


**Adaptive filtering of projective quantum measurements using discrete stochastic methods**Riddhi Swaroop Gupta<sup>\*</sup> and Michael J. Biercuk*ARC Centre of Excellence for Engineered Quantum Systems, School of Physics, The University of Sydney, New South Wales 2006, Australia* (Received 22 January 2021; accepted 20 April 2021; published 13 July 2021)

Adaptive filtering is a powerful class of control theoretic concepts useful in extracting information from noisy data sets or performing forward prediction in time for a dynamic system. The broad utilization of the associated algorithms makes them attractive targets for similar problems in the quantum domain. To date, however, the construction of adaptive filters for quantum systems has typically been carried out in terms of stochastic differential equations for weak, continuous quantum measurements, as used in linear quantum systems such as optical cavities. Discretized measurement models are not as easily treated in this framework, but are frequently employed in quantum information systems leveraging projective measurements. This paper presents a detailed analysis of several technical innovations that enable classical filtering of discrete projective measurements, useful for adaptively learning system dynamics, noise properties, or hardware performance variations in classically correlated measurement data from quantum devices. In previous work we studied a specific case of this framework, in which noise and calibration errors on qubit arrays could be efficiently characterized in space; here, we present a generalized analysis of filtering in quantum systems and demonstrate that the traditional convergence properties of nonlinear classical filtering hold using single-shot projective measurements. These results are important early demonstrations indicating that a range of concepts and techniques from classical nonlinear filtering theory may be applied to the characterization of quantum systems involving discretized projective measurements, paving the way for broader adoption of control theoretic techniques in quantum technology.

DOI: [10.1103/PhysRevA.104.012412](https://doi.org/10.1103/PhysRevA.104.012412)**I. INTRODUCTION**

Quantum computers in the noisy intermediate-scale quantum (NISQ) era face considerable challenges in mitigating the effects of decoherence on intermediate-scale multiqubit devices. In realistic operating environments subject to noise, difficulties arise in device calibration, control, and error mitigation and the complexity of these challenges increases with system size. As the number of qubits on a device increases, existing calibration and control techniques typically lead to an infeasible resource overhead at the expense of available compute time. In overcoming these contemporary challenges, insights from classical inference and control engineering literature generally appear to be relevant. Indeed, contemporary classical techniques for quantum systems characterization [1–3], adaptive tomography [4–8], and parameter estimation [9–12] add to a growing body of literature in the last decade which has focused on realizing inexpensive and scalable characterization and control methods. Classical protocols have also been used for implementing optimal or efficient experiments [13,14] by enabling adaptive measurement selection or qubit allocation [15,16].

However, applying concepts from classical control engineering to quantum systems is not straightforward due to the peculiar role of measurement in quantum mechanics. This complication has been typically addressed by focusing on weak measurement of quantum systems captured via continuous stochastic differential equations [17–19]. In contrast,

when considering projective measurement records, continuous stochastic filtering methods are no longer applicable as a quantum state is reset after each projective measurement. Instead, many approaches for analyzing discrete measurements rely on rapid averaging or batch postprocessing single-shot raw data [20,21]. The conversion of discretized measurement outcomes into continuous variables through these means ultimately discards useful time-domain information and adds a computational bottleneck that unnecessarily slows state-estimation algorithms. Despite the importance of discrete measurement analysis in a wide range of applications for the characterization, calibration, and control of quantum systems, single-shot projective measurements have not yet been directly incorporated into classical filtering techniques for quantum control. In particular, one requires a measurement model that is quantum mechanically accurate but also correctly captures the statistical properties of discrete observations of an otherwise continuous state space.

In this work we rigorously demonstrate how adaptive filtering incorporating quantum projective measurements can be understood through the theoretical framework of classical nonlinear filtering, and describe in full a set of algorithmic modifications enabling their use. First, we demonstrate an efficient computational technique to discretize the amplitude domain of a classical signal in a manner that preserves statistical compatibility with classical filtering theory; this is achieved by combining Born's rule with an appropriate ansatz for measurement noise in accordance with classical amplitude quantization. Second, we solve the resulting inference problem using a sequential Monte Carlo framework called particle filtering. Here, continuous probability distributions

<sup>\*</sup>Present address: IBM Quantum, IBM Australia, Sydney, New South Wales 2065, Australia; riddhi.sw@gmail.com

over a state space are approximately solved by a collection of discrete particles that undergo nonlinear transformations. We introduce a set of rules to perform particle transformations in a manner compatible with single-qubit projective measurements.

With these modifications, we numerically validate that discretizing the continuous amplitudes of classical random signals, and performing discrete approximations to continuous probability distributions permits the convergence properties of classical nonlinear filtering carry over to classical filtering of discretized single-qubit projective measurements. Building on the experimental demonstration of adaptive spatial dephasing-field characterization first presented in Ref. [22], we use simulations to study the true error scaling coefficient with particle number. This error scaling coefficient is associated with the rate at which discrete empirical distributions tend to the true, continuous Bayesian posterior as the number of particles increase, thereby providing a numerical characterization of the convergence behavior of the filter. We probe the properties of our proposed filter using two different types of numerical tests. First, our desired operation, where we run our particle filter using single-shot projective measurements. Here, the error scaling coefficient with particle number agrees with theoretically expected values from classical convergence analysis, which is unanticipated for the modifications introduced here for projective measurements. Second, we break our filter by uniformly randomizing (discarding) some state information before receiving the next single-shot projective measurement for every iteration of the filter. As the overall injection of random information progressively increases, the error scaling coefficient gradually increases above theoretically anticipated values. These empirical studies provide evidence that our methods approximate the true classical filtration generated by a sequence of projective measurements, and demonstrate model robustness for a range of challenging operating conditions.

The paper is presented in the following parts. In Sec. II we discuss the use of particle filters as a discrete approximation to continuous probability distributions in classical inference and their utility in solving difficult Bayesian inference problems. In Sec. III, we outline how the quantum mechanical Born's rule can be combined with insights from classical discrete signal analysis so that individual projective measurement outcomes can be analyzed by classical filtering techniques. Subsequently, we show that classical convergence properties of particle filters are retained even if discrete, projective measurements are used. In Sec. IV, we consider adaptive filtering with single-shot projective measurements, first presented in Ref. [22], now discussed with greater generality and a focus on the convergence properties of filtering distributions in a general physical setting. Using the specific example of Ref. [22], in Sec. V we present numerical evidence for favorable convergence characteristics. Concluding remarks are provided in Sec. VI.

## II. PARTICLE FILTERING METHODS

Particle filters belong to a broader class of classical algorithms, known as sequential Monte Carlo algorithms, but have featured in quantum characterization and control applications.

Quantum particle filters were developed in the context of continuous quantum measurements [23], while particle methods have also been used for adaptive Hamiltonian learning using projective measurements [9,24]. Outside of quantum systems characterization, these methods have been popularized in nonlinear engineering control theory and probabilistic robotics, for example, in classical simultaneous localization and mapping (SLAM) problems [25–27] where a robot must characterize (“map”) and physically navigate through an unknown terrain. A common theme arising from these diverse applications is that particle filters perform strongly in high-dimensional, non-Gaussian, and nonlinear state spaces [28–30] that typically arise in context of characterizing quantum systems.

The efficacy of these particle filtering methods in solving inference problems is due to their so-called particle branching mechanisms. These branching mechanisms are an essential part of assessing convergence, computational efficiency, and correctness for a particle filter, irrespective of the specific details about measurement or system dynamics in any physical application. The subset of particle filters discussed here have extremely convenient convergence characteristics that can be exploited for designing algorithms for quantum control. In particular, a convenient convergence property is that the statistical behavior of branching process determines the rate at which a particle filter converges to the true Bayesian posterior distribution as the number of particles increase [31]. Furthermore, these convergence characteristics do not place any major constraints on the dynamical evolution or measurement procedures for the system under consideration. This insight paves the way for using nonlinear classical filtering directly on discrete, single-shot outcomes obtained from quantum systems in a wide range of physical applications.

The key objective of any particle filter is to approximate a true continuous Bayesian posterior distribution [28–30,32,33]. A true continuous Bayesian posterior distribution, denoted  $\pi_t$ , is the conditional probability of observing  $X_t$  given a set of measurements  $Y_{0:t}$ . The distribution  $\pi_t$  is expressed as the conditional probability of  $X_t$  given the  $\sigma$  field generated by the observations  $Y_{0:t}$ . In general, a transformation from  $X_t \rightarrow Y_t$  is nonlinear, and in case of single-qubit measurements, the binary nature of  $Y_t \in \{0, 1\}$  further makes it difficult or impossible to derive an analytical filter update using measurement data.

In the particle filtering approximation, the  $i$ th particle represents a hypothesis about  $X_t = x_t^{(i)}$ , known as the “position” of the particle in  $\mathbb{S}_X$ , the state space associated with  $X_t$  (refer Table I). The collection of particle positions represents the empirical approximation to  $\pi_t$ . This approach permits a mechanism by which a filtering algorithm may be applied in order to obtain a numerical estimate of the posterior distribution by directly transforming particles at each iteration, rather than seeking analytical solutions using algebraic inversions or decomposition methods. This discrete approximation  $\pi_t^n$  for the true  $\pi_t$  is expressed as

$$\pi_t^n := \frac{1}{n} \sum_{i=1}^n \delta_{x, x_t^{(i)}}, \quad (1)$$

where  $n$  represent the total number of particles  $\{x_t^{(i)}\}_{i=1}^n$ , and each particle represents a hypothesis for  $X_t$ . In the above, the

TABLE I. State, function, and measure space notation (consistent with [31]). The abbreviation R.V. stands for any random variable and  $\mathcal{S} \equiv \sigma(\mathcal{S})$ , i.e., the  $\sigma$  field generated by the state space  $\mathcal{S}$ .

Sym.	Definition
$\mathcal{S}$	A complete, separable metric (state) space for a R.V.
$\mathcal{S}$	The Borel $\sigma$ algebra [34] generated by $\mathcal{S}$
$C(\mathcal{S})$	The space of real continuous functions on $\mathcal{S}$
$M(\mathcal{S})$	The space of $\mathcal{S}$ -measurable functions on $\mathcal{S}$
$B(\mathcal{S})$	The space of bounded $\mathcal{S}$ -measurable functions on $\mathcal{S}$
$C_b(\mathcal{S})$	The space of bounded continuous functions on $\mathcal{S}$
$P(\mathcal{S})$	The space of probability measures on $(\mathcal{S}, \mathcal{S})$ s.t. $\mu \in P(\mathcal{S})$ satisfies $\mu(\mathcal{S}) = 1$

Kronecker delta  $\delta_{x,(\cdot)}$  is used because the approximate probability measures represent discrete probability distributions. This discrete approximation to a continuous distribution is schematically depicted in Fig. 1(a), where a set of  $n = 5$  discrete particle positions are illustrated as points on a continuous probability density by colored circular markers.

We now provide an overview of the particle-filtering algorithm. During filtering, particles  $\{x_t^{(i)}\}$  are transformed by operations which represent dynamical or measurement processes, represented by  $K_t$  and the likelihood function  $g_t$ , respectively. In general the transformations represented by  $K_t$  and  $g_t$  are nonlinear, and the resulting transformed particles need not resemble the forms of analytic probability distributions. Under the additional assumption that  $X$  is Markov, one uses the transition kernel for a Markov chain to obtain the distribution at  $t$  if the distribution at  $t - 1$  is known. The result is called the predictive probability measure  $p_t$  or, equivalently, the dynamical model for the filtering problem if drift characterization is relevant to a system under consideration. Thus, Bayes rule for the conditional probability of  $X_t$  given observations  $Y_{0:t}$  is written in the typical recursive form as

$$\pi_t := g_t * p_t(A) := \frac{\int_A g_t(x) d p_t(x)}{p_t g_t}, \quad (2)$$

$$p_t := K_{t-1} \pi_{t-1}, \quad p_t \in P(\mathcal{S}_X), \quad (3)$$

$$p_t g_t := \int_{\mathcal{S}_X} g_t(x) d p_t(x) > 0. \quad (4)$$

The use of the projective product in the first line,  $*$ , is essentially a restatement of Bayes rule. While the distribution  $p_t$  is the true continuous predictive distribution, it can also be approximated by individually transforming particles in an empirical distribution  $p_t^n := K_{t-1} \pi_{t-1}^n$ .

For each incoming measurement at  $t$ , a particle weight, denoted  $G_t^{(i)}$ , is computed for all  $i = 1, 2, \dots, n$  particles. These weighted particles  $\bar{x}_t^{(i)}$  form the weighted distribution  $\bar{\pi}_t$ , expressed as

$$\bar{\pi}_t^n := \sum_{i=1}^n G_t^{(i)} \delta_{x, \bar{x}_t^{(i)}}, \quad \bar{x}_t^{(i)} \sim p_t^n. \quad (5)$$

Here, the particle weight  $G_t^{(i)}$  represents the probability of receiving a measurement  $Y_t = y_t$  if the hypothesis captured by the  $i$ th particle  $X_t = \bar{x}_t^{(i)}$  is taken to be true. The bar notation  $\bar{\cdot}$

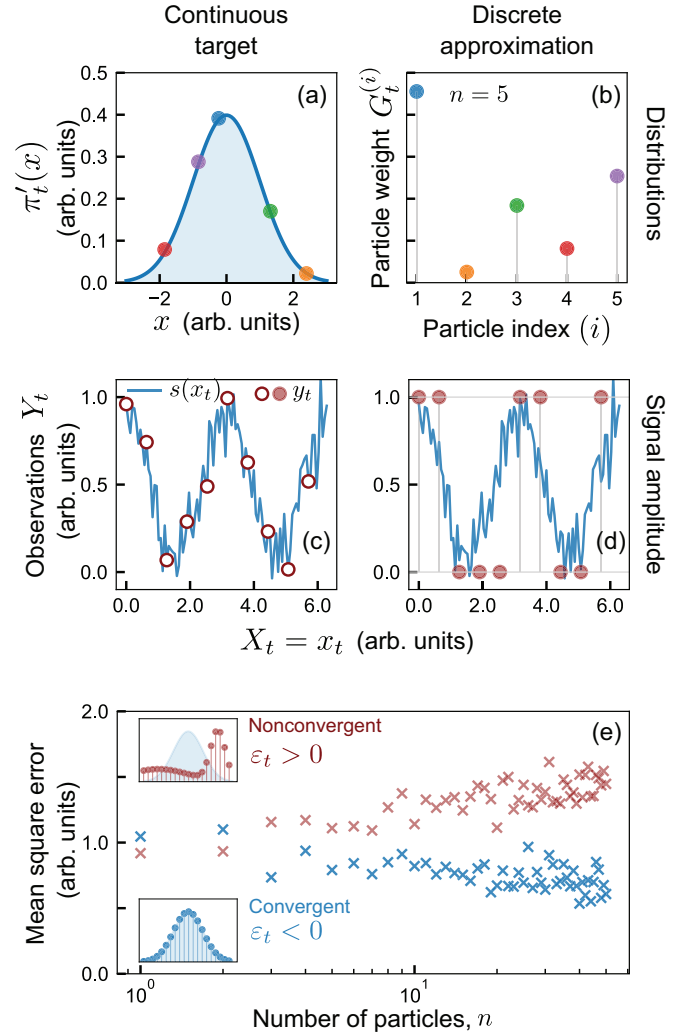


FIG. 1. Schematic overview of discrete approximations to continuous, classical quantities and their convergence behavior. (a) Illustration of a true continuous probability density  $\partial\pi_t/\partial x$  with respect to a state space  $x \in \mathcal{S}_X$  in one dimension (1D) (blue solid line). Discrete particle positions (colored markers) for  $n = 5$  correspond to particle indices in (b). (b) Illustrative plot of particle weights  $G_t^{(i)}$  vs particle index  $i$  for some fixed  $t$  computed from the particle filter. Weights are normalized,  $\sum_{i=1}^n G_t^{(i)} = 1$ , and the limit  $n \rightarrow \infty$  recovers continuous distributions from empirical approximations in the mean-square limit. (c) Idealized depiction of discrete-time observations (red open circles) of a continuous signal (blue solid). (d) Discretization of signal amplitude in (c) into two levels yields binary measurements (red solid markers). Both discrete-time and discrete-amplitude properties of the signal are emphasized by vertical and horizontal gray lines, respectively. (e) Schematic representation of a semilog plot of mean-square errors  $\mathcal{L}_t$  vs number of particles as  $n$  increases for fixed but sufficiently large  $t$ . Convergent (nonconvergent) behavior in blue (red) crosses corresponds to the extent of overlap of weighted particle positions and the true continuous distribution in the lower inset (upper inset). Error scaling coefficient  $\varepsilon_t$  describes rate of change of errors as  $n \rightarrow \infty$  for each  $t$  and negative values for  $\varepsilon_t$  indicate convergent behavior.

indicates that the distribution should be computed after evolving particles from  $t - 1$  into the current iteration at  $t$ , and  $G_t^{(i)}$  are calculated based on a single measurement  $Y_t$  received at  $t$ .

As a schematic illustration, for each colored circular marker in Fig. 1(a), the corresponding particle weight vs particle index  $i = 1, 2, \dots, n$  is shown in Fig. 1(b).

The frequency with which a particle position is represented in the next generation at  $t + 1$  is proportional to its particle weight. Here, the weighted empirical distribution of Eq. (5) is sampled according to the distribution of the particle weights  $\{G_t^{(i)}\}$ . The resulting offspring particles form posterior  $\pi_t^n$ ; equivalently, the prior distribution for the iteration at  $t + 1$  and weights are reset to uniform. Over many iterations, particles with higher weights are represented more frequently in the particle populations enabling empirical particle distributions to gradually converge to the true, continuous Bayesian posterior. In words, as  $n \rightarrow \infty$ , the set of weights in Fig. 1(b) approximate Fig. 1(a) for the schematic introduced earlier.

The efficacy of any particle filter is linked to particle branching processes. These depend on how particle weights  $\{G_t^{(i)}\}$  are calculated whenever new measurement information is received, and how particles are propagated from one iteration  $t$  to the next  $t + 1$  via resampling. Branching mechanisms therefore form the core algorithmic representation of the approach to providing an approximate solution to the Bayesian inference problem. Many different particle branching processes have been proposed in particle filtering literature [35–38]. We focus on branching processes satisfying Proposition 1 which benefit from a number of well-understood convergence characteristics [31]. One of these characteristics is that the correctness of a particle filter can be analyzed as convergence to the true Bayesian posterior distribution as the number of particles  $n$  increase for any  $t$ :

$$(\pi_t^n)_{n=1}^\infty \rightarrow \pi_t. \quad (6)$$

This equation describes the convergence of a particle filter to the true Bayesian posterior. The arrow schematically depicted in Eq. (6) can be interpreted as “convergence in expectation” and “almost-sure convergence” of the empirical to the true Bayesian distributions. Both these forms of convergence describe the expected distance between two probability distributions, where the first focuses on distances between moments of a distribution and the second describes overall convergence between distributions, subject to several technical considerations [31].

A schematic illustration of different convergence behavior is given in Fig. 1(e). Here, the square error between the means of the target and estimated conditional distribution of  $X_t$  given data  $Y_{0:t}$  increases (decreases) with particle number  $n$  for a divergent (convergent) particle filter, as depicted in red (blue) crosses. The consideration of higher-order moments of the posterior distribution is omitted from this schematic figure. Upper and lower insets plot particle locations (circular markers) against a target continuous distribution (shaded blue) for divergent vs convergent particle filtering, respectively, illustrating differences between the target and estimated distribution of  $X_t$  given data  $Y_{0:t}$  during particle filtering.

We now explore the technical considerations under which convergent particle filtering can be guaranteed. To achieve convergence during particle filtering applications, the two important technical conditions are that  $g_t$  is continuous and bounded and  $K_t$  is Feller [31]. Once these conditions are satisfied, a particle filter possessing branching properties of

Proposition 1 is guaranteed to have both convergence in expectation and almost-sure convergence to the true Bayesian posterior as  $n$  increases [31,39].

*Proposition 1.* Let  $i$  denote a particle label with  $i = 1, 2, \dots, n$ ,  $G_t^{(i)}$  denote a particle weight, and  $\xi^{(i)}$  denote the frequency of a particle position at  $t$ . Branching mechanisms for a particle filter satisfy [31] the following:

- (1) Constant particle number  $n = \sum_{i=1}^n \xi^{(i)}$  for all  $t$ .
- (2) Conditional mean proportional to  $G_t^{(i)}$ , that is,  $\mathbb{E}[\xi^{(i)} | \bar{\mathcal{G}}_t] = n G_t^{(i)}$ .
- (3) Conditional covariance matrix  $(A_t^n)_{ij} := \mathbb{E}[(\xi^{(i)} - n G_t^{(i)})^T (\xi^{(j)} - n G_t^{(j)}) | \bar{\mathcal{G}}_t]$  satisfy  $q^T A_t^n q \leq n c_t$  for some constant  $c_t$  and for any  $n$ -dimensional vector  $q$  with entries  $|q^{(i)}| < 1$ .

In Proposition 1, the quantity  $\xi^{(i)}$  is the number of times the parent particle  $x_t^{(i)}$  is copied and represented in the offspring generation of particles. The first proposition specifies that the total number of particles remains  $n$  for all  $t$  enabling a simpler analysis of the full branching random process from  $t = 0$  to  $t$ , i.e., only the branching transitions within each  $t$  need to be considered. The second proposition restates that empirical weight of the particle is associated with the true probability of observing that particle (state information) given some observed history via  $\bar{\mathcal{G}}_t$ . The last property places a constraint on the covariance matrix associated with the branching process. This constraint appears to have no *a priori* justification, but it is a condition associated with a particle filter’s convergence properties [31].

In particular, we focus on one aspect of convergence analysis which concerns the scaling behavior of empirical distributions with particle number. This scaling behavior can be associated with the behavior of true errors generated during filtering irrespective of the system under consideration, and can be compared with the actual performance of particle filters in numerical simulations. Using Proposition 1, one can derive conditions on empirical filtering distributions as

$$\mathbb{E}[(p_t^n f - \pi_{t-1}^n(K_{t-1}f))^2] \leq \frac{\|f\|_\infty^2}{n}, \quad (7)$$

$$\mathbb{E}[(\pi_t^n f - \bar{\pi}_t^n f)^2] \leq \frac{c_t \|f\|_\infty^2}{n}, \quad (8)$$

for all  $f \in B(\mathbb{S}_X)$ ,  $t \geq 0$ , where  $B(\mathbb{S}_X)$  is the space of bounded measurable functions on  $\mathbb{S}_X$  (see Table I), and  $\|f\|_\infty^2$  is a infinity norm for the function. These inequalities state that the expected distances between empirical distributions shrink as filtration proceeds. The two different types of distances under consideration are first, from posterior distributions at  $t - 1$  to predictive distribution at  $t$  [Eq. (7)], and second, the empirical distribution before and after particle resampling within each  $t$  [Eq. (8)]. The specific value of  $c_t$  in Eqs. (7) and (8) depends on the type of branching process and its value cannot always be deduced *a priori*. Supporting technical derivations for these equations and their relevance to the overall proofs for convergent particle filters is restated for completeness in Ref. [39].

Of the branching processes satisfying Proposition 1, “bootstrap” filters are a popular example; an example algorithmic implementation is outlined in Algorithm 1. Here, particle weights are calculated only using the likelihood function  $g_t(\cdot)$ .



**Algorithm 1** Bootstrap

---



---

```

if  $t = 0$  then
  Sample  $x_0^{(i)} \sim \pi_0$ ,  $i = 1, 2, \dots, n$ 
end if
if  $t > 0$  then
  Sample  $\bar{x}_t^{(i)} \sim K_{t-1} \pi_{t-1}^n$ ,  $i = 1, 2, \dots, n$ 
  Receive  $Y_t = y_t$ ; compute  $G_t^{(i)} = \frac{g_t^{(i)}(\bar{x}_t^{(i)})}{\sum_l g_t^{(i)}(\bar{x}_t^{(l)})}$ 
  Replace  $\bar{x}_t^{(i)}$  with  $\xi^{(i)}$  offspring, such that  $n = \sum_{i=1}^n \xi^{(i)}$ 
  Relabel offspring as  $x_t^{(i)}$ ; reset  $G_t^{(i)} = 1/n$  for  $i = 1, \dots, n$ .
end if

```

---



---

In Algorithm 1, one sees that the empirical distributions of the bootstrap particle filter follow the progression

$$K_{t-1} \pi_{t-1}^n \rightarrow \bar{\pi}_t^n \xrightarrow[n]{\text{resample}} \pi_t^n, \quad (9)$$

and all empirical distributions have a constant particle number  $n$ . In the above, the first arrow represents computing particle weights using the likelihood function. The second arrow represents particle resampling as summarized by the last two lines of Algorithm 1. For bootstrap particle filters of Algorithm 1, the progression depicted in Eq. (9) is a multinomial branching process and it is theoretically tractable to show that  $c_t = 1$  [39].

In the next section we introduce a model for projective measurements on quantum systems that may be employed in a data inference problem in which we must learn or estimate system dynamics based on a measurement record. Subsequently, we will proceed to establish a central result of this paper: that the use of the likelihood function introduced in Sec. III for projective measurements on quantum systems does not disrupt the essential convergence properties of these particle filters.

### III. NONLINEAR FILTERING OF SINGLE-QUBIT MEASUREMENTS

In many physical settings, it is often desired that some continuous-valued classical process  $X$  is inferred from a discrete-time sequence of measurements. The challenge posed by single-qubit measurements is that measurements can only assume certain allowed values  $Y \in \{0, 1\}$ . This challenge that a continuous  $X$  can only be observed as discrete outcomes  $Y$  is well known in classical literature as the quantization of signal amplitude. The key insight described in this section is that the combination of an analytic prescription of Born's rule with classical amplitude quantization theory to describe single-shot projective measurement outcomes provides compatibility with any classical filtering algorithm.

In classical signal processing, it is often the case that the continuous amplitude of some classical process is measured by a sensor that can only record discrete amplitude levels. In this context, amplitude-quantization theory specifies the statistical properties of discrete measurements of an otherwise continuous-amplitude signal [40]. The amplitude of a classical signal is said to be discretized by  $B$  bits if its continuous-amplitude is measured by a physical sensor which only has  $2^B$

discrete amplitude levels, up to some constant offset. These classical amplitude-quantized signals are analyzed via sampling a signal in the amplitude domain leading to an increase in the overall noise floor [40]. We illustrate the procedure of amplitude quantization in Figs. 1(c) and 1(d). An example of a continuous-amplitude discrete-time classical signal is first shown in Fig. 1(c) as red open markers, corresponding to discrete-time noisy measurements of a continuous-time signal (blue solid). This signal further undergoes a classical discretization of signal amplitude, where the  $y$  axis is discretized into two discrete levels ( $B = 1$  case). The resulting signal is a discrete-amplitude discrete-time signal given by the red filled markers in Fig. 1(d).

For concreteness, we treat the case  $B = 1$  and extend this classical analogy to single-qubit projective measurements. We consider a classical signal consisting of a sequence of projective measurements. Let single-qubit states be expressed in the  $\hat{\sigma}_z$  basis, and  $\hat{U}(t', t; X)$  be some single-qubit unitary interaction that depends on  $X$  for a qubit initially prepared in the ground state at the start of the procedure at  $t'$ . The Born probability for the outcome  $Y_t$  of the projective measurement commenced at  $t$  is  $|\langle Y_t | \hat{U}(t', t; X) | 0 \rangle|^2$  for  $Y_t \in \{0, 1\}$  and  $t' < t$ . Under these circumstances, the nonlinear measurement model for single-qubit measurements can be described as an outcome  $Y_t$  of a Bernoulli trial. This model is denoted in notation by the symbol  $\mathcal{Q}(\cdot)$  for taking a biased coin flip with the bias given by the argument

$$Y_t = \mathcal{Q}(|\langle Y_t | \hat{U}(t', t; X) | 0 \rangle|^2), \quad (10)$$

$$\mathcal{Q}(z) := \text{Binom}(p = z; n = 1, k = 1). \quad (11)$$

In the above, a binomial distribution has success probability  $z$ , number of trials  $n = 1$ , and  $k = 1$  successes. These repeated single-shot measurements spaced  $\Delta t$  apart gives rise to discrete classical random processes. We now interpret  $t$  to be a discrete-time index marking a set of repeated single-shot measurements  $Y_{0:t} := \{Y_0, Y_1, \dots, Y_t\}$  associated with  $X_{0:t} := \{X_0, X_1, \dots, X_t\}$ . The time step  $\Delta t$  is set by total time for system preparation, interaction, measurement, and reset, with  $\Delta t$  typically much greater than the unitary interaction period in practical experiments. The slowly varying assumption on  $X$  is that  $\Delta t$  is much faster than any variation in  $X$  and  $X$  is approximately constant over the interaction  $\hat{U}$ .

The key observation is that some classical continuous amplitude  $X_t$  yields only a discrete allowed value of  $Y_t$  upon observation. If a sensor measures a continuous amplitude signal  $s(X_t)$ , only as discrete allowed amplitude levels  $Y_t \in \{0, 1\}$ , then this sensor has the overall effect of adding noise in the  $t$  domain of the signal [40–43], represented by  $v_t$ . We express the association of the classical abstract signal with Born's rule as

$$|\langle Y_t | \hat{U}(t', t; X) | 0 \rangle|^2 \leftrightarrow s(X_t) + v_t + \frac{1}{2}, \quad (12)$$

where  $v_t$  represents uncertainty in our knowledge of the true Born probability inferred from single-shot measurements, and the term  $\frac{1}{2}$  is an arbitrary global rescaling factor so that  $s(X_t) + v_t$  is zero mean for a single-shot measurement of a maximally mixed qubit state. For typical single-qubit

measurements characterizing a quantum system, we assume commuting projective measurement procedures such that a joint probability density over the random variates  $Y_t, s(X_t), v_t$  exist for all  $t$ .

The statistical properties of  $v_t$  determine how effectively one can incorporate discrete-amplitude measurements into conventional classical filtering by proposing an appropriate noise model for capturing uncertainty in single-shot measurement information. Examples of classical amplitude quantized sensor information assume a variety of models, for example, where signal distributions are convolved ( $\star$ ) with a pulse train or uniform distribution, or integrated above and below each discrete amplitude level and renormalized to represent measurement errors [40,42,43]. Our departure from these approaches is to consider that any uncertainty in our knowledge of the true Born probabilities arise from truncated error distributions representing amplitude quantization into two discrete levels. Specifically, we assume a noise model to be zero-mean Gaussian distribution  $\mathcal{N}(0, \Sigma_v)$  with variance  $\Sigma_v$ , which is convolved with a uniform  $\mathcal{U}$  distribution as

$$\Pr[v_t] = \mathcal{N}(0, \Sigma_v) \star \mathcal{U}(a, b), \quad \forall t, \quad \mathbb{S}_v = \mathbb{R}, \quad (13)$$

where  $\Pr[Z]$  represents a probability measure for the real-valued, random variate  $Z$  defined over the space  $z \in \mathbb{S}_Z$ , and  $a, b$  represents finite bounds on the values of these errors due to amplitude discretization. In the above, the notation  $\Pr[Z]$  is interpreted as a probability mass function over discrete values in  $\mathbb{S}_Z$ , or density over continuous values in  $\mathbb{S}_Z$ .

Under these considerations, the continuous-amplitude measurement model takes the form

$$\begin{aligned} \Pr[Y_t = y|s(X_t)] \\ = \int_{\mathbb{S}_v} \left( [s(X_t) + v_t](\delta_{y,1} - \delta_{y,0}) + \frac{1}{2} \right) \Pr[v_t](v) dv, \end{aligned} \quad (14)$$

where  $\delta_{x,y}$  takes the value 1 if  $x = y$  or zero otherwise. As one example of the noiseless ideal case  $v_t \equiv 0$ , the ideal Born probability of observing the qubit in  $|1\rangle$  is then  $s(X_t) + \frac{1}{2}$  or  $\frac{1}{2} - s(X_t)$  for observing the qubit in the  $|0\rangle$  state.

Substituting Eq. (13) into (14), and performing the relevant integration yields the final form of likelihood function under amplitude discretization of Born probabilities,

$$\begin{aligned} \Pr[Y_t = y|s(X_t)] &= \frac{\rho_0}{2} + \rho_0 s(X_t)(\delta_{y,1} - \delta_{y,0}) \\ &\equiv g_t^{Y_t=y_t}(X_t), \end{aligned} \quad (15)$$

with the real-valued scalar  $\rho_0$  obtained from integration as

$$\rho_0 = \operatorname{erf}\left(\frac{2b}{\sqrt{2\Sigma_v}}\right) + \frac{\sqrt{2\Sigma_v}}{2b} \frac{e^{-\left(\frac{2b}{\sqrt{2\Sigma_v}}\right)^2}}{\sqrt{\pi}} - \frac{1}{2b} \frac{\sqrt{2\Sigma_v}}{\sqrt{\pi}}, \quad (16)$$

where  $\operatorname{erf}$  is the error function with values between  $[-1, 1]$ . For error sources that are symmetric with respect to how they affect single-qubit states, one sets  $-a = b$  in the calculation above. Asymmetric error distributions  $-a \neq b$  may arise, for example, when noise during state detection depends on the state of qubit at the start of a projective measurement procedure, e.g., state-dependent decay of hyperfine

qubits in trapped-ion quantum computers [44,45], but are not treated in this paper. In circumstances when  $b \leq 3\Sigma_v$ , model failure may occur as information in the original distribution is being discarded by the procedure for amplitude discretization.

While we have focused on single-qubit measurements with two possible discrete amplitude levels (“0” or “1”), one may extend to  $B$ -qubit measurements with  $2^B$  discrete levels if these states are individually discernible in experiments. In all of these cases, we assume that  $s(X_t)$  is continuous such that the properties of the resulting discrete-amplitude signal can be described via methods of Refs. [40,42]. Additionally, we will also assume that  $s(\cdot)$  is bounded and  $s(\cdot)$  has an inverse  $s^{-1}(\cdot)$  on  $\mathbb{S}_{X_t}$ , the space of allowed continuous values for  $X_t \in \mathbb{S}_{X_t}$ . As discussed below, the boundedness property ensures that our likelihood function can be safely incorporated into bootstrap particle filtering while preserving convergence properties of these filters. Subsequently, in Sec. IV, the inverse  $s^{-1}$  is used to share estimated state information in small regions for adaptive particle filtering.

We now establish that the likelihood for projective measurements proposed here can be incorporated within bootstrap particle filtering without affecting standard convergence theorems. Our likelihood function is given by Eqs. (15) and (16). As discussed in Sec. II, this function needs to be bounded and continuous for conventional convergence properties of bootstrap particle filters to hold. Examining Eq. (16), we see that for  $b \neq 0$ , the scalar value  $\rho_0$  is bounded, as evident by considering the following two limiting cases. The limit  $\Sigma_v \rightarrow 0$ , the scalar  $\rho_0 \rightarrow 1$ , and the ideal case of a coin flip with the win probability given by Born’s rule is obtained. In the opposite limit,  $\Sigma_v \rightarrow \infty$ , the scalar  $\rho_0 \rightarrow 0$  and no inference is possible. Assuming  $s(\cdot)$  is bounded and  $b \neq 0$ , the proposed likelihood function  $g_t^{Y_t=y_t}(X_t)$  is also bounded. For continuity, it is required that the likelihood function is continuous over the state space of  $X_t$  for a specific instance of data  $Y_t$  [31,42,46]. For a fixed instance,  $Y_t = y_t$ ,  $g_t^{Y_t=y_t}(X_t)$  in Eq. (15) is expressed by either  $\rho_0/2 + \rho_0 s(X_t)$  or  $\rho_0/2 - \rho_0 s(X_t)$ . Assuming  $s(\cdot)$  is continuous, the proposed likelihood function  $g_t^{Y_t=y_t}(X_t)$  is also continuous with respect to  $X_t$  for an instance of  $y_t$ . Thus, for continuous and bounded  $s(\cdot)$  and  $b \neq 0$ , the proposed likelihood satisfies the key properties required for conventional convergence properties. Further, our result is general in the sense that aside from the observation process, no further information about the physical application, system dynamics, or the noise environment is being assumed. The specific case  $b = \frac{1}{2}$  is considered in the remaining sections.

So far our work allows quantum projective measurements to be analyzed by fully exploiting the power of particle techniques for nonlinear, non-Gaussian, high-dimensional state spaces typically arising in the context of quantum characterization problems. Next, we provide technical details about an adaptive filtering framework as a variant of bootstrap particle filter with multinomial branching. However, our framework departs substantially from traditional bootstrap filters as it incorporates features for adaptive control using quantum projective measurements. We outline these features in the next section and subsequently investigate the numerical error scaling behavior of our protocol with the  $1/n$  behavior predicted by Eqs. (7) and (8).

#### IV. ADAPTIVE FILTERING FOR QUANTUM SYSTEMS CHARACTERIZATION

Our challenge in this section is to outline a theoretical framework capable of adaptively characterizing and predicting classical correlations arising in projective measurement records. These classical correlations may arise, for instance, due to the interaction of the quantum system with its ambient environment, unanticipated system dynamics, or intrinsic performance variations or noise in hardware. However, a naive application of multivariate filtering techniques to projective measurement records, even of commuting quantum observables, presents several difficulties. One issue is that quantum projective measurements are inherently local. In particular, Born's rule provides an unambiguous link between the measurement information and the elements of some multivariate  $X$  being inferred. In the language of classical estimation and mapping, this statement means that there is often no immediate benefit in defining a joint, classical Bayesian inference problem over elements of a multivariate  $X$  in filtering, a stark contrast to related classical literature, for example, for simultaneous localization and mapping (SLAM) applications [26,27,47]. Below, we present a deeper analysis of the implications of our adaptive methods, as first presented in [22], on overcoming these challenges and on filter convergence.

To accommodate predictive estimation of classical correlations in projective measurement records, we now associate points in some classical (continuous) parameter space with an index  $j$  as well as the discrete sequencing index  $t$ . This parameter space may arise in different physical applications where classical, continuous variables are sparsely sampled, as examples, due to geometric arrangement of qubits in space [16]; the choice different measurement procedures, tomography of continuous-variable systems [48], or noise spectroscopy. For a  $d$ -dimensional observation vector, we assume a classical joint probability distribution must exist over all  $d$  elements of  $Y_t^{(j)}$ ,  $j = 1, 2, \dots, d$ , i.e., quantum mechanical observables associated with  $Y_t^{(j)}$  commute for all  $j$  and  $t$ . If the labels  $j = 1, 2, \dots, d$  are measurements of *different* points in this parameter space, then each observation  $Y_t^{(j)}$  is local and provides information only about the elements of  $X_t^{(i)}$  uniquely associated with the label  $j$  at iteration  $t$ . If instead the labels  $j = 1, 2, \dots, d$  are repeated measurements of the *same* point in parameter space, then the empirical mean of repeated measurements  $\frac{1}{d} \sum_{j=1}^d Y_t^{(j)}$  is the empirical Born probability.

In order to efficiently learn classical correlations in projective measurement records and overcome these technical challenges, the adaptive filtering framework of this section shares estimated state information between elements of  $X$  during filtering, while behaving in accordance with the branching properties of Proposition 1. Our key observation is that many physical settings and noise sources lead to classical, continuously varying phenomena in  $j$  and  $t$ . In our framework, the outputs of classical state estimation at one coordinate point associated  $j$  can be spread locally about that location. The region or neighborhood within which information sharing occurs can also be estimated as part of the particle filtering process. Thus, in the language of classical mapping problems, for each  $t$  we estimate both map values at the point  $j$  and approximate map gradients in small regions about  $j$ . The

resulting output of the particle filter is a characterization of classical correlations over parameter space indexed by  $j$  and  $t$  using projective measurement records.

Framed in the language of classical map building, a true state vector  $X_t$  contains both the register of map values  $F_t$  and local approximate map gradient information  $R_t$ , that is,

$$X_t = [F_t^{(1)} \quad \dots \quad F_t^{(d)} \quad R_t^{(1)} \quad \dots \quad R_t^{(d)}] = [F_t \quad R_t], \quad (17)$$

$$\mathbb{S}_X = \mathbb{S}_F \times \mathbb{S}_R, \quad \forall t, \quad (18)$$

$$\mathbb{S}_F = [F_{\min}, F_{\max}], \quad \mathbb{S}_R = [R_{\min}, R_{\max}]. \quad (19)$$

In the above the quantities  $F_t$  and  $R_t$  represent  $d$ -dimensional, real, continuous-vector-valued random variables, and their outcomes take values between  $F_t^{(j)} \in [F_{\min}, F_{\max}]$  and  $R_t^{(j)} \in [R_{\min}, R_{\max}]$  for any location in parameter space labeled  $j = 1, 2, \dots, d$ .

Unlike typical particle filtering, our algorithm locally estimates the value of the field for a measured point at  $j$ , before sharing this information with neighboring points in the vicinity of  $j$ . The algorithm is responsible for determining the appropriate size of circular neighborhoods of radius  $R_t^{(j)}$  about the point labeled by  $j$ . The set of points inside the neighborhood  $Q_t$  shrinks or grows about  $j$  as the autonomous inference process progresses. Under these circumstances, this adaptive filtering protocol incorporates not only a local physical single-qubit projective measurement at  $j$  using Eq. (12),

$$Y_t^{(j)} = \mathcal{Q}(s(F_t^{(j)}) + v_t + \frac{1}{2}), \quad \mathbb{S}_Y = \{0, 1\}, \quad (20)$$

but also data messages generated by  $j$  for locations  $q_t$ ,

$$\hat{Y}_t^{(q_t)} = \mathcal{Q}(s(\chi_t^{(j,q_t)}) + \frac{1}{2}), \quad \mathbb{S}_Y = \{0, 1\}, \\ \forall q_t \in Q_t^{(j)}. \quad (21)$$

In the above,  $\chi_t^{(j,q_t)}$  is a convex combination of the existing estimate at  $q_t$  and new information due to a measurement  $Y_t^{(j)}$  received at  $j$ . The calculations associated with the term  $\chi_t^{(j,q_t)}$  invoke continuity of physical phenomena whereby new information at  $j$  is shared over a region about  $j$  via any choice of a sigmoidal function [49], here set to be a Gaussian function, parametrized by the estimate of  $R_t^{(j)}$ . The term  $\chi_t^{(j,q_t)}$  is computed using the posterior information at  $t$  and has the effect of introducing correlations between the elements of particle positions in the next iteration  $t + 1$ . Detailed technical information is provided in Sec. 2 of Ref. [39] for completeness.

Having modified conventional filtering with this information-sharing mechanism, we now focus on the branching properties of this framework and any potential implications on convergence properties of typical particle filters. In particular, two different types of particle species are used by the filter within a bootstrap filtering structure. Let  $\alpha$  particles be a set of  $n_\alpha$  number of particles. For each parent  $\alpha$  particle, let  $\beta_\alpha$  particles be a set of  $n_{\beta_\alpha}$  number of daughter particles useful for enabling neighborhood discovery and adaptation during filtering. The layer of  $\alpha$  particles  $\{x_t^{(\alpha)}\}_{\alpha=1}^{n_\alpha}$  carry a hypothesis about  $X_t$ ,

$$x_t^{(\alpha)} = [f_t^{(\alpha)} \quad r_t^{(\alpha)}], \quad (22)$$

where lowercase  $x_t, f_t, r_t$  refer to instances of the true process in uppercase  $X_t, F_t, R_t$ . Additionally in Eq. (26),  $\beta_\alpha$  particles are a set of  $n_\beta$  number of particles for each of the  $n_\alpha$  parents. A single  $\beta_\alpha$  particle carries a hypothesis for  $R_t^{(j)}$  assuming that  $F_t$  and neighborhoods at other locations  $R_t^{(j' \neq j)}$  are known, expressed in our notation as

$$x_t^{(j, \alpha, \beta_\alpha)} = [r_t^{(j, \alpha, \beta_\alpha)}], \quad (23)$$

where the distribution of  $\beta$  particles is the conditional distribution of  $R_t^{(j)}$  given  $X_t \setminus R_t^{(j)}$ . Here, the superscript notation  $(j, \alpha, \beta_\alpha)$  refers to the location label  $j$  for the parent  $\alpha$ -particle index,  $\alpha$ , and its associated  $\beta_\alpha$  particle. This empirical distribution of  $\beta_\alpha$  particles is related to the parent  $\alpha$  particle using the empirical mean

$$r_t^{(j, \alpha)} = \mathbb{E}_{\beta_\alpha} [x_t^{(j, \alpha, \beta_\alpha)}]. \quad (24)$$

The expression above relates the empirical mean of the  $\beta_\alpha$  particles for each parent  $\alpha$  particle to the element  $r_t^{(j, \alpha)}$ .

These manipulations lead to the following progression of empirical distributions for each  $t$ :

$$K_{t-1} \pi_{t-1}^{n_\alpha} \rightarrow \bar{\pi}_t^{(j, n_\alpha n_\beta)} \xrightarrow[N_1 = n_\alpha]{\text{resample}} \bar{\pi}_t^{(j, n_\alpha)} \xrightarrow[N_2 = n_\alpha]{\text{resample}} \pi_t^{n_\alpha}, \quad (25)$$

where the index  $j$  makes explicit that each iteration  $t$  receives physical measurements at the label  $j$ , and the superscript  $n_\alpha n_\beta$  (or  $n_\alpha$ ) indicates the total number of particles in the weighted distribution  $\bar{\pi}_t^{(j, \cdot)}$ . Two resampling steps are required to move from  $\bar{\pi}_t^{(j, n_\alpha n_\beta)} \rightarrow \bar{\pi}_t^{(j, n_\alpha)} \rightarrow \pi_t^{n_\alpha}$  corresponding to the arrows, where  $N_1$  and  $N_2$  represent the total number of particles in the new generation after resampling. This progression of empirical distributions of Eqs. (26) to (28) in Eq. (25) can be compared to the bootstrap particle filter in Eq. (9), where these measures are expressed as

$$\bar{\pi}_t^{(j, n_\alpha n_\beta)} := \sum_{\alpha=1}^{n_\alpha} \sum_{\beta_\alpha=1}^{n_\beta} G_t^{(j, \alpha, \beta_\alpha)} \delta_{x, \bar{x}_t^{(j, \alpha, \beta_\alpha)}}, \quad (26)$$

$$\bar{\pi}_t^{(j, n_\alpha)} = \sum_{\alpha=1}^{n_\alpha} \Omega_t^{(j, \alpha)} \delta_{x, \bar{x}_t^{(\alpha)}}, \quad (27)$$

$$\pi_t^{n_\alpha} = \sum_{\alpha=1}^{n_\alpha} \delta_{x, x_t^{(\alpha)}}. \quad (28)$$

In Eq. (26), the particle weights  $G_t^{(j, \alpha, \beta_\alpha)}$  are computed using a scoring function  $g_t^{(y_t^{(j)}, \alpha, \beta_\alpha)}(\lambda_1, \lambda_2, \Lambda_t)$ ,

$$G_t^{(j, \alpha, \beta_\alpha)} := g_t^{(y_t^{(j)}, \alpha, \beta_\alpha)}(\lambda_1, \lambda_2, \Lambda_t), \quad (29)$$

which incorporates the likelihood function of Sec. III and whose form and parameters are detailed in full in Sec. 2 of Ref. [39]. The weights  $G_t^{(j, \alpha, \beta_\alpha)}$  are rearranged into new weights  $\Omega_t^{(j, \alpha)}$  after the first resampling step in Eq. (27). The use of the overbar notation ( $\bar{\cdot}$ ) indicates that posterior particle positions at  $t-1$  have been propagated by the transition kernel to the step  $t$ , as indicated by sequence in Eq. (25).

Using the empirical definitions above, the pseudocode summarizing our proposed framework is given in Algorithm 2. As with standard particle filters, our algorithm is initiated by sampling from a prior distribution. At any iteration

---

**Algorithm 2** Adaptive filtering for quantum measurements
 

---

**if**  $t = 0$  **then**

Sample  $x_0^{(i)} \sim \pi_0, \quad i = 1, 2, \dots, n$

**end if**

**if**  $t > 0$  **then**

(i) Sample  $\bar{x}_t^{(i)} \sim K_{t-1} \pi_{t-1}^n, \quad i = 1, 2, \dots, n_\alpha$

(ii) Receive  $Y_t^{(j)} = y_t^{(j)}$ ; generate  $\beta_\alpha$  particles from Eq. (36) or (37)

(iii)–(v) Compute  $G_t^{(j, \alpha, \beta_\alpha)}$  using  $g_t^{(y_t^{(j)}, \alpha, \beta_\alpha)}(\lambda_1, \lambda_2, \Lambda_t)$

(vi) Replace  $\bar{x}_t^{(\alpha, \beta_\alpha)}$  with  $\xi_t^{(\beta')}$  offspring;  $N_1 = \sum_{\beta'=1}^{n_\alpha n_\beta} \xi_t^{(\beta')}$ .

Reset to uniform weights  $1/N_1$ .

(vii) Store  $r_t^{(j, \alpha)}$  and  $C_t^{(j)}$  from surviving particle pairs

(viii) Compute  $\Omega_t^{(j, \alpha)} = \frac{\text{num. of } \beta_\alpha \text{ survivors}}{N_1}$ ; discard  $\beta_\alpha$  particles,

$\forall \alpha$

(ix) Replace  $\bar{x}_t^{(\alpha)}$  with  $\eta_t^{(\alpha)}$  offspring;  $N_2 = \sum_{\alpha=1}^{n_\alpha} \eta_t^{(\alpha)}$ . Reset to uniform weights  $1/N_2$ .

(x) Relabel surviving particles as  $x_t^{(\alpha)}$ , for  $\alpha = 1, 2, \dots, n_\alpha$

(xi) Schedule next measurement  $j = \text{argmax}_k C_t^{(k)}$

(xii) Generate and update  $\hat{Y}_t^{(q)}$ , for all  $q_t \in \mathcal{Q}_t^{(j)}$

**end if**

---

$t$ , all particles from the posterior distribution at  $t-1$  are propagated to  $t$  via the transition kernel  $K_t$  in (i). Upon receiving measurements and data messages in step (ii), particles are subsequently scored using the likelihood function in (iii) and (iv) in a manner similar to bootstrap particle filtering. The subsequent steps involve particle resampling steps and adaptive control actions. In particular, step (vi) corresponds to computing the empirical variance estimate with respect to the  $\beta_\alpha$  particles for each  $\alpha$ . The resulting quantity  $C_t^{(k)}$  for  $k = 1, 2, \dots, d$  is a Fano factor and it is used in the control step (xi) by scheduling the  $t+1$  physical measurement for the label  $j'$  associated with maximal uncertainty  $j' = \text{argmax}_k \{C_t^{(k)}\}_{k=1}^d$ . Individual calculation steps for our code are fully specified in Sec. 2 of Ref. [39].

Under this adaptive protocol, we seek the convenient convergence properties of particle filtering discussed in earlier sections and we discuss the extent to which particle branching in Algorithm 2 satisfies Proposition 1. The following progression of empirical distributions in Algorithm 2,

$$\bar{\pi}_t^{(j, n_\alpha n_\beta)} \rightarrow \bar{\pi}_t^{(j, n_\alpha)} \rightarrow \pi_t^{n_\alpha}, \quad (30)$$

is found to be a multinomial process similar to conventional particle filtering. This process represents a rearrangement of particle weights into the weights  $\Omega_t^{(j, \alpha)}$  and it forms a multinomial random process if particle number is conserved during each resampling step. To see this, let  $\beta'$  be the labels over all particle pairs  $(\alpha, \beta_\alpha)$  so that these labels correspond to the indices  $\beta' = 1, 2, \dots, n_\alpha n_\beta$ . Let  $A_\alpha$  be the grouping of  $\beta_\alpha$ -particle weights for each  $\alpha$  parent, where  $\alpha$  is the label over parent particles  $1, 2, \dots, n_\alpha$  as before. This means that the labels  $\beta'$  are partitioned into  $n_\alpha$  nonoverlapping categories. Then the weights for each  $A_\alpha$  category are

$$\Omega_t^{(j, \alpha)} := \sum_{\beta' \in A_\alpha} G_t^{(j, \beta')}. \quad (31)$$



The recategorization given by the equation above occurs in (viii) of Algorithm 2 where the weights  $\Omega_t^{(j,\alpha)}$  are proportional to the count over surviving  $\beta_\alpha$  particles for each parent. Additionally, for  $N_1 = N_2 = N$ , the number of offspring in each resampled offspring generation satisfy

$$N = \sum_{\beta'=1}^{n_\alpha n_\beta} \xi_t^{(\beta')} \quad (32)$$

$$= \sum_{\alpha=1}^{n_\alpha} \sum_{\beta' \in A_\alpha} \xi_t^{(\beta')} \quad (33)$$

$$= \sum_{\alpha=1}^{n_\alpha} \eta_t^{(\alpha)}, \quad \eta_t^{(\alpha)} = \sum_{\beta' \in A_\alpha} \xi_t^{(\beta')}. \quad (34)$$

Thus, the resampling steps represent a recategorization of  $\beta_\alpha$  particle weights into nonoverlapping sets associated with each  $\alpha$  parent. If particle number is conserved  $N_1 = N_2 = n_\alpha$ , then two consecutive particle resampling steps in Algorithm 2 are multinomial and satisfy Proposition 1. These observations establish the second result that the adaptive filter of Algorithm 2 shares the same multinomial particle resampling process consistent with traditional particle filtering in Algorithm 1.

The departure of Algorithm 2 from conventional particle filtering lies in the following step of Eq. (25),

$$K_{t-1} \pi_{t-1}^{n_\alpha} \rightarrow \bar{\pi}_t^{(j, n_\alpha n_\beta)}, \quad (35)$$

which requires the generation of  $\beta_\alpha$  particles by creating samples of  $r_t^{(j, \alpha, \beta_\alpha)}$  at the start of each iteration  $t$ . We propose two methods for particle generation: ‘‘Uniform’’ or ‘‘Trunc. Gauss.’’ The Uniform method resets all  $\beta_\alpha$  particles to the initial distribution for  $R_0$  at any  $t$  or parent index  $\alpha$ ,

$$r_t^{(j, \alpha, \beta_\alpha)} \sim \mathcal{U}(\mathbb{S}_R), \quad \forall t, r_t^{(j, \alpha, \beta_\alpha)} \in \mathbb{S}_R. \quad (36)$$

This method represents a strong breakdown of the transfer of estimated state information about  $R_t$  from  $t$  to  $t+1$  during the estimation procedure. In contrast, Trunc. Gauss preserves some information about the estimated  $R_t$  from  $t$  to  $t+1$  for each parent index  $\alpha$ ,

$$r_t^{(j, \alpha, \beta_\alpha)} \sim \mathcal{N}(\bar{r}_t^{(j, \alpha)}, \bar{r}_t^{(j, \alpha)} C_{t-1}^{(j)}), \quad \forall t, r_t^{(j, \alpha, \beta_\alpha)} \in \mathbb{S}_R. \quad (37)$$

Here, one uses the approximation that the true distribution of  $R_t$  at each step can be summarized by the first two moments of a truncated Gaussian distribution. Second, one assumes that  $\bar{r}_t^{(j, \alpha)} C_{t-1}^{(j)}$  is an appropriate approximation for the true second moment of  $R_t$ . As before, the barred quantities  $\bar{r}_t$  denote that the posterior information at  $t-1$  have been propagated into the current  $t$  via the transition kernel  $K_t$ .

The impact of these departures on the convergence properties for particle filtering is now investigated numerically in the next section. In particular, the scaling behavior of true errors with particle number  $n$  will be explored via simulating the specific example in Ref. [22].

## V. NUMERICAL ANALYSIS

In previous sections, we discussed the convergence properties of particle filters. For the specific case of bootstrap particle filters with multinomial resampling, of which both Algorithms

1 and 2 are examples, the condition  $c_t = 1$  means that expected value of distance between the two empirical probability measures in Eqs. (7) and (8) decays as  $\frac{1}{n}$  as  $n \rightarrow \infty$  almost surely, where  $n$  is the particle number. However, Algorithm 2 additionally accommodates both single-qubit measurements (Sec. III) and adaptive control features (Sec. IV) that depart substantially from conventional filtering literature and thus it is not at all clear if the scaling behavior predicted by conventional convergence theory apply here. In this section, we numerically analyze whether the true error scaling behavior of Algorithm 2 with particle number accords with predictions from conventional convergence theory.

Instead of comparing the distance between empirical measures for the state  $X_t$  in Eqs. (7) and (8), in our analyses, we focus on the first moments associated with these empirical measures and compare the posterior estimate of  $F_t$  from the algorithm with the true  $F_t$  using simulations. Under these approximations, let  $\mathcal{L}_t$  be the expected value of the true mean-square error per label  $j$  at iteration  $t$ . From Eq. (8), let  $\mathcal{L}_t$  scale with particle number  $n \equiv n_\alpha$  according to the following postulated relationship,

$$\ln \mathcal{L}_t = \varepsilon_t \ln n_\alpha, \quad (38)$$

where  $\varepsilon_t$  is a real-valued scalar for finite values of  $t$ . Broadly, a value of  $\varepsilon_t < 0$  indicates that expanding the particle number improves the inference procedure (error decreases with greater  $n_\alpha$ ), while a value  $\varepsilon_t > 0$  indicates increased error with  $n_\alpha$ . We expect  $\varepsilon_t \in [-1, 0)$  for an algorithm that accords with conventional convergence theory, where  $\varepsilon_t = -c_t = -1$  holds if multinomial resampling satisfying Proposition 1 is the slowest contribution to overall algorithmic convergence.

For this empirical analysis, we focus on a specific example presented in Refs. [16,22]. In this example, one assumes that a set of independent qubits are subject to a classical, externally applied dephasing field,  $k \in \{1, 2, \dots, d\}$  labels coordinate positions of qubits in two-dimensional (2D) space, and  $s(\cdot)$  is given by a relative phase single-qubit Ramsey measurement. For classical dephasing in Ramsey measurements,

$$s(F_t) \equiv \frac{1}{2} \cos(F_t), \quad \mathbb{S}_F := [0, \pi], \quad (39)$$

where  $F_t$  has the physical interpretation of qubit phases at each location, giving rise to interference between quantum amplitudes of single-qubit states. In Ramsey measurements, the form of  $s(\cdot)$  is nonlinear, bounded, continuous over the half-cycle  $[0, \pi]$ , and this gives rise to a nonlinear particle filtering problem discussed in previous sections. Substituting Eq. (39) into (12) gives the measurement model for single qubits under dephasing:

$$Y_t = \mathcal{Q}\left(\frac{1}{2} \cos(F_t) + v_t + \frac{1}{2}\right), \quad \mathbb{S}_Y = \{0, 1\}. \quad (40)$$

We further assume that a slowly varying  $X_t$  is sampled rapidly using measurement data, and the approximation

$$K_t(x, A) = \Pr[X_{t+1} \in A | X_t = x] = \delta(x) \quad (41)$$

is used, where the symbol  $\delta(x)$  is interpreted as the Dirac delta at  $x$ . The equation above defines the assumption that  $F_t$  is approximately static relative to a high-measurement sample rate in  $t$ .

By taking the first  $d$  elements of the posterior  $X_t$ , and comparing it with a true dephasing field used during simulations,

the error is computed as

$$\mathcal{L}_t := \mathbb{E}[\|(\mathbb{E}_\alpha[f_t^{(\alpha)}] - F^*)\|_2^2/d]. \quad (42)$$

In the above, the true map used in simulations is the vector-valued  $F^*$ ,  $\mathbb{E}[\cdot]$  is an expectation taken over repetitions of each simulation,  $\mathbb{E}_\alpha[\cdot]$  captures the first moment of the particle distribution. Specifically,  $\mathbb{E}_\alpha[f_t^{(\alpha)}]$  are the first  $d$  elements of the posterior estimated  $X_t$  obtained as the mean of the posterior particle positions at  $t$ , and  $\|\cdot\|_2^2$  represents the total squared error on all  $d$  locations.

Using this error metric for the three case studies examined in [22], a plot of the logarithm true mean-square error per qubit against logarithm number of particles yields the estimated slope  $\varepsilon_t$  in Fig. 2. For each case study, the true map  $F^*$  over the arrangement of  $d = 25$  qubit locations is provided in the right insets in both 1D and 2D. In the main panel, we plot the extracted  $\varepsilon_t$  against  $t$  for both Uniform (red crosses) and Trunc. Gaussian (blue circles) expansion strategies. For each value of  $t$ , these  $\varepsilon_t$  values are calculated from the gradient of a line of best fit for the logarithm of true mean-square error per qubit in map reconstruction against the logarithm of  $n_\alpha$ , the total number of  $\alpha$  particles at the beginning and end of each  $t$ . Example raw and best fits for the case  $t = 75$  are shown in the left insets in Figs. 2(a)–2(c).

For  $t \lesssim d = 25$ , we observe  $\varepsilon_t < 0$  for both data sets. This means that increasing  $n_\alpha$  under any expansion strategy (Uniform or Trunc. Gaussian) improves the inference procedure when data are sparse, assuming that the correct initial distribution has been specified. For the high-data regime  $t \gg d$ , the values of  $\varepsilon_t$  diverge between the two expansion strategies. The Uniform approach in Figs. 2(a)–2(c) shows that  $\varepsilon_t > 0$  as  $t$  increases. In contrast, under a Trunc. Gaussian strategy, we see that  $\varepsilon_t \in [-1, 0)$  is satisfied for all values of  $t \gg d$  in all cases studied.

These observations are consistent with our expectations. For the Uniform strategy, we expect  $\varepsilon_t > 0$  since filter convergence does not hold asymptotically as information about  $R_t$  is reset to the prior distribution for  $R_0$  even for large values of  $t$ . In this limit, increasing particle number  $n_\alpha$  increases the level of randomness in the filtering distributions. By contrast, under a Trunc. Gaussian strategy, we expect that in some physical applications, it is reasonable to assume that that length-scale distributions are well described by the first two moments of an appropriately designed truncated Gaussian distribution at each  $t$ . Under these conditions, information transfer from  $t$  to  $t + 1$  occurs such that filter convergence may hold and we expect that the condition  $\varepsilon_t \in [-1, 0)$  is satisfied asymptotically.

Thus, true error scales in a predictable way for Algorithm 2 under a Trunc. Gaussian approach. Our results provide compelling numerical evidence that classical convergence behavior appears to hold even if single-qubit projective measurements and adaptive control features are incorporated into a classical filtering framework. Additional supporting numerical results are provided in the final section of Ref. [39].

## VI. CONCLUSION

In this work, we explore an alternative implementation of adaptive filters for quantum systems with projective measure-

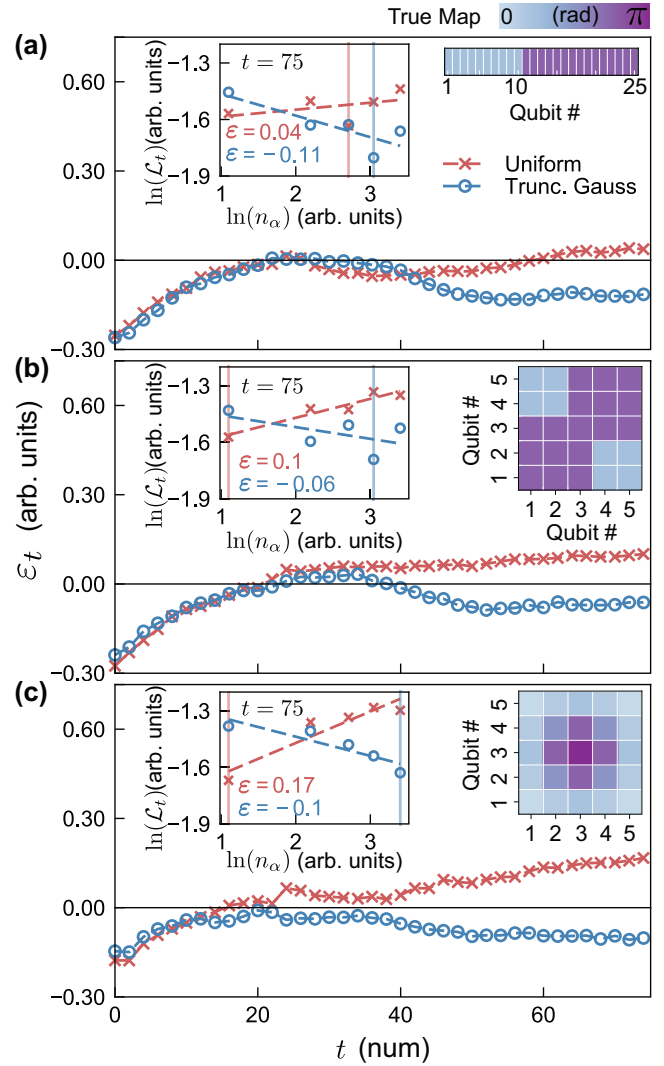


FIG. 2. Error scaling behavior for Uniform and Trunc. Gaussian. Rows represent 1D linear array, a 2D array with a square field, and a 2D array Gaussian field with  $d = 25$  (right insets); with high- and low-qubit phase values of  $0.25\pi$ ,  $0.75\pi$  radians depicted on color scales. (a)–(c) Main panels depict  $\varepsilon_t$  against  $t$  for tuned parameters.  $\varepsilon_t > 0$  for Uniform;  $\varepsilon_t \in [-1, 0)$  for Trunc. Gaussian for  $t \gg d$  agrees with typical convergence analysis. Data for Uniform (red crosses) and Trunc. Gaussian (blue circles). Left insets depict the natural logarithm of the expected mean-square map reconstruction error per qubit over 50 runs against the natural logarithm of  $n_\alpha$  number of  $\alpha$  particles. From left to right, the x axis shows increased particle number  $n_\alpha = 3, 9, 15, 21, 30$ ;  $n_\beta = \frac{2}{3}n_\alpha$ , for  $t = 75$ .  $\varepsilon_t$  is the gradient of the line of best fit (dashed lines). Vertical colored lines mark particle configurations yielding lowest empirical error for tuned parameters  $(\Sigma_v, \Sigma_F, \lambda_1, \lambda_2)$  for Uniform: (a)  $(6.0e^{-9}, 0.10, 0.88, 0.72)$ ; (b)  $(7.1e^{-7}, 0.04, 0.88, 0.72)$ ; (c)  $(5.9e^{-9}, 0.10, 0.72, 0.95)$ . Trunc. Gaussian: (a)  $(9.0e^{-8}, 2.6e^{-5}, 0.88, 0.72)$ ; (b)  $(8.9e^{-7}, 1.9e^{-9}, 0.88, 0.72)$ ; (c)  $(0.77, 4.6e^{-6}, 0.72, 0.95)$ .

ment models and rigorously demonstrate that the theoretical basis of classical nonlinear filtering applies in this context. Taking inspiration from classical signal processing, we combine discrete analysis of continuous amplitude signals with Born's rule and show that a likelihood function can be used

to individually filter a sequence of single-shot projective measurements. While this likelihood function can be incorporated in any classical filtering framework, we show that its inclusion into particle filtering methods preserves important convergence properties of particle-based solutions generalizable to a broad range of difficult inference problems encountered in quantum characterization and control.

Extending these insights, we investigate convergence of classical adaptive filtering of quantum projective measurements. These convergence properties are especially useful if practical implementations limit *a priori* knowledge typically required for filter tuning or training machine-learning methods. Indeed, the technical approach we introduce here for the modification of classical filtering algorithms is generalizable to a wide class of problems as we make minimal assumptions about measurement procedure, noise characteristics, or the dynamics of an open multiqubit system. Applications include adaptive measurement selection [16,22], but other examples could include classical noise spectroscopy, efficient tomography, spatiotemporal forecasting, or adaptive calibration and control tasks using time series of discrete projective measurements.

Focusing on numerical studies for one such example in Ref. [22], an empirical rate of convergence computed as the scaling factor  $\varepsilon_t$  of true error with particle number was shown to be theoretically expected to satisfy the condition  $\varepsilon_t \in [-1, 0)$ . This condition  $\varepsilon_t \in [-1, 0)$  appears to hold for a range of algorithmic and physical configurations under a Trunc. Gaussian particle expansion strategy in a manner similar to convergence properties for conventional particle

filtering. While these numeric studies represent only one type of application of what is a broadly deployable algorithmic framework, the empirical results provide compelling evidence that it may be possible to extend conventional convergence theorems to our methods.

Thus far, we have put forth the idea that effect of quantum projective measurements on classical filtering methods can instead be understood as the effect of a discrete likelihood function on convergence properties of the underlying branching processes. In the case that these branching processes can be viewed as classical random walks, for instance, in classification and regression tree analysis, the insights presented in this paper can be used to appropriately customize alternative stochastic frameworks for predictive control. All of these stochastic methods have wide-ranging implications for device calibration, crosstalk analysis, non-Markovian noise characterization, and automated system tuneup. We look forward to exploring how the rigorous analysis we have performed here may be applied to a broad class of adaptive filtering problems for near-term quantum computers.

Unrestricted access to the code base and data are provided in Ref. [52].

#### ACKNOWLEDGMENTS

R.S.G. would like to thank A. Doherty for extensive discussions. This work was partially supported by the U.S. Army Research Office under Contract No. W911NF-12-R-0012, and a private grant from H. and A. Harley.

- 
- [1] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunsworth, E. Farhi, B. Foxen, A. Fowler *et al.*, Quantum supremacy using a programmable superconducting processor, *Nature (London)* **574**, 505 (2019).
  - [2] D. T. Lennon, H. Moon, L. C. Camenzind, L. Yu, D. M. Zumbühl, G. A. D. Briggs, M. A. Osborne, E. A. Laird, and N. Ares, Efficiently measuring a quantum device using machine learning, *npj Quantum Inf.* **5**, 79 (2019).
  - [3] A. D. Tranter, H. J. Slatyer, M. R. Hush, A. C. Leung, J. L. Everett, K. V. Paul, P. Vernaz-Gris, P. K. Lam, B. C. Buchler, and G. T. Campbell, Multiparameter optimisation of a magneto-optical trap using deep learning, *Nat. Commun.* **9**, 4360 (2018).
  - [4] C. Ferrie, Quantum model averaging, *New J. Phys.* **16**, 093035 (2014).
  - [5] C. E. Granade, Characterization, verification and control for large quantum systems, Ph.D. thesis, University of Waterloo, Canada, 2015.
  - [6] N. Wiebe and C. Granade, Can small quantum systems learn? [arXiv:1512.03145](https://arxiv.org/abs/1512.03145).
  - [7] C. Granade, C. Ferrie, and D. G. Cory, Accelerated randomized benchmarking, *New J. Phys.* **17**, 013042 (2015).
  - [8] C. Granade, C. Ferrie, and S. T. Flammia, Practical adaptive quantum tomography, *New J. Phys.* **19**, 113017 (2017).
  - [9] C. E. Granade, C. Ferrie, N. Wiebe, and D. G. Cory, Robust online hamiltonian learning, *New J. Phys.* **14**, 103013 (2012).
  - [10] M. P. V. Stenberg, Y. R. Sanders, and F. K. Wilhelm, Efficient Estimation of Resonant Coupling Between Quantum Systems, *Phys. Rev. Lett.* **113**, 210404 (2014).
  - [11] S. Kimmel, G. H. Low, and T. J. Yoder, Robust calibration of a universal single-qubit gate set via robust phase estimation, *Phys. Rev. A* **92**, 062315 (2015).
  - [12] K. Rudinger, S. Kimmel, D. Lobser, and P. Maunz, Experimental Demonstration of a Cheap and Accurate Phase Estimation, *Phys. Rev. Lett.* **118**, 190502 (2017).
  - [13] F. Huszár and N. M. T. Houlshby, Adaptive bayesian quantum tomography, *Phys. Rev. A* **85**, 052120 (2012).
  - [14] K. S. Kravtsov, S. S. Straupe, I. V. Radchenko, N. M. T. Houlshby, F. Huszár, and S. P. Kulik, Experimental adaptive Bayesian tomography, *Phys. Rev. A* **87**, 062122 (2013).
  - [15] S. Majumder, L. A. de Castro, and K. R. Brown, Real-time calibration with spectator qubits, *npj Quantum Inf.* **6**, 19 (2020).
  - [16] R. S. Gupta, L. C. G. Govia, and M. J. Biercuk, Integration of spectator qubits into quantum computer architectures for hardware tuneup and calibration, *Phys. Rev. A* **102**, 042611 (2020).
  - [17] H. Carmichael, *An Open Systems Approach to Quantum Optics: Lectures Presented at the Université Libre de Bruxelles, October 28 to November 4, 1991*, Vol. 18 (Springer, New York, 2009).
  - [18] J. M. Geremia, J. K. Stockton, A. C. Doherty, and H. Mabuchi, Quantum Kalman Filtering and the Heisenberg Limit in Atomic Magnetometry, *Phys. Rev. Lett.* **91**, 250801 (2003).

- [19] H. M. Wiseman and G. J. Milburn, *Quantum Measurement and Control* (Cambridge University Press, Cambridge, 2009).
- [20] T. Proctor, M. Reville, E. Nielsen, K. Rudinger, D. Lobser, P. Maunz, R. Blume-Kohout, and K. Young, Detecting and tracking drift in quantum information processors, *Nat. Commun.* **11**, 5396 (2020).
- [21] S. Bravyi, S. Sheldon, A. Kandala, D. C. McKay, and J. M. Gambetta, Mitigating measurement errors in multi-qubit experiments, *Phys. Rev. A* **103**, 042605 (2021).
- [22] R. S. Gupta, C. L. Edmunds, A. R. Milne, C. Hempel, and M. J. Biercuk, Adaptive characterization of spatially inhomogeneous fields and errors in qubit registers, *npj Quantum Inf.* **6**, 53 (2020).
- [23] B. A. Chase and J. M. Geremia, Single-shot parameter estimation via continuous quantum measurement, *Phys. Rev. A* **79**, 022314 (2009).
- [24] C. Granade and N. Wiebe, Structured filtering, *New J. Phys.* **19**, 083014 (2017).
- [25] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age, *IEEE Trans. Robotics* **32**, 1309 (2016).
- [26] H. Durrant-Whyte and T. Bailey, Simultaneous localization and mapping: Part i, *IEEE Robotics Automation Mag* **13**, 99 (2006).
- [27] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics* (MIT Press, Cambridge, MA, 2005).
- [28] A. Doucet, N. De Freitas, and N. Gordon, An introduction to sequential monte carlo methods, *Sequential Monte Carlo Methods in Practice* (Springer, Berlin, 2001), pp. 3–14.
- [29] J. V. Candy, *Bayesian Signal Processing: Classical, Modern, and Particle Filtering Methods*, Adaptive and Cognitive Dynamic Systems (Wiley, Hoboken, NJ, 2016).
- [30] N. Bergman, Recursive Bayesian estimation, Doctoral dissertation, Linköping University, 1999.
- [31] A. Bain and D. Crisan, *Fundamentals of Stochastic Filtering*, Stochastic Modelling and Applied Probability (Springer, Berlin, 2009).
- [32] K. P. Murphy, Bayesian map learning in dynamic environments, *Advances in Neural Information Processing Systems* (MIT Press, Cambridge, MA, 2000), pp. 1015–1021.
- [33] J. Poterjoy, A localized particle filter for high-dimensional nonlinear systems, *Mon. Weather Rev.* **144**, 59 (2016).
- [34] P. Szekeres, *A Course in Modern Mathematical Physics: Groups, Hilbert Space and Differential Geometry* (Cambridge University Press, Cambridge, 2004).
- [35] T. Li, M. Bolic, and P. M. Djuric, Resampling methods for particle filtering: Classification, implementation, and strategies, *IEEE Signal Proc. Mag.* **32**, 70 (2015).
- [36] K. R. Beevers and W. H. Huang, Fixed-lag sampling strategies for particle filtering slam, in *Proceedings 2007 IEEE International Conference on Robotics and Automation* (IEEE, Piscataway, NJ, 2007), pp. 2433–2438.
- [37] G. Grisetti, C. Stachniss, and W. Burgard, Improving grid-based slam with rao-blackwellized particle filters by adaptive proposals and selective resampling, in *Proceedings of the 2005 IEEE International Conference on Robotics and Automation* (IEEE, Piscataway, NJ, 2005), pp. 2432–2437.
- [38] S. Godsill and T. Clapp, Improvement strategies for Monte Carlo particle filters, *Sequential Monte Carlo Methods in Practice* (Springer, Berlin, 2001), pp. 139–158.
- [39] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevA.104.012412> for background on convergence analysis for particle filtering, algorithm implementation details, and additional supporting numerical analysis and it includes Refs. [50,51].
- [40] B. Widrow, I. Kollar, and M.-C. Liu, Statistical theory of quantization, *IEEE Trans. Instrum. Meas.* **45**, 353 (1996).
- [41] S. P. Lipshitz, R. A. Wannamaker, and J. Vanderkooy, Quantization and dither: A theoretical survey, *J. Audio Eng. Soc.* **40**, 355 (1992).
- [42] R. Karlsson and F. Gustafsson, Filtering and estimation for quantized sensor information, Linköping University Technical Report No. LiTH-ISY-R-2674, 2005, p. 14.
- [43] F. Gustafsson and R. Karlsson, Generating dithering noise for maximum likelihood estimation from quantized data, *Automatica* **49**, 554 (2013).
- [44] S. Olmschenk, K. C. Younge, D. L. Moehring, D. N. Matsukevich, P. Maunz, and C. Monroe, Manipulation and detection of a trapped yb+ hyperfine qubit, *Phys. Rev. A* **76**, 052314 (2007).
- [45] S. Ejtemaee, R. Thomas, and P. C. Haljan, Optimization of yb+ fluorescence and hyperfine-qubit detection, *Phys. Rev. A* **82**, 063419 (2010).
- [46] D. Crisan and A. Doucet, A survey of convergence results on particle filtering methods for practitioners, *IEEE Trans. Signal Process.* **50**, 736 (2002).
- [47] T. Bailey and H. Durrant-Whyte, Simultaneous localization and mapping (slam): Part ii, *IEEE Robotics Automation Mag.* **13**, 108 (2006).
- [48] O. Landon-Cardinal, L. C. G. Govia, and A. A. Clerk, Quantitative Tomography for Continuous Variable Quantum Systems, *Phys. Rev. Lett.* **120**, 090501 (2018).
- [49] Y. Ito, Approximation of continuous functions on  $\mathbb{R}^d$  by linear combinations of shifted rotations of a sigmoid function with and without scaling, *Neural Networks* **5**, 105 (1992).
- [50] R. S. Gupta and M. J. Biercuk, Machine Learning for Predictive Estimation of Qubit Dynamics Subject to Dephasing, *Phys. Rev. Appl.* **9**, 064042 (2018).
- [51] M. J. Evans and J. S. Rosenthal, *Probability and Statistics: The Science of Uncertainty* (Macmillan, New York, 2004).
- [52] <http://github.com/qcl-sydney/nmqa>.