $r$-dependent unitary perturbation (stochastic Hamiltonian backaction). In what follows, we neglect any unitary backaction $Ut$ for simplicity to focus solely on the effects of the informational backaction. (See Refs. [19,34,35] for discussion about the role of such unitary phase backaction in measurements of superconducting qubits.)

The renormalized state after the observation of a subset $R$ of $r$ values on the detector (e.g., from classically coarse-grained resolution) is

$$\rho \rightarrow \int_{R} \frac{E_r(\rho)dr}{\int_{\mathbb{R}} P(r|\rho)dr} = \int_{R} \frac{M_rρM_r^†dr}{\int_{\mathbb{R}} P(r|\rho)dr}.$$  

(1)

We will restrict our discussion to a perfect detector with infinitely sharp resolution of individual points $r$ for simplicity, so that the measure factors in Eq. (1) simply cancel to yield the simplified expression

$$\rho \rightarrow \frac{E_r(\rho)}{P(r|\rho)} = \frac{M_rρM_r^†}{P(r|\rho)}.$$  

(2)

In the following, we assume Gaussian measurements of $A = σ_z = |1⟩⟨1| - |0⟩⟨0|$ describing the computational basis of a qubit. That is, the detector distributions for the distinct values $\pm 1$ of $σ_z$ are Gaussian, $P(r|\pm 1) = G_{\pm 1}(r) = \exp(-r^2/2\tau^2)/\sqrt{2\pi}\tau dr$, with equal variances $\tau^2$ but distinct means centered at their associated values of $\pm 1$. 062110-2
This parameterization is chosen so that $dt$ is a discretization time scale that specifies the duration of the coupling required to obtain the result $r$, and $\tau$ is a measurement collapse time scale that indicates the coupling duration needed to obtain a unit signal-to-noise ratio for the measurement. This variance scaling also guarantees that sequences of independent such measurements correctly average to coarsen the discretization time scale, i.e., $\text{Var}(r_j + r_{j+1})/2 = \text{Var}(r_j) + \text{Var}(r_{j+1})/4 = (\tau/dt + \tau/dt)/4 = \tau/(2dt)$, which will later permit a sensible continuum limit as $dt \to 0$ to yield a Markovian stochastic process [38]. The simplest Kraus operator for such a Gaussian measurement is $M_r = \sqrt{P(r)}|1\rangle\langle 1| + \sqrt{P(r)}|0\rangle\langle 0|$, which may be written in a more compact form as a function of the operator $\sigma_z$,

$$M_r = \left(\frac{dt}{2\pi \tau}\right)^{1/4} \exp\left[\frac{(r - \sigma_z)^2 dt}{4\tau}\right].$$

Despite its simplicity, this Gaussian model is a reasonable approximation for a variety of experimental situations, including double-quantum-dot measurements with a quantum point contact [9], and superconducting transmon measurements with microwave resonators [34].

Notably, the probability distribution $P(r|\rho)$ for causally obtaining a future $r$ from the current state $\rho$ may be conveniently expanded in terms of the single expectation value $z = \text{Tr}[\rho \sigma_z]$ as

$$P(r|\rho) = P(+1|\rho)G_z(r) + P(-1|\rho)G_{-z}(r) = \frac{1 + z}{2} G_z(r) + \frac{1 - z}{2} G_{-z}(r),$$

which allows all moments of $r$ to be easily calculated. For example, the first three moments are

$$\langle r \rangle = z, \quad \langle r^2 \rangle = 1 + \frac{\tau}{dt}, \quad \langle r^3 \rangle = \left(1 + \frac{3\tau}{dt}\right)z.$$

All such moments for future $r$ are characterized solely by the expectation value $z$ of the measured observable in the qubit state $\rho$ immediately prior to the measurement. We will see in the next section that this feature will no longer be true for moments of past $r$.

Let us now consider a sequence of $N$ such generalized measurements $M_r$, with outcomes $r_j$, with $j = 1, \ldots, N$. Between each measurement, the qubit independently evolves for the time step $dt$ with Hamiltonian $H$, which we model by a separate unitary operator $U \equiv \exp(-iHdt/\hbar)$. The state of the qubit at the time $T = N dt$, given an initial state $\rho$ at time $t = 0$ and the past set of outcomes $\tilde{r} = (r_1, \ldots, r_N)$, is then

$$\rho_{\tilde{r}} = \frac{(M_{r_N}U \cdots M_{r_1}U)(U^\dagger M_{r_N}^\dagger \cdots U^\dagger M_{r_1}^\dagger)}{P(\tilde{r} | \rho)},$$

where the joint probability $P(\tilde{r} | \rho) = \text{Tr}[E_{\tilde{r}} \rho]$ of all measured results is governed by the positive operator from a joint POVM,

$$E_{\tilde{r}} = (U^\dagger M_{r_N}^\dagger \cdots U^\dagger M_{r_1}^\dagger)(M_{r_N}U \cdots M_{r_1}U).$$

This model describes the periodic monitoring of the observable $\sigma_z$ at the times $t_j = j dt$. In the continuum limit as $dt \to 0$ and $N \to \infty$, keeping $T = N dt$ constant, the unitary and measurement operators will commute up to second order in $dt$ such that each pair of operators $M_{r_j} U$ effectively describes the evolution within the same time step $[t_j, t_j + dt)$, and the evolution in Eq. (6) becomes equivalent to a stochastic master equation [33,38,39,58] that describes truly continuous-in-time observable monitoring. In what follows, however, we retain the explicitly discrete time steps $dt$ for numerical stability and conceptual clarity. The discrete model has the added benefit of also modeling physically discrete sequences of impulsive Gaussian measurements [56,60].

When the continuum limit $dt \to 0$ is taken, the widths of the Gaussian distributions $P(r|\rho)$ broaden and mostly overlap, so the distribution $P(r|\rho)$ in (4) approximates a single Gaussian distribution centered at the expectation value $z(t) = \text{Tr}[\rho \sigma_z]$:

$$P(r(t)|\rho_t) \approx G_{z(t)}(r(t)) = \sqrt{\frac{dt}{2\pi \tau}} \exp\left[\frac{-(r(t) - z(t))^2 dt}{2\tau}\right],$$

where we have replaced the discrete index $j$ with the continuous time $t$. It follows that in this limit, the future (uncollected) readout can itself be approximated as a moving-mean Markovian stochastic process centered at the evolving expectation value $z(t)$,

$$r(t) = z(t) + \sqrt{\tau} \xi(t).$$

where $\xi(t)$ is zero-mean additive white noise [38,39], satisfying $\langle \xi(t)\xi(t') \rangle = \delta(t - t')$. This understanding of the readout as an expectation value cloaked by additive noise is standard in the literature of continuous quantum measurements [5,33,39,41] and has been applied with tremendous success in a variety of experiments [16,18,29].

Note that the white noise expression in Eq. (9) seems to give a simple prescription for how to learn information about the qubit evolution solely from the readout $r(t)$. For example, classical signal processing methods can reduce the zero-mean noise $\xi(t)$ and thus approximately recover the dynamics of the observable expectation value $z(t)$. This feature has been demonstrated for the observation of quantum jumps between $z = \pm 1$ [24,40]. Moreover, by concurrently monitoring the three qubit Pauli operators $\sigma_x$, $\sigma_y$, and $\sigma_z$ and applying exponential filtering to the collected readouts, the dynamics of all three expectation values $x(t)$, $y(t)$, and $z(t)$ that determine the evolving qubit state $\rho(t)$ may be recovered simultaneously with reasonably high fidelity for individual measurement realizations [41]. This latter result is particularly startling since it seems to challenge an interpretation of the quantum state $\rho$ as pertaining solely to an ensemble of realizations.

The relation in Eq. (9) is misleading, however, since it pertains only to an as-yet-uncollected future readout and does not yet describe the temporal structure of a readout that was collected in the past. Due to the informational backaction of the measurement, the readout and state evolution become temporally correlated, which effectively refines the distribution in Eq. (8) of the past readout and shifts the mean of Eq. (9). Strictly speaking, the relation in Eq. (9) only holds at the final collected time of the readout, which still has an uncertain future.
These two quantities generalize the pure "time-symmetric state" (|ψ⟩, ⟨φ|) from the "two-vector formalism" pioneered in [37,47,48].

Applying Bayes’ rule to the joint distribution \( P(\vec{r}_p, r, \vec{r}_f | ρ) \) yields the desired distribution,

\[
P(ρ | \vec{r}_p, r, \vec{r}_f) = \frac{P(\vec{r}_p, r, \vec{r}_f | ρ)}{\int_{\mathbb{R}} P(\vec{r}_p, r, \vec{r}_f | ρ) dr}.
\]  \( (12) \)

From the preceding section, the joint distribution is

\[
P(\vec{r}_p, r, \vec{r}_f | ρ) = \text{Tr} [E_{\vec{r}_f} M_r ρ M_r †].
\]  \( (13) \)

Combining Eqs. (3), (12), and (13) thus permits explicit calculation of the desired distribution,

\[
P(ρ | \vec{r}_p, r, \vec{r}_f) = \frac{(G_1(r) - G_{−1}(r))z_w}{(1 + e^{−2π dt})} + \frac{(G_1(r) + G_{−1}(r) + 2e^{−2π dt} G_0(r))}{(1 + e^{−2π dt})}z_e + \frac{(G_1(r) + G_{−1}(r) - 2e^{−2π dt} G_0(r))}{(1 + e^{−2π dt})}z_e
\]  \( (14) \)

which depends on only two quantities containing the monitored observable,

\[
z_w = \text{Re} \left[ \frac{\text{Tr} [E_{\vec{r}_f} σ_z ρ]}{\text{Tr} [E_{\vec{r}_f} ρ]} \right], \quad z_e = \frac{\text{Tr} [E_{\vec{r}_f} σ_z ρ]}{\text{Tr} [E_{\vec{r}_f} ρ]},
\]  \( (15) \)

neither of which are an expectation value. Instead, \( z_w \) is the real part of a (generalized) weak value [37,43–45] in its role as a first-order conditioned expectation value [44,45] and \( z_e \) is a second-order contribution [61]. Due to the assumption of Gaussian statistics, these first two orders are sufficient to fully characterize the distribution.

For comparison with Eq. (5), the first three moments of the distribution \( P(ρ | \vec{r}_p, r, \vec{r}_f) \) are

\[
⟨r⟩_S = z_S,
\]  \( (16a) \)

\[
\langle r^2⟩_S = \frac{τ}{dt} + \frac{1}{2} (1 + z_c),
\]  \( (16b) \)

\[
\langle r^3⟩_S = \left( 1 + 3\frac{τ}{dt} \right) z_S.
\]  \( (16c) \)

Remarkably, after taking into account subsequent measurement outcomes, the mean of the intermediate \( r \) shifts to a refined (smoothed) estimate \( z_S \) instead of the traditionally accepted expectation value \( z \) that we obtained in the previous section [compare with Eq. (5)]. This smoothed estimate of \( σ_z \) is the central quantity of this paper,

\[
z_S ≡ \frac{z_w}{2} + \frac{z_e}{2} (1 + e^{−π dt}) + \frac{z_e}{2} (1 - e^{−π dt})z_e,
\]  \( (17) \)

and depends upon both \( z_w \) and \( z_e \).

Note that for weak measurements with large variance, \( τ/dt ≫ 1 \), the second-order contribution \( z_e \) is suppressed and the smoothed estimate converges to the first-order weak value state”.
The smoothed estimate therefore inherits the behavior of the weak value, to a degree that depends on the coarseness of the time steps. For example, the smoothed estimate can take values outside the range \([-1, 1]\) of possible observable values for \(\sigma_z\) [60,62,63], as shown in the next section. Importantly, this convergence to the weak value becomes exact in the continuum limit as \(dt \to 0\). Moreover, the continuum limit of the full distribution in Eq. (14) is a single Gaussian distribution similar to Eq. (8) but centered on the smoothed estimate \(z_S\) of \(\sigma_z\), which in turn converges to \(z_w\):

\[
P(r(t)|\vec{r}_f, \rho) \approx G_{z,w}(r(t)) = \sqrt{\frac{dt}{2\pi\tau}} \exp\left\{-\frac{(r(t) - z_w(t))^2dt}{2\tau}\right\},
\]

(18)

Therefore, arguments identical to the preceding section imply that the past (already collected) continuous readout still has the structure of a moving-mean stochastic process, but instead following a weak value of \(\sigma_z\),

\[
r(t) = z_w(t) + \sqrt{\tau} \xi(t).
\]

(19)

This key result implies that using classical signal processing techniques to reduce the zero-mean white noise \(\xi(t)\) on a collected readout \(r(t)\) will not recover the expectation value \(z(t)\) as might be expected from the previous section. Instead, such techniques will recover the smoothed (weak-valued) estimate \(z_S(t) \rightarrow z_w(t)\) that properly takes into account the temporal correlations in the signal caused by the measurement backaction. This discrepancy explains the limited fidelity of the state reconstruction seen in [41] when exponentially filtering simultaneous observable readouts. The reasonably high fidelities that were still obtained are explained by the fact that the smoothed estimate \(z_w(t)\) can often remain close to the expectation value, as shown in the next section.

IV. A BETTER DESCRIPTION OF PAST OBSERVABLE DYNAMICS

The previous section established that a smoothed observable estimate more closely describes the observed readout than an expectation value. In this section, we numerically simulate an explicit example of a monitored qubit Rabi oscillation to demonstrate the practical significance of this result. Specifically, we define two figures of merit that contrast an expectation value with a smoothed estimate and show that the smoothed estimate systematically outperforms the expectation value. Importantly, we consider individual measurement realizations, not ensemble averages.

Consider the periodic monitoring of \(\sigma_z\) at time steps \(dt\) for a total duration \(T\), with characteristic collapse time scale \(\tau\), on a qubit driven by a Hamiltonian

\[
H = \hbar \Omega \frac{\sigma_x}{2},
\]

(20)

where the Pauli matrix \(\sigma_x\) generates Rabi oscillations in the \(x\)-\(z\) plane and \(T_R \equiv 2\pi/\Omega\) is the period of these oscillations. Figure 2 illustrates the distinction between the expectation value \(z\) (blue solid line), the smoothed value \(z_S\) (black dotted line), and the weak value \(z_w\) (red hashed line) for such a monitored oscillation. In both plots, \(dt/T_R = 1/20\) is held constant, with the upper plot showing a weaker measurement with \(t/T_R = 2\) and the bottom plot showing a stronger measurement with \(t/T_R = 1/10\). The weaker measurement in the upper plot exhibits noisy Rabi oscillations since the dynamics are only weakly perturbed by the monitoring. For this upper plot, \(dt/\tau = (1/20)/2 = 1/40\), so the smoothed estimate \(z_S\) and weak value \(z_w\) are essentially indistinguishable. The stronger measurement in the lower plot exhibits quantum jumps between measurement eigenstates, to which the qubit is pinned by the quantum Zeno effect [24,40]. For this lower plot, \(dt/\tau = (1/20)/(1/10) = 1/2\), so the smoothed estimate \(z_S\) is visibly distinct from the weak value \(z_w\) when the latter becomes sufficiently large. Note that the weak value \(z_w\) can exceed the eigenvalue range of \([-1, 1]\) when the readout \(r(t)\) is statistically unlikely.

To answer the question of whether the expectation value \(z\) or the smoothed value \(z_S\) better follows a single readout realization \(r\) quantitatively, we establish two figures of merit. First, we consider the mean squared error between the \(N\)-dimensional readout vector \(\vec{r}\) for all \(N\) time steps \(dt\), and the dynamical estimate vectors \(\vec{z}\) and \(\vec{z}_S\). The mean squared error is defined for any two vectors \(\vec{v}\) and \(\vec{w}\) of length \(N\) as

\[
\text{MSE}(\vec{v}, \vec{w}) = \sum_{j=1}^{N} (v_j - w_j)^2/N.
\]

Notably, the mean squared error is the primary figure of merit used in classical filtering and estimation theory [64] to find optimal estimates for the “true” value of a noise-polluted signal. We treat the readout \(\vec{r}\) as such a noise-polluted signal, and define the relative mean
squared error,
\[ Q(z_S, z) = \frac{\text{MSE}(\tilde{r}, \tilde{z}) - \text{MSE}(\tilde{r}, \tilde{z}_S)}{\text{MSE}(\tilde{r}, \tilde{z})}, \]

(21)
such that \( Q(z_S, z) > 0 \) if and only if the smoothed estimate \( \tilde{z}_S \) is a better fit to the measured readout \( \tilde{r} \) than the expectation value \( \bar{z} \).

The second figure of merit we adopt is a hypothesis test. Namely, let us assume prior probabilities \( P(z) \) and \( P(z_S) \) for models in which the readout \( \tilde{r} \) fits \( z \) or \( z_S \), respectively. Bayes’ rule allows us to express the probability of the estimate \( \tilde{z} \) given the observed measurement record \( \tilde{r} \) as
\[ P(\tilde{z} | \tilde{r}) = \frac{P(\tilde{r} | \tilde{z})P(z)}{P(\tilde{r})} \]
Similarly, \( P(\tilde{z}_S | \tilde{r}) = \frac{P(\tilde{r} | \tilde{z}_S)P(z_S)}{P(\tilde{r})} \).
Assuming equal prior probabilities for both hypotheses, \( P(z) = P(z_S) \), we can then define the hypothesis test ratio as
\[ R(z_S, z) = \frac{P(\tilde{z}_S | \tilde{r})}{P(\tilde{z} | \tilde{r})} = \frac{P(\tilde{r} | \tilde{z}_S)P(z_S)}{P(\tilde{r} | \tilde{z})P(z)}, \]
where \( P(\tilde{r} | \tilde{z}) = \prod_j P(r_j | \rho_{ij}^z) \) and \( P(\tilde{r} | \tilde{z}_S) = \prod_j P(r_j | \rho_{ij}^{z_S}, E_{j}) \) can be calculated from Eqs. (4) and (14).

The ratio \( R \) discriminates the likelihood of the estimates \( z \) or \( z_S \), given the observed record \( \tilde{r} \). We use its natural logarithm as a figure of merit to decide between the two alternatives: that is, \( \ln R > 0 \) if and only if the smoothed estimate is more probable than the expectation value.

Figure 3 shows histograms of both the relative mean squared error \( Q(z_S, z) \) (top row) and the hypothesis test logarithmic ratio \( \ln[R(z_S, z)] \) (bottom row), computed for \( 10^5 \) realizations using fixed time steps of \( dt/T_R = 1/100 \) to approximate the continuum limit in two regimes: a weaker regime with \( \tau/T_R = 2 \) (left column) and a stronger regime with \( \tau/T_R = 1/10 \) (right column). For both figures of merit, positive values indicate that \( z_S \) is a better estimate than \( z \). The ratio of realizations for which \( Q(z_S, z) > 0 \) (i.e., \( \ln [R(z_S, z)] > 0 \)) for the longer trajectory \( T/T_R = 5 \) (blue circles) is higher than for the shorter trajectory \( T/T_R = 1/10 \) (red hashed line). For both figures of merit, positive values indicate that \( z_S \) is a better estimate than \( z \).

FIG. 3. Estimate comparisons. Normalized histograms comparing the expectation value \( z \) to the smoothed value \( z_S \), for \( 10^5 \) realizations monitoring \( \sigma \) of a qubit oscillating with Rabi period \( T_R \) for a duration \( T \), with time steps \( dt = T_R/100 \). Top row: Relative mean squared error \( Q(z_S, z) \); see Eq. (21). Bottom row: Hypothesis test logarithmic ratio \( \ln [R(z_S, z)] \) [see Eq. (22)] scaled by \( 2dt/T \) to show correspondence to \( Q(z_S, z) \).

Left column: Weaker measurement regime with \( \tau/T_R = 2 \). Right column: Stronger measurement regime with \( \tau/T_R = 1/10 \). For both figures of merit, positive values indicate that \( z_S \) is a better estimate than \( z \). The improvement in the estimate is approximately linear in the inverse collapse time scale \( \tau^{-1} \) since the \( T/\tau \) determines the signal-to-noise ratio of the stochastic readout. The factor of 20 improvement in performance expected between the weaker and stronger regimes is confirmed by the shift in mean of the histograms in Fig. 3—note that the means do not shift with the duration \( T \) since the plotted discriminators are effectively normalized per unit \( T \). These results confirm that one should consider the collected readout to better follow the smooth estimate \( z_S(t) \approx z_w(t) \) given in Eq. (15), nor the expectation value as naively expected from Eq. (9).

Observe that in Fig. 3, we have scaled the hypothesis test logarithmic ratio \( \ln[R(z_S, z)] \) by a factor \( (2dt/T) \) to make its correspondence to the relative mean squared error \( Q(z_S, z) \) evident and to serve as a consistency check for the simulations. This correspondence may be explained by noting that in the time-continuous limit, the probabilities \( P(\tilde{r} | \tilde{z}) = \prod_j P(r_j | \rho_{ij}^z) \) and \( P(\tilde{r} | \tilde{z}_S) = \prod_j P(r_j | \rho_{ij}^{z_S}, E_j) \) may be approximated by products of Gaussian distributions, as in Eqs. (8) and (18), respectively. It follows that the hypothesis test simplifies,
\[ \frac{2dt}{T} \ln [R(z_S, z)] \approx \frac{dt \text{MSE}(\tilde{r}, \tilde{z}_S)}{\tau} Q(z_S, z) \approx Q(z_S, z), \]
(23)
where in the last step we have used that in the continuum limit \( \text{MSE}(\tilde{r}, \tilde{z}_S) = \sum_j (r_j - \tilde{z}_S, j)^2/N \approx \tau \sum_j (\xi_j)^2/N = \tau/dt \) from Eq. (18) and the fact that the white noise \( \xi \) at any time step has variance \( 1/dt \). This relationship between \( Q(z_S, z) \) and \( \ln [R(z_S, z)] \) in the time-continuous limit is correctly confirmed in Fig. 3. Importantly, this numerical equivalence between the two \textit{a priori} distinct figures of merit confirms the white noise relation in Eq. (19), and thus that \( z_S \approx z_w \) is in fact the minimum mean squared error estimate for individual realizations of the readout \( r(t) \).

V. SMOOTHED ESTIMATES BY AN IGNORANT THIRD PARTY

Crucially, the smoothed observable estimate derived in the preceding sections is not merely an artificial best fit to a past record, but is also a \textit{predictive} quantity with operational meaning that extends beyond the original collected record. To
see this, we now consider a situation where two observers monitor different observables on the same system. The task at hand will be for the first observer to estimate what was measured by the second observer. For this task, we now show that a smoothed estimate using partial information is not only operationally better than an expectation value that uses partial information, but can even be better than an expectation value that uses all available information.

For specificity, consider an agent $Z$ who monitors $\sigma_z$ on a qubit, as described in the previous sections, while a second agent $X$ simultaneously monitors the distinct observable $\sigma_x$, in a similar way (as considered in Refs. [25,41,42,65]). We assume characteristic collapse time scales $\tau_z$ and $\tau_x$ for Gaussian Kraus operators $M_z$ and $N_x$ measuring $\sigma_z$ and $\sigma_x$, respectively, with $\tau_z < \tau_x$ so that the agent $Z$ causes the majority of the measurement backaction. After both $X$ and $Z$ monitor for a duration $T = N \Delta t$, they each possess one measurement record, $\vec{r}_z$ or $\vec{r}_x$. We now consider two distinct scenarios: (A) An omniscient third agent $O$ examines both measurement records and estimates both $\sigma_z$ and $\sigma_x$ using all information, and (B) the agent $Z$ uses only the record $\vec{r}_z$ to estimate $\sigma_z$ without knowledge of what agent $X$ actually measured.

For the omniscient observer $O$ in scenario (A), the access to both sets of outcomes allows the derivation of smoothed estimates $\vec{r}_z$ and $\vec{r}_x$ precisely as in Sec. III from the joint probability of obtaining $r_z$ and $r_x$ conditioned on both past and future outcomes, $P(r_x, r_z | \vec{r}_x, \vec{r}_z, \vec{r}_x, \vec{r}_z, \rho)$. The form of the smoothed estimates is as is in Eqs. (17) and (15):

$$x_w \equiv \frac{1}{N} \sum_{\vec{r}_x, \vec{r}_z} P(\vec{r}_x, \vec{r}_z | \vec{r}_x, \vec{r}_z) \left[ \frac{\mathcal{E}_{\vec{r}_x, \vec{r}_z} \sigma_z \rho_{\vec{r}_x, \vec{r}_z}}{\mathcal{E}_{\vec{r}_x, \vec{r}_z}} \right],$$

$$x_z \equiv \frac{1}{N} \sum_{\vec{r}_x, \vec{r}_z} P(\vec{r}_x, \vec{r}_z | \vec{r}_x, \vec{r}_z) \left[ \frac{\mathcal{E}_{\vec{r}_x, \vec{r}_z} \sigma_x \rho_{\vec{r}_x, \vec{r}_z}}{\mathcal{E}_{\vec{r}_x, \vec{r}_z}} \right] \frac{x_w}{\frac{1}{2}(1 + e^{\Delta t/2 \tau_z}) + \frac{1}{2}(1 - e^{\Delta t/2 \tau_z}) x_z},$$

but with a modified bidirectional state consisting of

$$\rho_{\vec{r}_x, \vec{r}_z} \equiv \frac{(U M_{\vec{r}_x, \vec{r}_z} U \ldots M_{\vec{r}_x, \vec{r}_z} U) \rho(\ldots)}{P(\vec{r}_x, \vec{r}_z, \vec{r}_x, \vec{r}_z | \rho)}$$

and

$$\mathcal{E}_{\vec{r}_x, \vec{r}_z} \equiv \left( U^\dagger M_{\vec{r}_x, \vec{r}_z}^\dagger \ldots U^\dagger M_{\vec{r}_x, \vec{r}_z}^\dagger \ldots U^\dagger M_{\vec{r}_x, \vec{r}_z}^\dagger \ldots (\ldots) \right).$$

Note that our model here interleaves the measurements of $\sigma_y$ and $\sigma_z$ for simplicity; in the continuum limit, $\Delta t \to 0$, the measurements become effectively simultaneous [25,42]. As in the last section, the smoothed estimates obtained in this way fit the measurement output better than the expectation values

$$x \equiv \text{Tr} \left[ \rho_{\vec{r}_x, \vec{r}_z} \sigma_z \right], \quad z \equiv \text{Tr} \left[ \rho_{\vec{r}_x, \vec{r}_z} \sigma_z \right],$$

obtained from the most informationally complete causal state of the qubit.

The more interesting case is scenario (B), where agent $Z$ has incomplete information from which to construct an estimate. Let $\vec{x}_Z$ be the smoothed estimate of $\sigma_z$ defined in an analogous way to Eqs. (24), but based on the bidirectional state known to $Z$, which takes into account only the measurement collapses from the monitoring of $\sigma_z$.

$$\rho_{\vec{r}_z} \equiv \frac{(U M_{\vec{r}_z, \vec{r}_z} U \ldots M_{\vec{r}_z, \vec{r}_z} U) \rho(\ldots)}{P(\vec{r}_z, \rho)},$$

$$\mathcal{E}_{\vec{r}_z} \equiv \left( U^\dagger M_{\vec{r}_z, \vec{r}_z}^\dagger \ldots U^\dagger M_{\vec{r}_z, \vec{r}_z}^\dagger \ldots U^\dagger M_{\vec{r}_z, \vec{r}_z}^\dagger \ldots (\ldots) \right).$$

Similarly, let $x^2$ be the expectation value of $\sigma_z$ based on the causal state $\rho_{\vec{r}_z, \vec{r}_z}$ known to $Z$. How good are the two ignorant estimates $\vec{x}_Z$ and $x^\perp_Z$ compared to those made by the omniscient observer $O$?

Using the omniscient expectation value $x$ as a reference, a fixed long duration $T / T_R = 100$, fixed time steps $dt / T_R = 1 / 100$, and fixed collapse time scale $\tau_x / T_R = 1 / 10$, Fig. 4 shows the relative mean squared error for the ignorant expectation value $Q(x, z)(x)$ (red solid line), the ignorant smoothed estimate $Q(x, z_x)$ (black circles), and the omniscient smoothed estimate $Q(x, z_x)$ (blue starred line). This latter quantity shows the maximum improvement for reference, with ~99% of realizations consistently favoring the omniscient smoothed estimate. Figure 4(a) shows the case when the monitoring of $X$ is substantially weaker than $Z$, $\tau_x / \tau_z = 25$, so perturbs the evolution less in comparison. Figure 4(b) shows slightly less weak monitoring by $X$, $\tau_x / \tau_z = 15$. Figure 4(c) shows the fraction of cases that are better than the omniscient expectation value $x$ (i.e., where $Q > 0$) as the ratio between monitoring strengths $\tau_x / \tau_z$ varies. Unsurprisingly, the ignorant expectation value $x^2$ is always worse on average than the omniscient expectation value $x$ because of the loss of information. Surprisingly, however, the ignorant smoothed estimate $x^\perp_Z$ can outperform the omniscient expectation value for predicting what the agent $X$ actually measured when $\tau_x / \tau_z \gg 1$. In Fig. 4(a), more than 72% of the realizations favor the ignorant smoothed estimate, even though the monitoring of $\sigma_z$ produces backaction that is not negligible, with $T / \tau_x = 40$. In Fig. 4(b), more than 54% of the realizations favor the ignorant smoothed estimate, showing a reduction in advantage as the sources of backaction become comparable. In the opposite limit as $\tau_x / \tau_z \to \infty$, the monitoring by $X$ no longer perturbs the system and the ignorant estimates converge to the omniscient estimates, $x^\perp_Z \to x$ and $x^2 \to x_Z$.

To emphasize the significance of this result, we note that a similar situation to scenario (B) has been discussed by Guevara and Wiseman [46], who conclude that using an entire collected measurement record allows one to construct a “smoothed state” $\rho_s$ that is closer in fidelity to the (typically pure) maximally informative state $\rho_{\vec{r}_z, \vec{r}_z}$ known to an omniscient observer $O$ than the (mixed) state $\rho_{\vec{r}_z, \vec{r}_z}$ constructed from the incomplete information known to $Z$. The surprising extension to this result that we show here is that the informationally incomplete smoothed estimate $x^\perp_Z$ can outperform even the best expectation value $x$ known to the omniscient observer $O$. Evidently, the omniscient state $\rho_{\vec{r}_z, \vec{r}_z}$ is not informationally complete when it comes to the content of the collected readout.

VI. DISCUSSION

We have shown that, contrary to traditional wisdom, the collected readout of a continuous quantum measurement is not
centered on the expectation value of the monitored observable. Instead, the readout is centered on a modified moving mean, which is a smoothed observable estimate that converges to the real part of a generalized weak value in the time-continuous limit. The physical reason that the accumulated data follows this smoothed estimate, as opposed to the expectation value used to causally generate the same data, is that the partial measurement collapses create nontrivial correlations between past and future measurements that are only exposed in retrospect. The smoothed observable estimate provides an objectively better description of the readout than what can be accounted for solely from knowledge of the causal state of the qubit. Notably, this correspondence applies to single measurement realizations, without the need for ensemble averages and without the need for additional postselection. Importantly, this result implies that applying classical signal processing techniques to the measurement output will not reveal information about the causal state of the system, but rather information about the smoothed estimate of the monitored observable, which bounds the fidelity of any state tomography scheme based on classical signal processing of the readout.

We have also shown that the smoothed estimate from the readout has operational meaning beyond the scope of a single measured observable. That is, an agent with access only to their own measurement record can still construct a meaningful smoothed estimate for a second observable being concurrently monitored by a second agent. Provided the second measurement is sufficiently weak compared to the first measurement, this informationally incomplete smoothed estimate will still be more consistent with the experimental output of the second agent than any quantity derived from the informationally complete causal qubit state. This observed improvement over the best causal quantum state estimate could have interesting applications for experimental parameter estimation and model verification in situations where one only has partial access or incomplete information about a system, which remains to be investigated.

ACKNOWLEDGMENTS

We thank Alexander Korotkov and Howard Wiseman for helpful discussions. This work was supported by US Army Research Office Grant No. W911NF-15-1-0496. We also acknowledge partial support by the Perimeter Institute for Theoretical Physics. Research at the Perimeter Institute is supported by the Government of Canada through Industry Canada and by the Province of Ontario through the Ministry of Economic Development and Innovation.


FIG. 4. Informationally incomplete estimate comparisons. Agents $Z$ and $X$ monitor $\sigma_z$ and $\sigma_x$, respectively, of a qubit oscillating with Rabi period $T_R$ for a duration $T/T_R = 100$, with fixed time steps $dt/T_R = 1/100$ and collapse time scale $\tau_z/T_R = 1/10$. (a), (b) Normalized histograms with $10^4$ realizations, comparing the mean squared errors $Q$ relative to the reference expectation value $x$ that uses all information contained in the causal quantum state evolving from both measurement records $r_z$ and $r_x$. The ignorant expectation value $x^S$ (red solid line) and ignorant smoothed estimate $x^S$ (black circles) use incomplete information contained only in the record $r_z$, while the optimal smoothed estimate $x^S$ (blue starred line) also uses the information in $r_x$. (a) For weaker $X$ monitoring, with $\tau_x/\tau_z = 25$, the ignorant smoothed estimate $x^S$ still beats $x$ in more than 72% of the realizations, while the best smoothed estimate $x^S$ beats $x$ in more than 98% of the realizations and the ignorant expectation value $x^Z$ beats $x$ in only 32% of the realizations. (b) For stronger $X$ monitoring, with $\tau_x/\tau_z = 15$, the advantage of the ignorant smoothed estimate $x^Z$ is reduced, with only 54% of the realizations beating $x$. (c) Relative mean square error $Q$ as a function of the relative monitoring strength $\tau_x/\tau_z$, showing that for $\tau_x/\tau_z \gg 1$ the ignorant smoothed estimate $x^Z$ reliably outperforms the expectation value $x$, despite having restricted information.


