Convergence analysis for autonomous adaptive learning applied to quantum architectures

Riddhi Swaroop Gupta* and Michael J. Biercuk
ARC Centre of Excellence for Engineered Quantum Systems, School of Physics, The University of Sydney, New South Wales 2006, Australia

We present a formal analysis and convergence proofs for an autonomous adaptive learning algorithm useful for the tuneup and stabilization of quantum computing architectures. We focus on the specific application of spatial noise mapping in a "spectator" qubit paradigm, in which these qubits act as sensors to provide information useful for decoherence mitigation on proximal data qubits. In earlier work, the authors introduced and experimentally demonstrated this framework, Noise Mapping for Quantum Architectures (NMQA), as applied in autonomous adaptive scheduling of sensor-qubit measurements across multi-qubit architectures [1]. This methodology has several unique features: a classical measurement model that incorporates discretized single-qubit projective measurements, and a two-layered particle filtering structure to facilitate adaptive information sharing between qubits in small spatial regions. In this work, we formalize the NMQA problem definition and build direct links to conventional non-linear filtering theory. Taking into account NMQA's departure from existing literature, we show that NMQA satisfies axioms of theorems for asymptotic convergence in specific parameter regimes. Outside of these parameter regimes, we augment our theoretical analysis with numerical studies to estimate rates of convergence for NMQA. We find that these numerical estimates match our theoretical expectations for a variety of physical configurations. Our work ensures that the methodology encapsulated by NMQA permits comparative analysis with existing filtering techniques in the literature, while highlighting unique directions for future automation of quantum computer tuneup, calibration, and stabilization.

The development of mesoscale quantum computers with many dozens of interacting quantum devices [2] necessitates a move away from manual approaches to tuneup, calibration, and operation. This is relevant across a range of tasks, from identifying performance variations across a chip to characterizing crosstalk between many device pairs. Chief among these remains the task of mitigating sources of decoherence, frequently arising from semi-classical noise processes which manifest as perturbations to a quantum system and its associated dynamics [3]. Moving beyond brute-force device-by-device tuneup and stabilization procedures to achieve quantum advantage for realistic applications requires these issues to be confronted directly; control theory is a natural discipline to which we may turn for inspiration in developing autonomous, adaptive learning techniques.

Focusing on decoherence mitigation, low-level hardware stabilization through sensor information is challenging in the quantum setting in large part due to the problem of quantum state collapse under projective measurement [4, 5], despite its general popularity in classical hardware control. New multi-qubit device architectures have considered the integration of additional qubits as sensors to help mitigate decoherence on nearby data qubits via indirect detection of the perturbing noise process [6]. These sensor qubits yield spatio-temporally correlated projective measurement records from which information about the underlying classical noise process can be inferred using state estimation and filtering techniques [1]. Having access to this information across a device permits "unsupervised" actuation on proximal data qubits used in a computation targeting the mitigation of noise-induced decoherence [6], and can serve as the first step in developing optimized error-robust quantum control techniques [4, 5, 7]. The functionality of this paradigm critically depends on the spatial correlation of noise experienced by the sensor and data qubits, but inhomogeneities in space are regularly apparent due to e.g. interference phenomena or spatial gradients of ambient fields.

Making the spectator-qubit paradigm practically useful - even in medium-scale architectures - requires the development of efficient and autonomous routines to schedule sensor measurements in a way that identifies and maximally exploits the presence of spatial correlations in the underlying noise fields. One specific task in this direction is the spatial mapping of the noise [8] in order to determine which qubits may be actuated upon using information from a specific sensor. This approach is complicated by both the "expense" of qubit measurements, and the limited utility of individual single-shot measurement results arising from projective measurements on qubits. In earlier work, the authors presented a new framework for autonomous learning, denoted "Noise Mapping for Quantum Architectures (NMQA)", to efficiently schedule sensor measurements across a mesoscale device [1]; the general approach bears similarity in its objective to the efficient calibration routines employed in Ref. [2]. In our demonstrations we targeted the efficient (in time or measurement number) reconstruction of a map of unknown spatial fields in multi-qubit arrays, although simple changes allow this approach to be applied to calibration of performance variations across a device or other similar tasks. In any application the al-

* riddhi.sw@gmail.com

gorithm runs in an iterative fashion and in each iteration, NMQA's state estimates are shared with a controller that adaptively selects which qubit to measure next in order to maximise information utility from each physical measurement. Meanwhile, local state information at one qubit location is shared with neighbouring qubits on the array. Compared with brute-force approaches to characterizing a spatial noise field, NMQA showed order-of-magnitude gains in reconstructing a spatially inhomogeneous magnetic field to a target fidelity. Efficiency gains were linked to NMQA's adaptive measurement scheduling and its unique information sharing mechanisms.

NMQA relies on several novel theoretical features that enable efficient measurements scheduling and hence noise mapping in two dimensions. First, NMQA incorporates single-qubit projective measurements as part of its measurement model, enabling this classical algorithm to take as inputs discretized measurement data. Second, NMQA is implemented as a unique two-layer particle filter to perform approximate, iterative likelihood maximization; while particle filters are often used to solve non-linear filtering problems, NMQA departs substantially from conventional particle filters which typically do not have a two-layered structure. Aside from a handful of simultaneous mapping and localization demonstrations in robotics [9–11], an iterative likelihood maximisation approximation is also uncommon in literature. Third, the NMQA information sharing mechanism departs from the typical assumptions in many classical techniques for state estimation and filtering. Thus, the novel algorithmic structure required to adapt NMQA to the problem of measurement scheduling and noise mapping with discretized qubit-based sensors merits deeper quantitative analysis.

In this work, we describe NMQA in the language of discrete-time, stochastic non-linear filtering and present convergence results for our algorithm. We explain the physical intuition for mathematical objects and introduce two major approximations underpinning the NMQA framework. Our key insight is that NMQA satisfies properties for conventional particle filtering at the *beginning* and *end* of an ordered set of computational filtering equations for each iteration t. This means that state of information in NMQA at the start and end of each iteration t can be compared to standard particle filters, even though radically different computations occur in NMQA. We show that in specific parameter regimes the NMQA algorithm behaves in accordance with the properties required for asymptotic convergence while carrying out iterative state estimation from t to $t+1$. Outside of these regimes, we find that it is not clear whether convergence theorems for conventional particle filtering could continue to apply to NMQA. Here, we use numeric simulations to show that the error-scaling behaviour of NMQA still conforms to convergence bounds set by conventional theory across a range of physical device configurations.

The structure of this document comprises of three major sections, as follows. In Section I, we state the physical setting and provide a schematic overview of the NMQA framework. Next, we introduce NMQA as a discrete, non-linear filter and establish technical notation for this manuscript. In Section II, we show how the true NMQA filtering problem is solved approximately through a particle filtering approach. We first demonstrate how we employ particles to represent continuous probability distributions, and second, describe our use of *two* different types of particles to implement iterative likelihood maximization. In particular, we introduce technical details required to manipulate the second layer of particles under NMQA. Collectively, Section I and Section II provide a full mathematical description of the NMQA algorithm. The remainder of the manuscript presents detailed convergence analyses for NMQA drawing heavily from conventional non-linear particle-filtering literature. In Section III, we outline parameter regimes of NMQA that satisfy of axioms of asymptotic convergence theorems. Beyond these parameter regimes, we present numeric evidence of asymptotic convergence. A discussion on the practical operability of NMQA and our conclusions are provided in Section IV.

I. NMQA FRAMEWORK

The framework underlying NMQA enables autonomous adaptive scheduling of measurements on a multi-qubit array, incorporating discretized measurement results and exploiting spatio-temporal correlations in order to gain efficiency. When tailored to noise characterization, NMQA considers the problem of mapping an unknown, classical noise field in 2D using projective measurements performed on an arbitrary spatial arrangement of qubits located within that field. In this section, we first provide an overview of the physical setting and schematically describe how NMQA *adaptively* schedules measurements to efficiently map classical noise fields using far fewer measurements than a brute force measurement strategy. We will establish the true non-linear filtering problem that constitutes the NMQA framework, and in so doing, we will introduce notation and mathematical objects that form the foundation for analyses presented in Section II and Section III.

A. Physical setting and algorithm overview

The physical setting for NMQA is an arbitrary spatial arrangement of d qubits corresponding to a particular choice of hardware. An unknown, classical field exhibiting spatial correlations extends over all qubits on our notional device, and corresponds to ambient noise fields in realistic operating architectures. Our general task is to determine the spatial variation in this noise field as experienced at each qubit, in order to provide useful information to any secondary stabilization scheme (e.g. incoherent measurement feedback) [6].

Each measurement is performed on a single qubit, and

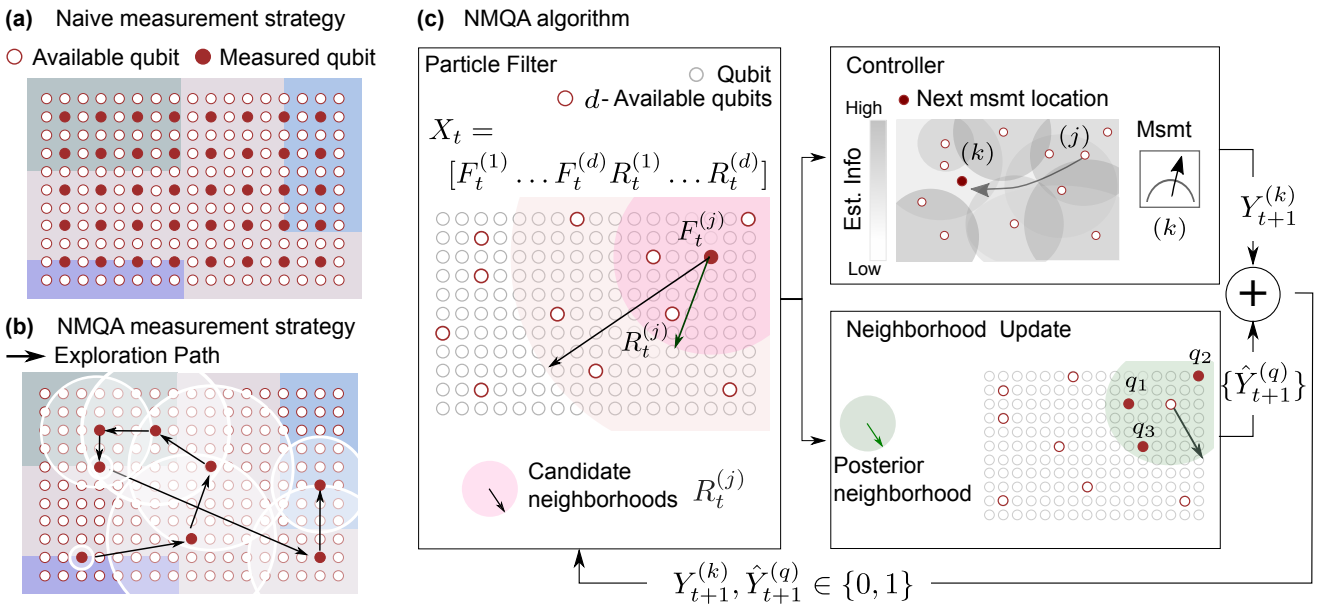

FIG. 1. NMQA framework for inhomogeneous spatial field reconstruction using single qubit measurements (filled red circles) on a 2D array of available qubits (open red circles). (a) Naive brute force measurement strategy distributes measurements uniformly on an array; to be contrasted with NMQA in (b) which selects the next measurement location to maximize information utility for map reconstruction [1]. (c) Schematic overview of a NMQA iteration from t to $t+1$. A particle filter estimates the map, F_t, and discovers neighborhoods (circular shaded) parameterized by $R_t^{(j)}$ for sharing state information about site j. Posterior state estimates from the particle filter are used by the controller, to choose site k as the location of the next physical measurement, $Y_{t+1}^{(k)}$ based on regions of highest estimated uncertainty (middle top). Meanwhile, posterior state estimates are also used to share information at j within the posterior neighborhood, Q via data messages, $\hat{Y}_{t+1}^{(q)}$, between all neighbouring qubits $q \in Q$ (middle bottom) before commencing the next iteration. The input / output information loop is the outermost loop formed by a single physical measurement (in notation, Y) and a set of data messages (in notation, $\hat{Y}$). The arrow to site k from j does not relate to the transfer of any information, but rather, adaptive measurement scheduling by the controller to maximize information utility from the next physical measurement choice k.

contains information about the unknown field at that location; a collection of spatially distributed measurements provides a means to determine the field's spatial structure. For concreteness we consider an ambient magnetic field coupling to the qubit through a Hamiltonian term $\propto \sigma_z$, with σ a Pauli matrix. A single-shot Ramsey-like experiment performed on a single qubit in the presence of the unknown field results in a measurable phase shift between the qubit's basis states at the end of a fixed interrogation period. This phase is not observed directly, but rather depends parameterizes the probability of observing a "0" or "1" outcome in a projective measurement on the qubit depending on the value of the unknown field. Our algorithm takes these discretized, binary measurement results as input data.

The objective we pursue is to build a map of the underlying spatial variation of the noise field with the fewest possible single-qubit measurements, performed sequentially (though the algorithm also can in principle accommodate parallelism). The desired output at any given iteration t is a map of the noise field, denoted as a set of qubit phases, F_t, inferred from the binary measurement record up to t. The NMQA strategy can be compared against a brute-force measurement strategy in which we infer the field using measurements on randomly selected qubits over the entire array, as depicted schematically in panel (a) of Fig. 1.

NMQA employs an adaptive strategy to efficiently build the same underlying noise map by exploiting spatial correlations in the single-shot measurement outcomes, accessible only through a data inference procedure. The algorithm advances iteratively; following each measurement an autonomous controller selects the next measurement location to maximize overall information utility. This approach can yield, depending on operating regimes, order-of-magnitude resource savings in the number of measurements required to reconstruct the noise map to a target fidelity [1].

As depicted in Fig. 1(b), NMQA locally estimates the value of the field for a measured qubit at j, before sharing this information with neighboring qubits in the vicinity of j. The algorithm is responsible for determining the appropriate size of the neighborhood, shown as shaded circular regions in panel (b) of Fig. 1. For each j, the set of neighbouring qubits, Q_t, shrinks or grows as the autonomous inference process progresses. The size of this

neighbourhood and the approximate magnitude of the field gradient over this region is parameterized by a single value, $R_t^{(j)}$, denoting the estimated field-correlation length-scale at the location of the qubit j. The black arrows in (b) depict the sequence of measurements scheduled by the controller, as described below.

Under the NMQA framework, the quantity we wish to infer from measurement data is the true state vector, containing both the register of map values F_t and local approximate map length-scale information, R_t, that is, $X_t = \{F_t, R_t\}$, as shown in Fig. 1 (c). The terminology 'state vector' represents information that is statistically inferred from data, rather than an actual physical quantum state. The role of the particle filter in panel (c) Fig. 1 is to provide an estimate of X_t at t which represents our best knowledge of the map given measurement data. Once all the particle-filtering calculations are completed for an iteration t, the state X_t provided by the particle filter is called the posterior state; the true uncertainty associated with this state is called the posterior-state variance.

The posterior X_t is passed to an autonomous measurement scheduler, the NMQA controller, which attempts to maximize the information utility from each measurement. The NMQA controller adaptively selects the location k for the next physical measurement by choosing the region of the map where posterior state variance is maximally uncertain (Fig. 1 (c), top-right panel); the new measurement outcome, once collected, is denoted $Y_{t+1}^{(k)}$. Meanwhile, the posterior state information at iteration t is also shared with the set of qubits Q_t in the vicinity of qubit j (Fig. 1 (c), bottom-right panel). The shared information within this neighborhood is denoted by the set $\{\hat{Y}_{t+1}^{(q)}, q \in Q_{t+1}\}$, and serves as an effective state estimate that is taken as an input to the algorithm in a manner similar to a physical measurement. Jointly, the new actual physical measurement, $Y_{t+1}^{(k)}$, and the set of shared information $\{\hat{Y}_{t+1}^{(q)}\}$ form the inputs for the next iteration of the algorithm at $t+1$. Through this iterative procedure it becomes possible to infer the spatial features of the underlying noise field with greater efficiency than a brute-force approach.

B. Casting NMQA as nonlinear filtering

Building on our schematic overview, we now present NMQA using the technical nomenclature of classical nonlinear filtering. Introducing these linkages provides an important foundation for constructing numerical solutions via particle approximations in Section II, and establishing axioms for convergence in Section III.

We proceed as follows. First, NMQA is a 'filter' because we wish to infer a hidden state X from a set of observations Z. Second, NMQA is a *non-linear* filter, since the measurement model for Z incorporates single-qubit projective measurements, and this model is not linearizable. In general, we do not work directly with the state X, but rather, with the conditional distribution of X given Z, denoted π. The measure π can be conceived mathematically as a single point in the space of all possible probability measures. In the materials below, we will introduce random processes relevant to the NMQA filtering problem (such as X and Z) as well probability measures (such as π) associated with these random processes.

For NMQA, the true state vector at iteration t is given by $X_t = (F_t, R_t) \in \mathbb{S}_X := \mathbb{S}_F \times \mathbb{S}_R$ and the notation $\mathbb{S}$ refers to the state-space of the respective variable. The quantities F_t and R_t represent d-dimensional, real, continuous, random variables, and their outcomes take values between $F_t^{(j)} \in [0, \pi]$ radians and $R_t^{(j)} \in [R_{min}, R_{max}]$ with units of distance, for any qubit location labelled by the index j. The 'filtration' generated by the process X is a family of σ-fields at each iteration t, where each σ-field represents all subsets (Borel sets) of X_t which can be assigned probabilities under probability theory. In notation, a filtration is given by $\mathcal{F}_t := \sigma(\{X_s, s \in [0, t]\})$ and these filtrations are used to formally delineate the different types of conditional probability distributions used in particle filtering analysis. We denote $X_{0:t} := (X_0, \ldots, X_t)$ as the set of state variables until t.

The observation vector at t consists of a single physical measurement performed at location j and shared data messages between qubits, namely, $Z_t = (Y_t^{(j)}, \{\hat{Y}_t^{(q_t)}\}_{q_t \in Q_t})$. Here, the set of messages $\{\hat{Y}_t^{(q_t)}\}_{q_t \in Q_t}$ refers to the posterior neighborhood Q_{t-1} about j_{t-1}. The full observational vector is $Z := \{Z_t, t = 1, 2, \ldots\}$ and $Z_{0:t} := (Z_0, \ldots, Z_t)$ denotes the measurement record until t.

We will now establish the link between vectors X and Z and the measure π_t, using the language of non-linear filtering in discrete time established in [12], as this notation appears in foundational theory for particle filtering [13–15]. In the absence of the controller, the filtering problem is to compute the specific measure π_t that is the conditional probability of X given the σ-field generated by the observation process $Z_{0:T}$:

$$\pi_t := \mathbb{P}[X_t \in A | \sigma(Z_{0:T})], \quad \forall A \in \mathcal{S}_X \tag{1}$$

$$\pi_t f = \mathbb{E}[f(X_t) | \sigma(Z_{0:T})] \quad \forall f \in B(\mathbb{S}_X), A \in \mathcal{S}_X \tag{2}$$

$$\pi_0 \sim \mathcal{U}(\mathbb{S}_X) \tag{3}$$

In the relations above, π_t is identified with the posterior distribution in Bayesian analysis and $\mathbb{P}[\cdot]$ denotes a probability measure. Since X_t and Z_t are random, the specific π_t is a random measure. The appropriate state space is given by $\mathbb{S}_X$ and $\mathcal{S}_X$ is the Borel σ-field generated by $\mathbb{S}_X$. The notation $X_t \in A$ means that some instance of X_t is an event in the Borel σ-field generated by the state space of X. Similar comments apply to the σ-field generated by the observation vector. In the second line, the term $B(\mathbb{S}_X)$ refers to a space of bounded, $\mathcal{S}_X$-measurable functions which correspond to transformations of the state

Sym.	Definition
$\mathbb{S}$	A complete, separable metric (state) space for a R.V.
$\mathcal{S}$	The Borel σ-algebra [16] generated by $\mathbb{S}$
$C(\mathbb{S})$	The space of real continuous functions on $\mathbb{S}$
$M(\mathbb{S})$	The space of $\mathcal{S}$-measurable functions on $\mathbb{S}$
$B(\mathbb{S})$	The space of bounded $\mathcal{S}$-measurable functions on $\mathbb{S}$
$C_b(\mathbb{S})$	The space of bounded continuous functions on $\mathbb{S}$
$P(\mathbb{S})$	The space of probability measures on $(\mathbb{S}, \mathcal{S})$ s.t. $\mu \in P(\mathbb{S})$ satisfies $\mu(\mathbb{S}) = 1$

TABLE I. State, function and measure space notation (consistent with [12]). The abbreviation R.V. stands for any random variable and $\mathcal{S} \equiv \sigma(\mathbb{S})$ i.e. the σ-field generated by the state space $\mathbb{S}$.

(e.g. dynamical evolution of X_t, measurement model for X_t). Thus, $\mathbb{E}[\cdot]$ refers to an expectation of a random variable or expectation of bounded, Borel-measurable functions of random variables. In the last line, the notation $\mathcal{U}$ conveys that the initial condition π_0 is a uniformly distributed probability measure on the state space of X. A list of useful definitions are summarized in Table I.

Under a typical filtering framework, we define the measure p_t, a 'predictive probability measure'. The word 'predictive' is used to capture the idea that true state at any t is conditional on its observed history until $t-1$:

$$p_t := \mathbb{P}[X_t \in A | Z_{0:t-1}], \quad \forall A \in \mathcal{S}_X \tag{4}$$

$$p_t f = \mathbb{E}[f(X_t)|Z_{0:t-1}] \quad \forall f \in B(\mathbb{S}_X), A \in \mathcal{S}_X \tag{5}$$

As such, the predictive probability distribution captures assumptions about the dynamics of X. Along with π_t, p_t represents another point in measure space. This general framework permits us to incorporate time-varying noise maps, although we do not consider them explicitly in this manuscript.

The predictive probability measure p_t is then combined with a function, g_t, which captures the essential consequences of the measurement model for X. When g_t describes measurement noise probability density, it is often called the 'likelihood' function under certain conditions [12]. We will adopt this convention for ease of reading and refer to g_t as the likelihood function. The resulting product of p_t and g_t yields the true measure π:

$$\pi_t := g_t * p_t(A) := \frac{\int_A g_t(x) dp_t(x)}{p_t g_t} \tag{6}$$

$$p_t g_t := \int_{\mathcal{S}_X} g_t(x) dp_t(x) > 0$$

The use of the projective product in the first line, $*$, is essentially a restatement of Bayes rule. The integral in the numerator exists only over the outcomes A, whereas the denominator represents a normalization over all outcomes in the state space of X. In this framework, $g_t \in B(\mathbb{S}_X)$ is free to be any non-negative density function that describes the density of the random observation vector, Z_t.

Having defined the two measures, p_t and π_t, one establishes a recursive relation for π_t using an additional Markov assumption for X_t. A Markov assumption has the effect that the conditional on all values, $[0, t]$, is not required in analysis and the NMQA algorithm can process measurement information iteratively. We define the initial (prior) distribution π_0 for the initial state X_0, and a model, K_t, to govern state transitions from t to $t+1$. The recursion for π_t is obtained by substituting the following expression into the projective product:

$$p_t := K_{t-1}\pi_{t-1}, p_t \in P(\mathbb{S}_X) \tag{7}$$

$$K_t := \mathbb{P}[X_{t+1} \in A | X_t = x] \tag{8}$$

$$K_t(x, A) : \mathbb{S}_X \times \mathcal{S}_X \to P(\mathbb{S}_X) \times B(\mathbb{S}_X), \quad \forall t = 0, 1, \ldots, A \in \mathcal{S}_X, x \in \mathbb{S}_X. \tag{9}$$

In the above, the transition kernel $K_t := K_t(x, A)$ satisfies the property that $K_t(\cdot, A) \in B(\mathbb{S}_X)$ is a bounded Borel-measurable function for any $A \in \mathcal{S}_X$, and $K_t(x, \cdot)$ is a probability measure over all possible final states at $t+1$ if $X_t = x$ (Markov assumption). Substitution of Eq. (7) into the projective product can be used to show that π_t satisfies recursion relations under a strong convergence law, namely, with probability one [12]. For simplicity, we assume time-invariant maps in this manuscript and set the transition kernel to be an identity matrix, $K_t := \mathbb{I}$.

The observation record $Z_{0:t}$ is a random observation vector. In any single experimental run, we collect data by measuring instances of the true random process $Z_{0:t}$ *i.e.* we obtain the 'fixed' realization of the observation record, $z_{0:t} := \{Z_0 = z_0, Z_1 = z_1, \ldots, Z_t = z_t\}$. The lowercase z_t represents instances of the true random variate Z_t. This means that once data, $z_{0:t}$, has been collected for a single experimental run, the measures $\pi_t^{z_{0:t}}$ and $p_t^{z_{0:t}}$ can be computed using this data. For any fixed path $(z_{0:t})$, these computations will in-principle yield non-random quantities for $\pi_t^{z_{0:t}}$ and $p_t^{z_{0:t}}$. Thus, $\pi_t^{z_{0:t}}$ and $p_t^{z_{0:t}}$ should be distinguished from the true theoretical quantities π_t, p_t, which are *random measure sequences* revealed only through many experimental runs, and depend on the random observation record *i.e.* $\pi \equiv \pi_t^{Z_{0:t}}, p_t \equiv p_t^{Z_{0:t}}$. These distinctions will be used in the particle-filtering convergence analysis presented in the final section of this manuscript.

Thus far, we have linked the output of NMQA with a true point in measure space, π, and we introduced recursive filtering equations for the measure π. We have also established notation for key mathematical objects in the NMQA algorithm associated with filtering. The discussion of NMQA so far has been generalized and abstracted away from the physical system under consideration. In the next subsection, we bring physical intuition into our discussion of NMQA by introducing a specific measurement model.

C. Measurement model in NMQA

In general filtering theory the transformation of the hidden state X to the observations Z is known as the measurement model. The measurement model explicitly captures the connection between raw data and the quantity which we are trying to infer. In practice, measurement models do not influence an inference procedure directly, but rather these models govern the mathematical form of the probability density of additive measurement noise, described by likelihood functions. Here, the concept of measurement noise captures an uncertainty, often expressed as a difference between observed data, Z, and the estimated output generated by applying our measurement model to the hidden state, X (assuming both that X is known and our model of the measurement model is correct).

Our global likelihood function, g_t, is a product of two different likelihood functions, incorporating both the physical measurement model and the information sharing mechanism. We also introduce two free parameters of g_t in the NMQA framework, $\lambda_1, \lambda_2 \in [0, 1]$. Physically, the λ_1, λ_2 parameters regulate the influence that the sharing mechanism has on the overall NMQA inference procedure, where $\lambda_1, \lambda_2 = 0$ ensures that no information sharing takes place and NMQA effectively reduces to a brute-force measurement procedure [1]. Below, greater technical detail about the likelihood functions for physical measurement and the shared data-messages are discussed in turn.

The physical measurement model for our state vector incorporates a projective single-qubit measurement. For concreteness we determine that each projective measurement will follow a Ramsey interrogation period; for a given location j and iteration t, the measurement model is then formally expressed

$$Y_t^{(j)} := \mathcal{Q}\left(\frac{1}{2}\cos(F_t^{(j)}) + v_t + \frac{1}{2}\right) \tag{10}$$

Here, $\mathcal{Q}(\cdot)$ represents a single 'coin-toss' experiment yielding a '0' or a '1', where the bias on the coin is given by the argument of $\mathcal{Q}(\cdot)$. In the equation above, this bias depends on the map-value $F_t^{(j)}$ and a zero-mean white Gaussian noise term, v_t with Σ_v variance for all t. In principle the argument of $\mathcal{Q}(\cdot)$ can be an arbitrary interrogation procedure which is then followed by projective measurement. The resulting observation has a probability distribution resembling that obtained from enforcing a symmetrically truncated Gaussian error model for v_t. This measurement model leads to a likelihood function given by:

$$g_1(\lambda_1, Y_t^{(j)}) := \frac{\rho_0}{2} + \frac{\rho_0 \cos(F_t^{(j)})}{2}\Big(\delta(Y_t^{(j)} - 1) - \delta(Y_t^{(j)})\Big) \tag{11}$$

$$\rho_0 := \mathrm{erf}(\frac{2b}{\sqrt{2\Sigma_v}}) + \frac{\sqrt{2\Sigma_v}}{2b}\frac{e^{-(\frac{2b}{\sqrt{2\Sigma_v}})^2}}{\sqrt{\pi}} - \frac{1}{2b}\frac{\sqrt{2\Sigma_v}}{\sqrt{\pi}} \tag{12}$$

$$b := 1/2 \tag{13}$$

In the likelihood function above, the parameters ρ_0, b are set by the properties of measurement noise when sensor data is only allowed to take discrete binary values. The rationale for Eq. (10) and the form of Eqs. (11) and (12) are derived from a study of amplitude-quantized sensor error in classical signal processing, as detailed in the Appendix. The prior for F_0 is uniformly distributed.

We will now describe how data-messages are created and shared between qubits within a local neighbourhood. Given a map F_t, we associate a likelihood function for the length-scale at j, $R_t^{(j)}$, that scores the most appropriate length-scale for sharing information between qubits. Both the size of the neighbourhood for a qubit at j and the data-message generation from j to its neighbours $q \in Q_t$ is parameterized by this length-scale. In notation, the data message sent to qubit q at the beginning of $t+1$ (posterior at t) due to the physical measurement at j and t is the random variable $\hat{Y}_{t+1}^{(q_{t+1})} \in \{0, 1\}$, generated using the model:

$$\hat{Y}_{t+1}^{(q_{t+1})} := \mathcal{Q}(\frac{1}{2}\cos(\mathcal{X}_{q_t}) + \frac{1}{2}) \tag{14}$$

The shared information, $\mathcal{X}_{q_t}$, uses a weighted average of both posterior map estimate on the neighbouring qubit and posterior state information for the physically measured qubit at j, mediated by a factor $\lambda_2^{\tau_{q_t}}$. The quantity $\mathcal{X}_{q_t}$ will be introduced below and it is to be interpreted as the posterior Eq. (19), where $q_{t+1} \in Q_{t+1}$ (above) equals the Q_t set by posterior state estimate $R_t^{(j)}$ and the posterior F_t is used for all calculations.

The length-scale parameterizes a Gaussian function with amplitude $F_t^{(j)}$ that spreads the point-value of the phase $F_t^{(j)}$, at j, into a small neighbourhood. Let $\nu_{(j,q)}$ be the separation distance between the qubit at j and a point q in a neighbourhood around j. Then the smeared phase value is given by:

$$\bar{F}(\nu_{(j,q)}) := F_t^{(j)} \exp\left(\frac{-\nu_{(j,q)}^2}{(R_t^{(j)})^2}\right) \tag{15}$$

The point value of the phase is regained in the limit $R_t^{(j)} \to 0$, and the smallest inter-qubit separation distance sets the value of R_{min}. The neighbourhood, Q_t of the measured qubit at j with length-scale $R_t^{(j)}$ is the set:

$$Q_t := \{q_t | \nu_{(j,q_t)} \leq k_0 R_t^{(j)} \quad \forall q_t \in \{1, 2, \ldots d\} \setminus \{j\}\}, \quad k_0 \in [1, \infty). \tag{16}$$

That is, 'neighbors' are defined as the set of qubits whose separations fall within a radius of $k_0 R_t^{(j)}$ about the location j. Here, k_0 is an arbitrary constant that truncates the neighbourhood when smeared phase values at the boundary of the neighborhood are dissimilar to those at the center; a conservative approach sets $k_0 = 1$. The index q_t denotes the location labels for all qubits excluding the measured qubit, $q_t \in \{1, 2, \dots d\} \setminus \{j\}$.

For a constant $\lambda_2 \in [0, 1]$, and a non-negative natural number τ_{q_t}, the likelihood for the length-scale at j compares the estimate $\mathcal{X}_{q_t}$ to the best available phase information on each neighbour:

$$g_2(\lambda_2, Q_t) := \prod_{q_t \in Q_t} \frac{1}{k_1 \sqrt{2\pi\Sigma_F}} \exp\left(-\frac{(F_t^{(q_t)} - \mathcal{X}_{q_t} - \mu_F)^2}{2\Sigma_F} \right) \tag{17}$$

$$F_t^{(q_t)} = \mathcal{X}_{q_t} + w_t, \quad \forall q_t \in Q_t \tag{18}$$

$$\mathcal{X}_{q_t} := (1 - \lambda_2^{\tau_{q_t}}) F_t^{(q_t)} + \lambda_2^{\tau_{q_t}} \bar{F}(\nu_{(j,q_t)}) \tag{19}$$

$$w_t \sim \mathcal{N}(\mu_F, \Sigma_F) \tag{20}$$

$$k_1 := \frac{1}{2}\left(\mathrm{erf}(\frac{\pi + \mu_F}{\sqrt{2\Sigma_F}}) + \mathrm{erf}(\frac{\pi - \mu_F}{\sqrt{2\Sigma_F}}) \right) \tag{21}$$

This likelihood function determines whether a proposed length-scale, $R_t^{(j)}$, is viable, given that the map F_t is known. The viability of a length-scale is assessed by comparing the proposed quantity $\mathcal{X}_{q_t}$, with the known map F_t, over all neighbours, $q_t \in Q_t$. The number of total neighbours in Q_t depends on the proposed candidate for $R_t^{(j)}$. The proposed quantity $\mathcal{X}_{q_t}$ is based on a weighted average of the most recent map value on the neighbouring qubit and $\bar{F}(\nu_{(j,q_t)})$ generated by the physically measured qubit at j, mediated by $\lambda_2^{\tau_{q_t}}$. The noise parameters μ_F, Σ_F represent the true error in approximating a continuously varying spatial field with overlapping Gaussian functions. The error term is a scaled, truncated Gaussian random variable. Error values lie in the finite interval $[-\pi, \pi]$, giving rise to the constant k_1. The scaling factor $\lambda_2^{\tau_{q_t}}$ ensures that the effect of $\bar{F}(\nu_{(j,q_t)})$ is ignored as the number of physical measurements accumulate at the location q_t; the constant $\tau_{q_t} \leq T$ is the tally of the total number of times q_t has been physically measured. The prior for R_0 is uniformly distributed.

Thus, the estimated map value at j, $F_t^{(j)}$, depends on previous physical measurements performed on qubit j up to iteration t, and on previous data messages received by qubit j. The extent to which the map estimate at a physically-measured qubit j (with neighbouring qubit q) incorporates these data messages, depends on the parameter λ_1 in Eq. (30) in Section II (similarly, λ_2 in Eq. (19)). This means that λ_1, λ_2 can be numerically tuned to determine how NMQA performance varies along the continuum between no information sharing ($\lambda_1, \lambda_2 \to 0$) and equivalence between data messages and physical measurements ($\lambda_1, \lambda_2 \to 1$).

The essence of any numerical solution to the NMQA filtering problem will require the introduction of a set of empirical measures to approximate the true measure π_t. The manner in which NMQA manipulates these empirical measures will utilize the likelihood functions discussed above. We will use a unique, two-layered particle filter to approximately solve the NMQA filtering problem as presented in the next section, along with numeric tuning of the parameters λ_1, λ_2.

II. NMQA PARTICLE FILTER

Particle-filtering is a class of numerical approximation methods useful for solving non-linear filtering problems. The objective of a particle filter is, via numeric evolution, to asymptotically converge to the true measures, π_t, p_t, as total data (t) and the total number of particles (n) increase. In particle-filtering implementations, the posterior distribution π_t is approximately represented as a discrete collection of weighted 'particles', each representing information about a hypothesis for X_t for some true measure (particle position), as well an estimate of that hypothesis's likelihood or importance in the estimation procedure (particle weight). All weights are initially equal at $t = 0$ and after receiving measurement data, the particles are 're-sampled'. This means that the original set of particles at t is replaced by a set of 'offspring' particles, where the probability that a parent is chosen to represent itself in the next iteration (with replacement) is directly proportional to its weight. Over many iterations, only the particles with the highest weights survive and these surviving particles form the estimate of the true X_t in our algorithm. The mechanism by which parent particles are replaced by offspring particles is called a 'branching' mechanism; many particle branching methods have been proposed in literature [12, 13, 17].

The use of particles to approximate a continuous probability distribution necessitates that a particle filter employs empirical measures to approximate the true π_t, p_t. These approximate measures should be thought of as a random measures that form a sequence, indexed by the number of particles, n, and the limit of this sequence is the true measure for any choice of t. We write these approximate measures as π_t^n, p_t^n, to be interpreted as approximations to the true π_t, p_t using a finite n. Our approximate measures, π_t^n, p_t^n, are random, and this randomness comes from *two* sources: first, the observation vector, Z_t; second, *additional* randomness from the particle branching mechanism of the algorithm at each t.

Below, we define the approximate measures π_t^n, p_t^n in terms of particles inside the filter. The notation, x_t, represents a sample from a probability measure at t, and constitutes the position of the particle while i is the particle index. The Kronecker delta, $\delta(\cdot)$, is used because the approximate probability measures represent discrete

probability distributions:

$$\pi_0^n := \frac{1}{n}\sum_{i=1}^{n}\delta(x_0^{(i)}), \quad x_0^{(i)} \sim \pi_0 \tag{22}$$

$$p_t^n := \frac{1}{n}\sum_{i=1}^{n}\delta(\bar{x}_t^{(i)}), \quad \bar{x}_t^{(i)} \sim K_{t-1}(x_{t-1}^{(i)}) \tag{23}$$

$$\pi_t^n := \frac{1}{n}\sum_{i=1}^{n}\delta(x_t^{(i)}) \tag{24}$$

In the above, π_0^n is the particle approximation to the true prior, since the particles $x_0^{(i)} \sim \pi_0$ are sampled from the true initial distribution, π_0. These particles are evolved from t to $t+1$ using K_t and this yields the predictive distribution p_t^n. The output of the particle filter at any t, is π_t^n, and by definition this measure is the collection of particles $\{x_t^{(i)}\}$.

Within this structure, we now wish to incorporate the effect of receiving a measurement. We input measurement data into the likelihood function g_t to weight each particle; a high particle weight, $w_t^{(i)}$ results in a high probability of receiving this measurements result for the i-th particle given the true X_t. Raw weights for each particle are empirically normalized across n particles for each t. Hence, we obtain the weighted empirical distribution, $\bar{\pi}$, as:

$$\bar{\pi}_t^n := \sum_{i=1}^{n} w_t^{(i)}\delta(\bar{x}_t^{(i)}), \;\; \bar{x}_t^{(i)} \sim K_{t-1}(x_{t-1}^{(i)}) \tag{25}$$

The $\bar{x}_t^{(i)}$ indicates that $\bar{\pi}_t^n$ should be computed after evolving particles from $t-1$ into the current iteration at t, and $w_t^{(i)}$ are calculated based on a single measurement, Y_t, received at t.

In Fig. 2, we schematically represent how these measures are implemented by the NMQA particle filter. The empirical predictive distribution, p_t^n, brings the posterior from $t-1$ into t via the dynamical state evolution described by the transition kernel (column (i) in Fig. 2). The particle weights, $w_t^{(i)}$ for the weighted empirical measure, $\bar{\pi}_t^n$, are calculated from the likelihood function g_t in columns (iii) to (v) of Fig. 2. These empirical weights for both particle layers are regrouped in columns (v) to (viii) into a single layer at the end of each iteration t. The empirical posterior is given by π_t^n at the end of each iteration t (column (x) in Fig. 2).

In order to suit the needs of NMQA, we implement our particle filter using *two* different types of particles. These are required so that NMQA can numerically implement an 'iterative maximum likelihood procedure', simultaneously estimating the qubit state as well as the optimal neighborhood for information sharing, based on discretized measurements, discussed below.

A. Two-layer particle filter and iterative maximum likelihood

We rely on an 'iterative maximum likelihood procedure' within each iteration of the particle filter to solve the NMQA inference problem, as the size of the state-space generally impedes analytic solutions even in classical settings [10, 11]. The 'iterative' component of the procedure refers to the assumption that the state vector is known at the previous iteration $t-1$ before proceeding to t, significantly reducing the size of the state-space to be explored. Meanwhile, the 'maximum likelihood' component refers to the concept of using likelihood functions to score both types of particles inside NMQA.

In each iteration of NMQA, we update F_t assuming X_{t-1} is known, and subsequently update R_t based on F_t. As introduced above, the structure of these computations requires two particle types; α particles carry information about the full state vector, X_t, while β particles discover the optimal information-sharing neighborhood size, $R_t^{(j)}$, around qubit j. These two different types of particles are then used to manipulate the joint probability distribution defined over F_t and R_t.

The two layers of the particle filter are structured as follows. We label α-particles by the set of numbers $\{1, 2, \ldots, n_\alpha\}$ (column (iii) in Fig. 2). For each α particle, we also associate a set of β particles denoted $\beta^{(\alpha)}$, with β taking values $\{1, 2, \ldots, n_\beta\}$ (column (iv) in Fig. 2), yielding an exponential scaling of total particles. Strategies for generating β-particles, *i.e.* the expansion of the particle tree in column (iii) of Fig. 2, can be designed either to inject information about $R_t^{(j)}$ sampled from the prior distribution, or to use posterior information about R_{t-1} from the previous iteration. These strategies will be analyzed in Section II B. As a consequence, a single $\beta^{(\alpha)}$-particle inherits the state from its α-parent, but additionally acquires a single sample for $R_t^{(j)}$.

In our iterative maximum likelihood procedure we associate the functions, h_1 and h_2 (commonly known as 'data association mechanisms' [10, 11]), with α and β particles respectively. The full iterative maximum likelihood update is then captured as a set of recursive equations which conform to a Markov description of our state space. In the equations below, $h_1(\lambda_1, Z_t)$ updates the state variable $F_t^{(j)}$ given X_{t-1}, and h_2 updates the state variable $R_t^{(j)}$ given F_t. The h_1, h_2 are local and should be interpreted as identity for state estimates at other lo-

cations, $j' \neq j$:

$$F_t^{(j)} := \begin{cases} F_{t-1}^{(j)}, & \tau_t^{(j)}, \varphi_t^{(j)} = 0 \\ h_1(\lambda_1, Z_t), & \text{otherwise} \end{cases} \tag{26}$$

$$R_t^{(j)} := h_2(F_t) \tag{27}$$

$$h_1(\lambda_1, Z_t) := \cos^{-1}(2P_t^{(j)}(\lambda_1, Z_t) - 1) \tag{28}$$

$$h_2(F_t) = \mathbb{E}[\{R_t^{(j),n}\}_{n=1}^{n_\beta} | F_t] \tag{29}$$

$$P_t^{(j)}(\lambda_1, Z_t) := \begin{cases} \left(1 - \frac{\lambda_1^{\tau_t^{(j)}}}{2}\right)\kappa_t^{(j)} + \left(\frac{\lambda_1^{\tau_t^{(j)}}}{2}\right)\gamma_t^{(j)}, \\ \text{for } \tau_t^{(j)}, \varphi_t^{(j)} \neq 0. \\ \kappa_t^{(j)}, \text{ for } \tau_t^{(j)} \neq 0, \varphi_t^{(j)} = 0. \\ \gamma_t^{(j)}, \text{ for } \tau_t^{(j)} = 0, \varphi_t^{(j)} \neq 0. \end{cases} \tag{30}$$

The first data association mechanism, $h_1(\lambda_1, Z_t)$, updates map values via the quantity $P_t^{(j)}$. Here, $P_t^{(j)}$ is defined as an empirical Born probability estimate obtained by averaging the binary measurement data obtained at a single qubit, though in principle it may reflect any measurement strategy. The binary measurement data is averaged over $\tau_t^{(j)}$ physical measurements at qubit j, and $\varphi_t^{(j)}$ shared messages induced by measuring the neighboring qubits of j. The role of shared messages is reduced as the number of physical measurements, $\tau_t^{(j)}$, increases, at a rate governed by the choice of λ_1. For the Ramsey interrogation assumed in this specific application, the form of $h_1(\lambda_1, Z_t)$ holds if the uncertainty in our knowledge of the qubit phase, v_t, is represented by a zero mean Gaussian white noise source. Under this assumption, one updates the estimate for F_t using $h_1(\lambda_1, Z_t)$ and invoking Born's rule for a single Ramsey measurement.

The second data association mechanism h_2 locally updates the length-scale value $R_t^{(j)}$ as the expectation over re-sampled β-particles given the state at F_t. The dependence of h_2 on F_t is implicit, as β-particles are weighted via the $g_2(\lambda_2, Q_t)$ likelihood function for which it is assumed F_t is known. The counters for physical measurements, $\tau_t^{(j)}$, and shared data messages, $\varphi_t^{(j)}$, are recursively updated. The empirical averages of this data are maintained by the recursive means, κ_t and γ_t respectively:

$$\kappa_t^{(j)} := \frac{\tau_{t-1}^{(j)}}{\tau_t^{(j)}}\kappa_{t-1}^{(j)} + \frac{1}{\tau_t^{(j)}} Y_t^{(j_t)} I_{(j_t=j)}, \quad \tau_t^{(j)} \neq 0,\ \kappa_0^{(j)} = 0 \tag{31}$$

$$\gamma_t^{(j)} := \frac{\varphi_{t-1}^{(j)}}{\varphi_t^{(j)}}\gamma_{t-1}^{(j)} + \frac{1}{\varphi_t^{(j)}} \hat{Y}_t^{(q_t)} I_{(q_t=j, q_t \in Q_t)}, \quad \varphi_t^{(j)} \neq 0,\ \gamma_0^{(j)} = 0 \tag{32}$$

$$\tau_t^{(j)} := \tau_{t-1}^{(j)} + I_{(j_t=j)}, \quad \tau_0^{(j)} = 0 \tag{33}$$

$$\varphi_t^{(j)} := \varphi_{t-1}^{(j)} + I_{(q_t=j, q_t \in Q_t)}, \quad \varphi_0^{(j)} = 0 \tag{34}$$

Here, I_A is the indicator function equalling unity if the statement A is true, and zero if the statement A is false. Fig. 2 encompasses the order in which the mathematical equations for NMQA should be implemented, taking into account the iterative maximum likelihood equations depicted above.

Establishing a joint distribution over both particle layers is the first step in the branching mechanism for NMQA, as depicted in Fig. 2 columns (v). The first re-sampling procedure samples particles from the joint distribution in (v) to give offspring particle pairs. Low probability (α, β) particle pairs are lost as schematically depicted by a tree structure that is non-uniform; two examples of uneven trees are illustrated in column (vii). After re-sampling, particles regain equal weights. We take the mean and the variance of length scales across the β layer for each α particle and store them as the posterior length-scale information to be used in future computations; we then collapse the β-layer and adjust the scores of the remaining α particles to yield the empirical distribution in column (viii). The final re-sampling step in column (ix) draws n_α particles from this empirical distribution to yield α particles in (x). These re-sampled α particles represent posterior-state information with uniform weights, forming the prior for the $t+1$ iteration (a return to column (i)). We note that the particle number is the same at the beginning and the end of each iteration.

The additional randomness introduced by the particle approximations and re-sampling procedures means that we have additional filtrations generated by the algorithm:

$$\mathcal{G}_t = \sigma(x_s^{(i)}, \bar{x}_s^{(i)}, s \leq t, i = 1, \ldots, n) \tag{35}$$

$$\bar{\mathcal{G}}_t = \sigma(x_s^{(i)}, \bar{x}_s^{(i)}, \bar{x}_t^{(i)}, s < t, i = 1, \ldots, n) \tag{36}$$

Here, $\bar{\mathcal{G}}_t \subset \mathcal{G}_t$, where $\bar{\mathcal{G}}_t$ includes particles $\bar{x}_t^{(i)}$ at start of iteration t but excludes the posterior particles at the end of t. These additional filtrations will be used in Section III.

We have now established how NMQA's α and β particles are updated under an iterative maximum likelihood approximation. We have also used the likelihood functions from Section I C to show how these particles are weighted globally within the NMQA particle filter, and we have discussed how high-weight particles survive through the NMQA branching mechanism at each iteration t. In the next subsection, we discuss how the β-particle layer is generated in each iteration t, before any of the particle weighting or branching procedures come into effect. The generation of a β-particle layer has no analogue in classical particle filtering literature, and we discuss two different theoretical approaches and their implications as follows.

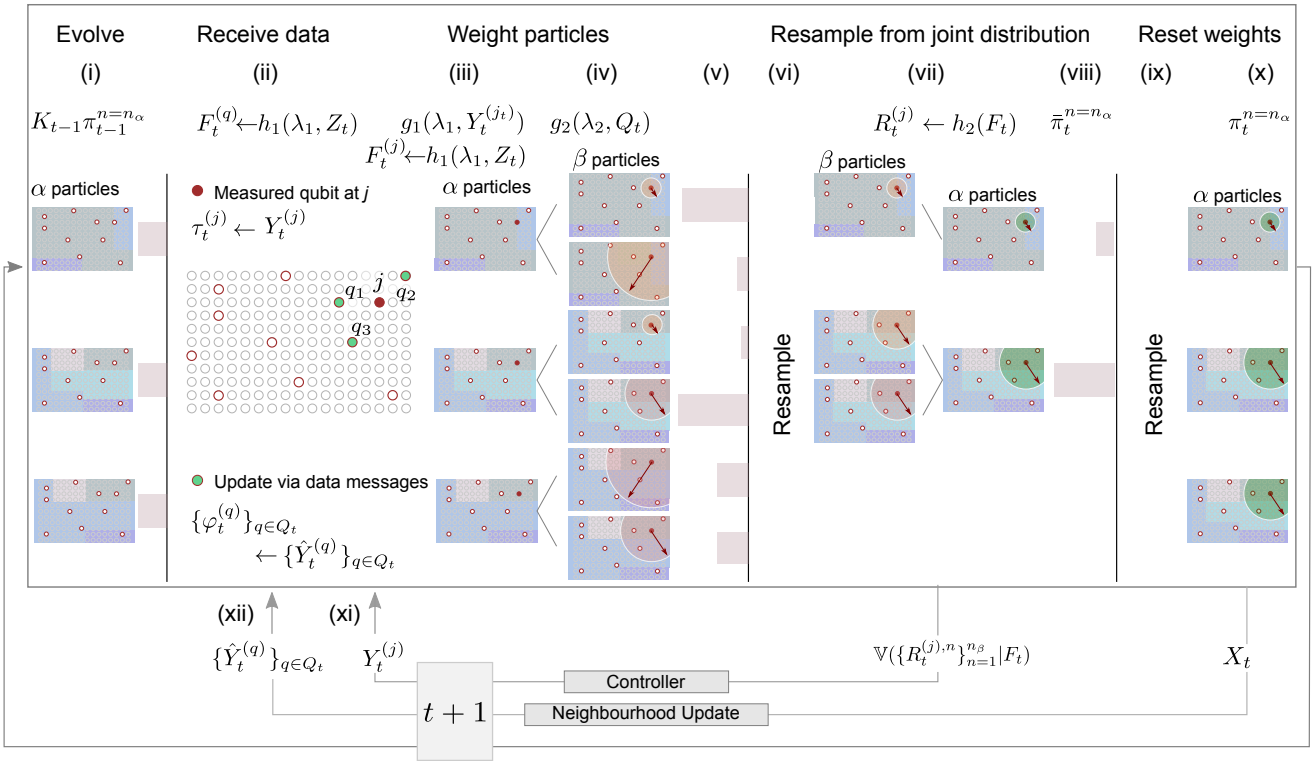

FIG. 2. NMQA's particle filter under an iterative likelihood approximation at t. α-particles carry map information, F_t, and β-particles represent length-scale information, $R_t^{(j)}$, at a measured location j; $n_\alpha = 3$, $n_\beta = 2$. (i) The posterior from $t-1$ is dynamically evolved to t via K_{t-1} yielding p_t^n. (ii) A physical measurement at t. Shared data messages from $t-1$ are updated via $h_1(\lambda_1, \hat{Y}_t^{(q)})$ and $\varphi_t^{(q)}$ for $q \in Q_{t-1}$ the posterior neighbourhood at $t-1$. (iii)-(iv) Likelihood functions score individual particles. (iii) Each α-particle is scored using $g_1(\lambda_1, Y_t^{(j)})$. Assuming X_{t-1} is known, $F_t^{(j)}$ is updated for each α-particle using $h_1(\lambda_1, Y_t^{(j)})$ and $\tau_t^{(j)}$. (iv) Assuming F_t for each α-particle is known, a β-particle is scored via $g_2(\lambda_2, Q_t)$. (v) The global likelihood $g_t = g_1(\lambda_1, Y_t^{(j)})g_2(\lambda_2, Q_t)$ is computed for $n_\alpha n_\beta$ particles. (vi) Particles are re-sampled, resulting in new off-spring particles depicted as leaves in (vii). (vii) The empirical mean and variance estimates of $R_t^{(j)}$ are stored for the β-layer of each α particle; the posterior estimate is passed onto the controller in (xi). (viii) The β-layer is collapsed for each α-particle giving rise to the weighted empirical distribution, $\bar{\pi}_t^n$. (ix) A second re-sampling step occurs with n_α particles using the empirical distribution $\bar{\pi}_t^n$ in (viii). (x) Output of (ix) results in the final posterior distribution π_t^n at t; the mean of this distribution reflects the best information about X_t conditioned on data, and this is shared with neighbouring qubits in the posterior neighborhood in (xii). (xi) Controller picks the location for the next measurement where estimated variance is highest. (xii) Data messages are generated for the neighbours of j in the posterior neighborhood using Eq. (14).

B. β-particle expansion strategy

A unique feature of NMQA is the role of the β-layer in estimating length-scale information about the underlying field (spatial correlations), captured via R_t. The inference facilitated by β particles governs both the size of the neighborhoods and the extent of the information sharing that occurs between neighboring qubits. A critical step of this process is the generation of β-particles for each α-parent at any t; the choice of such a strategy completes the definition of the NMQA algorithm.

In NMQA, the existence of diversity in β particles helps the algorithm to explore a very large state-space for R_t, given a hypothesis for F_t. If particle diversity is maintained at high t, the algorithm can respond to changes in the estimated map as measurements accumulate. On the other hand, linking β-particle generation at t to previously learned information at $t-1$ simplifies implementation and facilitates particle convergence at the risk of restricting NMQA from exploring the full state-space.

For standard particle filters with a single type of particle, these issues are addressed using techniques such as injecting new particles from uniform distributions [18], and sampling particles from probability distributions that change during filtering [17, 19, 20]. For NMQA, we will address these issues by focusing on how β-particles are revived in each iteration t. We compare two strategies for β-particle expansion: what we term 'Uniform' (introduced in [1]) and 'Trunc. Gaussian'. Each of these expansion strategies addresses the manner by which sam-

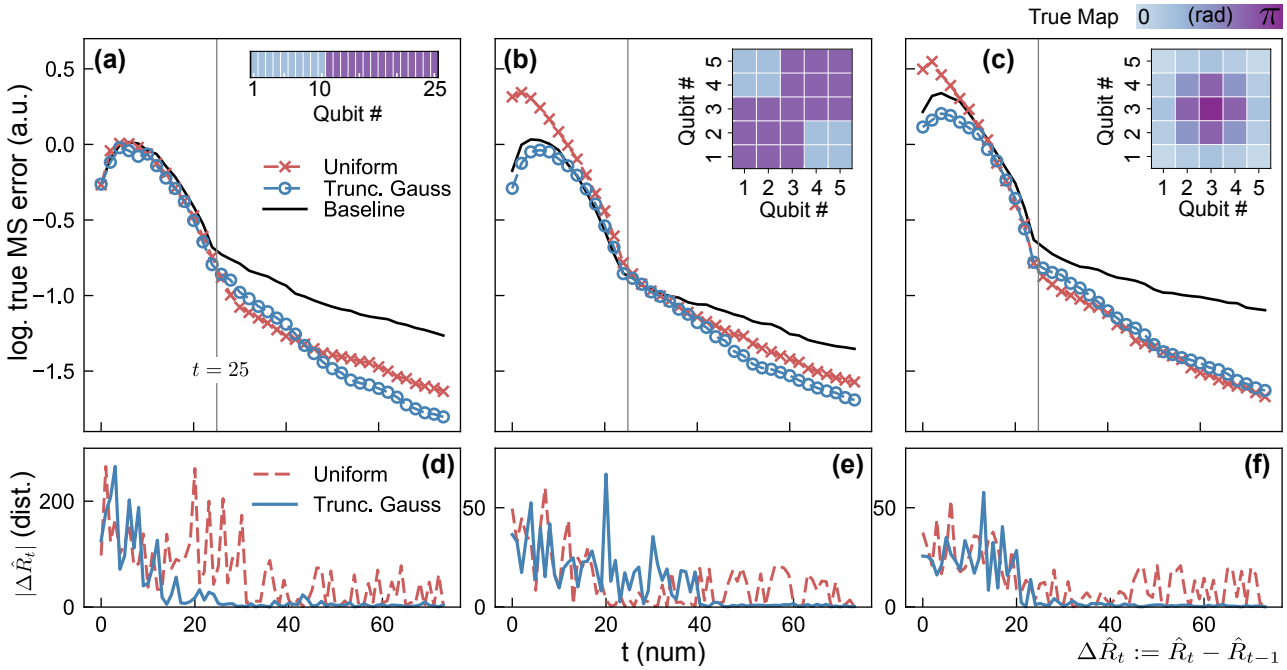

FIG. 3. NMQA performance by choice of 'Uniform' v. 'Trunc. Gaussian' expansion strategy to generate β-particle layer. Vertical columns represent 1D linear array, a 2D array with a square field, and a 2D array Gaussian field with $d = 25$, depicted as heat-maps of qubit phases (insets of (a)-(c)), with low and high field values at 0.25π and 0.75π radians respectively. (a)-(c) main panels are semi-log plots where the y-axis is the log of the expected mean square map reconstruction error per qubit, over 50 runs; x-axis represents the number of iterations, t. Each iteration corresponds to one physical measurement. Uniform (Trunc. Gaussian) expansion strategy shown as red crosses (blue circles). Baseline data (solid black) represents equivalent simulation in [1]. Vertical grey lines mark t for which number of physical measurements equals system size, $t = d$. (d)-(i) depict absolute rate of change of estimated length-scales vector $|\Delta\hat{R}|$ v. t under Uniform (red dashed) and Trunc. Gaussian (blue solid) for a single run. Tuned parameters $(\Sigma_v, \Sigma_F, \lambda_1, \lambda_2)$: (a),(d) Uniform $(6.0e^{-9}, 0.10, 0.88, 0.72)$, $n_\alpha = 15$; Trunc. Gaussian $(9.0e^{-8}, 2.6e^{-5}, 0.88, 0.72)$, $n_\alpha = 21$. (b),(e) Uniform $(7.1e^{-7}, 0.04, 0.88, 0.72)$, $n_\alpha = 3$; Trunc. Gaussian $(8.9e^{-7}, 1.9e^{-9}, 0.88, 0.72)$, $n_\alpha = 21$. (c),(f) Uniform $(5.9e^{-9}, 0.10, 0.72, 0.95)$,$n_\alpha = 3$; Trunc. Gaussian $(0.77, 4.6e^{-6}, 0.72, 0.95)$, $n_\alpha = 30$. For all data, $n_\beta = \frac{2}{3}n_\alpha$.

ples for $R_t^{(j)}$ are obtained when β-particles are generated for every α parent at the location of the measured qubit, j (columns (iii)-(iv) of Fig. 2).

The 'Uniform' expansion strategy [1] involves β-particles sampling $R_t^{(j)}$ from a uniformly distributed *prior* for all t. It has the effect of discarding previously learned information and repeating the learning procedure for $R_t^{(j)}$ at each t, under the assumption that the estimated map is uncorrelated from some previous iteration when qubit at location j was measured. From a practitioners' standpoint, a Uniform strategy for expanding the β-layer represents a conservative approach for any particular physical application.

An alternative expansion strategy, denoted 'Trunc. Gaussian', uses the estimated distribution of length-scales from $t-1$ to generate samples at t. A candidate $R_t^{(j)}$ is generated from a truncated Gaussian described within the limits $[R_{min}, R_{max}]$, with a mean and variance estimated empirically from posterior particles distribution at $t-1$. This strategy assumes that the length-scale distribution of $R_t^{(j)}$ for each j and t can be accurately described by the first two moments of the distribution.

The impact of these two strategies on mapping performance is compared numerically in Fig. 3. In Fig. 3(a)-(c), we consider three different physical configurations, with the true field depicted as a colorscale over $d = 25$ qubits. For these panels, the y-axis depicts the log of the expected value of the mean-square-error per qubit. This error is defined as the difference between the map reconstruction from NMQA and the true map. The x-axis depicts the number of iterations, t, with a single physical measurement performed at each t. The Baseline result (solid black) is obtained by conducting NMQA simulation with the same parameters as used in [1].

For data presented in colored markers in Fig. 3, the NMQA algorithm was numerically tuned over a much larger configuration space than investigated originally in [1]. For each true field and particle expansion strategy, we uniformly and randomly sample values between zero and one, for thirty random pairs across 11 uniformly chosen orders of magnitude $(\Sigma_F, \Sigma_v) \in [10^{-10}, 10]^2$ and 29 random pairs of $(\lambda_1, \lambda_2) \in [0, 1]$. The configuration $\lambda_1 = \lambda_2 = 0$ is always included within the procedure. For each

$(\Sigma_F, \Sigma_v, \lambda_1, \lambda_2)$, we run NMQA for five different particle configurations, setting $n_\alpha = 3, 9, 15, 21, 30, n_\beta = \frac{2}{3} n_\alpha$.

In each run, $t_{max} = 75$ measurements are provided sequentially to NMQA, with a single measurement performed in each t. The configuration $(\Sigma_F, \Sigma_v, \lambda_1, \lambda_2, n_\alpha)$ with the lowest expected error per qubit at $t = t_{max}$ is set as the numerically 'tuned' configuration for NMQA for a given truth and particle expansion strategy. These tuned parameter configurations for NMQA were then used alongside a Uniform or a Trunc. Gaussian particle expansion strategy and true field. In a single run, the estimated map after each iteration t is extracted and stored, and the run terminates at $t = t_{max} = 75$. The log of the expected value of the mean-square error per qubit is plotted against t for each expansion strategy, with the expectation computed over 50 independent runs. Data are compared against a 'Baseline' representing the equivalent simulation performed in [1].

In these demonstrations the tuned Trunc. Gaussian strategy for β-particle expansion shows the benefits of maintaining prior learnings in the small-t regime. However, the differences diminish between the Uniform and Trunc. Gaussian strategies across the different configurations tested in the large-t regime. The difference between these two regimes is empirically demarcated by the vertical grey lines in (a)-(c), which marks the position $t = d$, where the NMQA algorithm has the potential to measure each qubit on the grid at least once. A distinctive kink in the trajectory of true error consistently appears at this point, confirmed in additional simulations using larger values of d. In bottom panels (d)-(f), we plot the absolute rate of change in estimated R_t in a single run, for each of the three configurations. Under Trunc. Gaussian (blue solid), fluctuations in estimated R_t converge to zero; whereas these fluctuations persist with t under Uniform even at high t (red dashed). We observe that convergence of the estimated R_t with respect to t in a single run occurs only under the choice of Trunc. Gaussian strategy in (d)-(f).

While the differences in performance of NMQA accruing from the particle expansion strategy are small, we note that the Trunc. Gaussian strategy reduces in performance relative to Uniform as the spatial variations in the field become short compared to the inter-qubit spacing. In (a) and (b), groups of qubits exist over which a neighbourhood can be defined such that the true field has the same value across the neighbourhood i.e. the true field varies 'slowly' in space relative to inter-qubit spacing. In contrast, the true field has a different value for every qubit location in (c) *i.e.* the true field varies 'rapidly' relative to inter-qubit spacing. Hence, these numerical investigations provide evidence that the Trunc. Gaussian assumption works as expected as measurement information increases, with information sharing giving a greater benefit in map estimation when field correlations are longer than inter-qubit separation. Finally, we confirm tuned $\lambda_1, \lambda_2 \neq 0$ for all cases presented, suggesting non-trivial information sharing occurs within NMQA for all cases presented, irrespective of choice of β-particle expansion strategy.

III. CONVERGENCE ANALYSIS

Fixing the choice of Uniform or Trunc Gaussian as a β-particle expansion strategy for any practical application thus completes technical specification of the NMQA algorithm. With this structure in hand, this section will address the extent to which NMQA satisfies the axioms of asymptotic convergence theorems in the stochastic nonlinear particle filtering literature. Performing such an analysis represents a major advantage relative to many extant classical filtering approaches; the particle filtering literature consists of numerous algorithmic constructions where the choice of particle sampling techniques [21, 22] leaves convergence properties unresolved, despite wide use of these algorithms in the community [12, 14, 23, 24]. We present theoretical results for asymptotic convergence of NMQA in specific parameter regimes; outside of these regimes, we provide numerical evidence for convergence behaviour of NMQA. Following [12], we introduce convergence conditions in the literature for these particle filters.

A. General convergence theory for particle filters

Our notion of 'convergence' will focus on the established notions of 'convergence in expectation' and 'almost sure' (a. s.) convergence [12], which both imply convergence in probability and distribution [25]. For a sequence of random measures, $(\mu^n)_{n=1}^{\infty}$, and another random measure μ, these are formalized as:

$$\lim_{n\to\infty} \mathbb{E}[|\mu^n f - \mu f|] = 0, \quad \forall f \in C_b(\mathbb{S}) \tag{37}$$

$$\lim_{n\to\infty} \mu^n = \mu, \quad \mathbb{P}-\text{a.s.} \tag{38}$$

The first line defines convergence in the expectation values for all continuous, bounded functions f. The second line defines a. s. convergence.

We will begin by stating a general theorem for particle approximations:

Theorem 1 (Bain & Crisan, 2009)**.** *For all $f \in B(\mathbb{S}_X)$ and all $t \in [0, T]$ the limits:*
(a0) $\lim_{n\to\infty} \mathbb{E}[|\pi_t^{n,z_{0:t}} f - \pi_t f|] = 0$
(b0) $\lim_{n\to\infty} \mathbb{E}[|p_t^{n,z_{0:t-1}} f - p_t f|] = 0$
hold if and only if for all $f \in B(\mathbb{S}_X)$ and all $t \in [0, T]$:
(a1) $\lim_{n\to\infty} \mathbb{E}[|\pi_0^{n,z_{0:t}} f - \pi_0 f|] = 0$
(b1) $\lim_{n\to\infty} \mathbb{E}[|p_t^{n,z_{0:t-1}} f - K_{t-1}\pi_{t-1} f|] = \lim_{n\to\infty} \mathbb{E}[|\pi_t^{n,z_{0:t}} f - \bar{\pi}_t^{n,z_{0:t}} f|] = 0$

The statements above capture the intuition that if the particle filter starts out with the correct initial distribution and stays reasonably close to the true posterior at each t, then the particle-filter approximation will improve as a the number of total particles increases (using

(a0) and *(b0)*). 'Staying close' at each t involves two concepts that are captured by the first and second equations in *(b1)* respectively. Firstly, the transition kernel brings the posterior measure from the previous into the current iteration approximately correctly. Secondly, the likelihood function and the branching mechanism ensures that the weights in $\bar{\pi}_t^{n,z_{0:t}}$ are a good approximation to $\pi_t^{n,z_{0:t}}$.

The theorem above refers to a single fixed path, $Z_{0:t} = z_{0:t}$, constituting a single run of measurements. The same results hold for a random observation vector, $Z_{0:t}$ under the additional condition that there exists some constant k_t such that $p_t g_t \geq k_t$. With these substitutions, the conditions (a1) and (b1) imply convergence in expectation of the sequences $\pi_t^{n,Z_{0:t}}$ ($p_t^{n,Z_{0:t-1}}$) to π_t (p_t), and the proof is provided in the Appendix.

In order to use Theorem 1 to make necessary and sufficient statements about convergence (of any form), we restrict our discussion to the mapping of continuous bounded functions, f, and place additional requirements on the transition kernel and likelihood functions. One of these additional requirements is that the transition kernel K_t is Feller [12, 14]. This means that $K_t f \in C_b(\mathbb{S}_X)$ for all $f \in B(\mathbb{S}_X)$, *i. e.* the transition kernel does not bring f outside the space of continuous bounded functions. The following theorem states these additional requirements and establishes the link with the branching mechanism for the particle filter using results from [12]:

Theorem 2. *Let $(\pi_t^{n,Z_{0:t}})_{n=1}^{\infty}$ and $(p_t^{n,Z_{0:t-1}})_{n=1}^{\infty}$ be measure valued sequences produced by particle approximations and branching mechanisms satisfying Proposition 3 (below). Assume there exists some constant k_t such that $p_t g_t \geq k_t$. Then:*
(i) Theorem 1 holds for all $f \in B(\mathbb{S}_X)$ and all $t \in [0,T]$ with the substitutions $\pi_t^{n,z_{0:t}} \to \pi_t^{n,Z_{0:t}}$ and $p_t^{n,z_{0:t-1}} \to p_t^{n,Z_{0:t-1}}$.
(ii) Assume that the transition kernel for X is Feller and that the likelihood functions are all continuous for all $t \in [0,T]$. Then, the sequences $\pi_t^{n,Z_{0:t}}$ ($p_t^{n,Z_{0:t-1}}$) converges in expectation to π_t (p_t) for all $t \in [0,T]$ if and only if conditions (a1) and (b1) are satisfied for all $f \in C_b(\mathbb{S}_X)$ and all $t \in [0,T]$.
(iii) Assume that the transition kernel for X is Feller and that the likelihood functions are all continuous for all $t \in [0,T]$. If the branching mechanism satisfying Proposition 3 is a multinomial random process, then, $\lim_{n\to\infty} \pi_t^{n,Z_{0:t}} = \pi_t$, $\mathbb{P}-$a.s. and $\lim_{n\to\infty} p_t^{n,Z_{0:t-1}} = p_t$, $\mathbb{P}-$a.s..

In the above, part (i) implies convergence in expectation, but parts (ii) and (iii) are 'if and only if' statements for our two notions of convergence; both place additional requirements on the form of the likelihood and transition kernels. The theorem summarises several key results in [12] and full details are re-stated for completeness in the Appendix.

In this context, the branching mechanism of any particle filter can be designed arbitrarily but has important implications in convergence analysis. Each parent particle is replaced by a number of offspring, given by the random number $\xi^{(i)}$. Following [12], a branching mechanism satisfying the following properties can be used to show convergence as follows.

Proposition 3. *Branching mechanisms for a particle filter satisfy:*

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

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 is re-stating that empirical weight of the particle is associated with the true probability of observing that particle (state information) given some observed history via the filtration $\bar{\mathcal{G}}_t$. The last property places a constraint on the covariance matrix associated with the branching process. This constraint appears to have no apriori justification, but it is a condition associated with a particle filter's convergence properties. The link with convergence is discussed at length in [12], and clearly seen in the proof of Theorem 7 in the Appendix.

Our next task is to identify the extent to which the structures of NMQA satisfy these established convergence criteria.

B. Convergence properties of NMQA

In analyzing NMQA's convergence properties, we make global simplifications where the environment does not change with time. To enable theoretical analysis in this sub-section, we suppress the controller by instead envisioning uniformly random sampling of all qubit locations to align with scope of Theorem 2. (In subsequent numerical analysis of convergence, the NMQA controller is reinstated with full functionality.) With these simplifications, the convergence implications of NMQA's algorithm design components are discussed below:

1. *Transition kernel and likelihood functions*

The properties of the likelihood functions and transition kernels of NMQA satisfy convergence conditions under Theorem 2.

There, the constraint on the transition kernel merely states that dynamical evolution cannot bring the true state outside the space over which functions of random measures are defined - namely, continuous, bounded Borel measurable functions. For time-invariant noise maps, we assume an identity transition kernel such that the Feller property is automatically satisfied.

The constraint on the likelihood function is that is must be continuous. This continuity condition is imposed so that any change in the conditional predictive distribution of X_t will not result in a large variation in posterior distribution [14]. Hence, the requirement on the continuity of likelihoods is interpreted as two specific statements. First, any likelihood function, $g(Y_t, X_t) > 0$ is strictly positive such that $p_t g_t$ is never zero in Eq. (6). Second, the function, $g(Y_t, \cdot)$ is bounded and continuous [14]. For single-qubit projective measurements in $g_1(\lambda_1, Y_t^{(j)})$, the function is continuous in state variables once λ_1 is fixed for an application, and $Y_t^{(j)}$ is fixed by an incoming measurement. Boundedness is satisfied for $b = 1/2$.

2. Empirical distributions

Under a 'Trunc. Gaussian' expansion strategy, we expect Theorem 2 to hold from t to $t+1$, and we invoke Theorem 2 with $c_t = 1$.

Two β-particle expansion strategies were proposed in earlier sections, Uniform and Trunc. Gaussian, of which only the latter approach enables learned-state information to transfer from t to $t+1$, satisfying Theorem 2. In general, the transfer of learned information from t to $t+1$ is mediated by the sampling procedure for β-particles. The choice of this sampling distribution cannot cause large perturbations from the trajectory NMQA would have otherwise followed under consecutive applications of its transition kernel and its likelihood function. A 'Trunc. Gaussian' expansion strategy samples β-particles at $t+1$ based on estimated state information at t, thereby ensuring that any additional randomness inserted into the filter is on order of the filters' own estimated state variance. Numerical investigations support the approximation of length-scale distributions using only the first two moments of a truncated Gaussian distribution, particularly where an unknown field varies slowly in space with respect to the interqubit distance in a regular array.

Aside from the particle expansion strategy, the particle branching mechanism affects the distance between the empirical particle filter distribution before and after a re-sampling step. In our case, the particle branching mechanism (columns (iv)-(x)) for NMQA corresponds to a multi-nomial branching process, with proof appearing in the Appendix. For multi-nomial branching, we can invoke Theorem 2 with $c_t = 1$; this implies that the squared distance between empirical distributions in the particle filter before and after the branching step, $\mathbb{E}[(\pi_t^n - \bar{\pi}_t^n)^2] \leq \frac{c_t \|f\|_\infty^2}{n}$, decays as $\frac{1}{n}$ as $n \to \infty$ 'almost surely'. The proof of the second statement follows [12] and is restated in full in the Appendix.

Next, we assess the impact of NMQA's unique information-sharing mechanism on asymptotic convergence by considering specific choices for the parameters λ_1, λ_2.

3. Convergence for λ_1, λ_2 regimes:

The simplest regime for λ_1, λ_2 for which NMQA converges by Theorem 2 removes information sharing asymptotically through setting $\lambda_1, \lambda_2 = 0$. Hence Theorem 2 applies straightforwardly, and NMQA converges in expectation and almost surely to the true posterior with $c_t = 1$. In the limit that every qubit at every location, j, is measured a very large number of times, $\tau_t^{(j)}$, the role of data messages at each location decays at a rate set by λ_1, λ_2 and $\tau_t^{(j)}$. For sufficiently large $\tau_t^{(j)}$ at every j, the general case for $\lambda_1, \lambda_2 \in [0, 1)$ eventually approaches the case $\lambda_1, \lambda_2 = 0$.

In most practical circumstances, however, tuning procedures show that values of $\lambda_1, \lambda_2 \neq 0$ result in low errors, and evidence for non-trivial sharing exists [1] at low-to-moderate values of $\tau_t^{(j)}$. In regimes of practical interest, some locations are never measured, while other locations receive some physical measurements *i.e.* $\tau_t^{(j)}$ relative to $\lambda_1, \lambda_2 \neq 0$ for some j, t ensures data messages are contributing non-trivially to the inference procedure. In these 'intermediate' parameter regimes, a simplified theoretical analysis is not possible and we depart from Theorem 2 in that we can no longer assume that underlying noise processes relating to measurement are uncorrelated with the state estimation procedure. This is because the data messages, $\hat{Y}_t$, represent random variables that 'correlate' state estimation at one location, j, with the observation vector corresponding to some other location k.

It is important to note that these correlations arise between *different* locations, and NMQA does not correlate state information and measurement error at the same location. A data-message at qubit k could be received from a physically measured qubit at j by sharing a large estimated value using a short estimated length-scale, or by sharing a small estimated value over a long length-scale. The algorithm builds a map by iteratively combining the effect of numerous, non-negative contributions implied by both data-messages and physical measurements. If measurement error and state estimation were correlated at the same location j and iteration t, then state estimation and measurement noise filtering would become intractable. By correlating different locations, NMQA can build a map even when data is sparse; nonetheless the algorithm may face non-unique, equal-cost solutions to the inference problem in the low-t limit. As the number of measurements on all qubits increases in the high-t limit, previous data messages received at any given lo-

cation j have asymptotically less influence on the state estimates at j, enabling an asymptotic transition to a regime in which NMQA converges by Theorem 2.

Thus far, we have established that NMQA's transition kernel and likelihood functions satisfy properties required by Theorem 2, and that this theorem continues to apply under the Trunc. Gaussian expansion strategy we have introduced with multi-nomial particle-branching processes. For $\lambda_1, \lambda_2 \neq 0$, and in the high measurement regime at every qubit, we see that NMQA converges 'almost surely' and in expectation by Theorem 2, for $c_t = 1$.

For $\lambda_1, \lambda_2 \neq 0$ in intermediate regimes, NMQA is difficult to analyse theoretically due to non-trivial correlations generated by data-messages between the observation vector and the estimated state. In the next subsection we demonstrate that NMQA satisfies error scaling bounds implied by Theorem 2, despite operating outside of this reference's formal scope, through numeric analysis.

C. Numerical analysis of convergence

In order to analyze the regime of practical interest where the considerations of Theorem 2 are difficult to apply analytically, we perform numerical simulations and investigate the extent to which Theorem 2 continues to hold for NMQA. In this subsection, we implement NMQA in a natural, unrestricted fashion. Simplifying theoretical assumptions of the previous sub-section are now removed: the NMQA controller is reinstated; λ_1, λ_2 are numerically tuned with nonzero values; and NMQA cannot receive a large number of measurements at every qubit location. This unrestricted NMQA implementation is not guaranteed to be within the scope of traditional convergence theory.

Our approach will be to use Theorem 2 to set theoretical expectations on the true-error scaling with particle number for NMQA, and to use numerical analysis to determine if an unrestricted NMQA implementation yields empirical error-scaling parameters that are compatible with these bounds. In particular the condition that $c_t = 1$ means that the quantity $\mathbb{E}[(\pi_t^n - \bar{\pi}_t^n)^2]$ decays as $\frac{1}{n}$ as $n \to \infty$ almost surely. This quantity is the expected value of distance between the two empirical probability measures as the number of particles n increases and will be the target of our investigations.

In our numerical analyses, we compute the expected value of the true mean-square-error per qubit in map reconstruction, $\mathbb{E}[(\mathbb{E}[\pi_t^n] - \mathbb{E}[\pi_\infty])^2/d]$ as a proxy metric to investigate how errors scale with n. A plot of the log true mean-square-error per qubit against log number of particles yields the estimated slope, Δ, that estimates how error scales with n for finite, fixed values of t. Broadly, a value of $\Delta < 0$ indicates that expanding the particle number improves the inference procedure (error decreases with greater n), while a value $\Delta > 0$ indicates increased error with n. We expect $\Delta \in [-1, 0)$ for an unrestricted NMQA implementation that accords with Theorem 2, where the equality holds if the branching mechanism is the slowest contribution to convergence.

Fig. 4 shows a numerical convergence analysis for the three mapping case-studies addressed in Fig. 3, with maps duplicated in the right insets. The choice of β-particle expansion strategy impacts our expectations with respect to convergence and we explore both strategies introduced above. In the main panel, we plot the extracted Δ against t for both Uniform (red crosses) and Trunc. Gaussian (blue circles) expansion strategies. For each value of t, these Δ values are calculated from the gradient of a line of best fit for the log of true mean-square-error per qubit in map reconstruction against the log of n_α, the total number of α 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 (a)-(c).

For $t \lesssim d = 25$, we observe $\Delta < 0$ for both data sets. This means that increasing n_α under any expansion strategy (Uniform or Trunc. Gaussian) improves the inference procedure when data is sparse, assuming that the correct initial distribution has been specified. For the high-data regime, $t \gg d$, the values of Δ diverge between the two expansion strategies. The Uniform approach in (a)-(c) shows that $\Delta > 0$ as t increases. In contrast, under a Trunc. Gaussian strategy, we see that $\Delta \in [-1, 0)$ 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 $\Delta > 0$ since Theorem 2 does not hold asymptotically as information about length-scales is reset to the prior distribution even for large values of t. In this limit, increasing particle number 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 Theorem 2 holds. Further, as mentioned above our branching mechanism is multinomial, and we expect that the condition $\Delta \in [-1, 0)$, is satisfied asymptotically.

With these results, we establish numerical evidence that Theorem 2 holds for NMQA for intermediate regimes under a Trunc. Gaussian approach, even in the large-data limit. These investigations also suggest that true error in NMQA scales in a predictable way for a range of different physical configurations despite NMQA's unique algorithmic structure.

IV. CONCLUSION

In this work we provided a detailed quantitative analysis of the underlying algorithmic structure of NMQA using the specific application of noise mapping in mesoscale

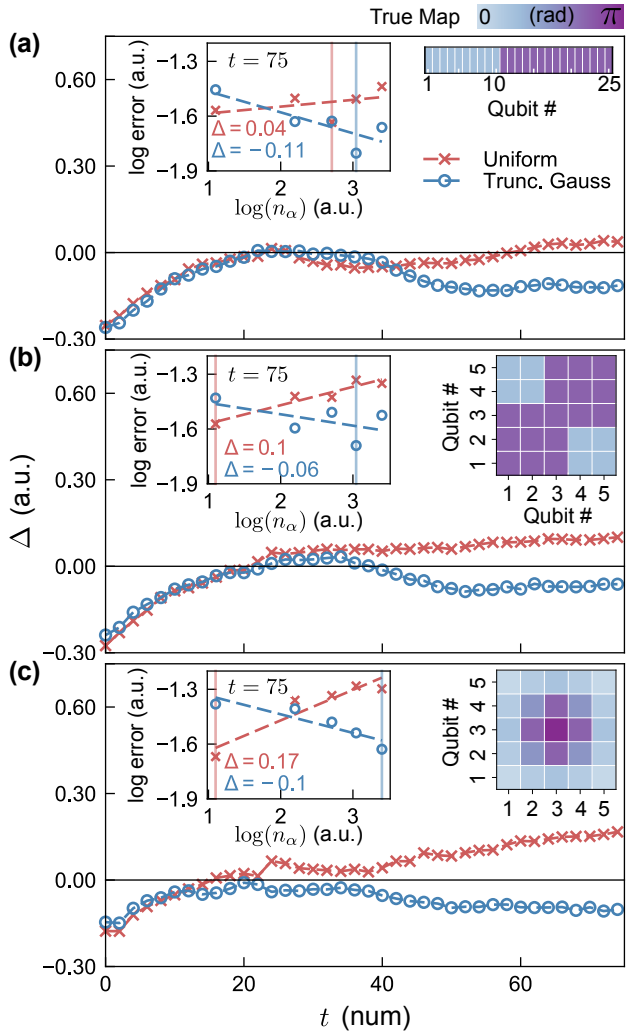

FIG. 4. Error scaling behaviour for Uniform and Trunc. Gaussian. Rows represent 1D linear array, a 2D array with a square field, and a 2D array Gaussian field with $d = 25$ (right insets); with high and low qubit phase values of $0.25\pi, 0.75\pi$ radians depicted on heat-maps. (a)-(c) main panels depict Δ against t for tuned NMQA parameters. $\Delta > 0$ for Uniform; $\Delta \in [-1, 0)$ for Trunc. Gaussian for $t \gg d$ agrees with Theorem 2. Data for Uniform (red crosses) and Trunc. Gaussian (blue circles). Left insets depict the log of the expected mean square map reconstruction error per qubit over 50 runs against the log of n_α number of α-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 $t = 75$. Δ is the gradient of the line of best fit (dashed lines). Vertical colored lines mark tuned particle configuration plotted in Fig. 3. 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) $(0.77, 4.6e^{-6}, 0.72, 0.95)$.

quantum computing architectures. With substantial departures from typical non-linear filtering literature, we presented technical details for the NMQA framework and its unique implementation through an iterative, particle-based filtering algorithm. Our methodology possesses several novel features. First, we designed classical likelihood functions that enable us to work with single qubit projective measurements. Second, we implemented an iterative maximum likelihood approach to approximately solve a non-linear filtering problem via a novel two-layered particle filter in a large state space consisting of both maps and their approximate values over local neighbourhoods. We proposed unique particle generation and branching mechanisms to enable NMQA to explore a large state-space using both particle layers. Third, we implemented a tunable information sharing mechanism where estimated state information is shared between qubits in small spatial neighbourhoods. The ability to tune the strength of information sharing via parameters λ_1, λ_2 allows a practitioner to test, using measurement data, whether shared information in NMQA improves inference over a brute force approach for any physical application.

Under simple parameter regimes for a specific choice of particle expansion strategies, we further established that NMQA has a multi-nomial branching mechanism and converges in expectation and 'almost surely' in the limit of large data and total particle numbers. The rate of convergence computed as the scaling factor of true error with particle number, Δ, was shown to be theoretically expected to satisfy the condition $\Delta \in [-1, 0)$. Outside of these simple parameter regimes, where conventional convergence theorems are not easy to apply, we used simulations to provide numerical evidence for asymptotic convergence: namely, that the condition $\Delta \in [-1, 0)$ appears to hold for a range of algorithmic and physical configurations under a Trunc. Gaussian particle expansion strategy. These numeric results suggested that it may be possible to extend conventional convergence theorems to intermediate regimes for NMQA. Establishing analytic convergence for intermediate regimes remains an open question for this algorithm.

Our analyses indicated that NMQA represents an approximation or a noise filtering problem in low data regimes (low-t), and gradually evolves to a mathematical interpolation problem in 2D when one obtains a set of qubit locations at which the noise field has been accurately estimated (high-t regime). This leads to new open questions about the optimal spatial arrangement of qubits in 2D for the reconstruction of spatial fields on proximal data qubits. For interpolation problems on a 2D square, it is known that Padua points represent an ideal arrangement of point values for performing Lagrange interpolation [26, 27]. Padua-inspired techniques for approximation and filtering are an active area of research (*e.g.* a recent classical example [28] and application to reconstructing probability distributions for quantum systems [29]).

Exploration of such optimal interpolation challenges has both practical relevance to the design of multi-qubit

devices, and also helps to extend NMQA's convergence analyses to formally accommodate spatial-sampling requirements. In addition, time-varying environments and drift in experimental parameters are commonly experienced in laboratory operation of multi-qubit devices; we believe it is possible to incorporate such environmental dynamics into this framework through an appropriately designed transition kernel. Finally we emphasize that the specific problem of noise mapping via spectator qubits represents only a single application of what appears to be a broadly deployable algorithmic framework. We look forward to exploring how the rigorous analysis we have performed here may be applied to problems in device calibration, cross-talk analysis, and automated system tuneup.

ACKNOWLEDGEMENTS

R. Gupta would like to thank Andrew Doherty for extensive discussions. This work was partially supported by the US Army Research Office under Contract W911NF-12-R-0012, and a private grant from H. and A. Harley.

[1] Riddhi Swaroop Gupta, Alistair R Milne, Claire L Edmunds, Cornelius Hempel, and Michael J Biercuk, "Adaptive scheduling of noise characterization in quantum computers," arXiv preprint arXiv:1904.07225 (2019).

[2] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C. Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando G. S. L. Brandao, David A. Buell, Brian Burkett, Yu Chen, Zijun Chen, Ben Chiaro, Roberto Collins, William Courtney, Andrew Dunsworth, Edward Farhi, Brooks Foxen, Austin Fowler, Craig Gidney, Marissa Giustina, Rob Graff, Keith Guerin, Steve Habegger, Matthew P. Harrigan, Michael J. Hartmann, Alan Ho, Markus Hoffmann, Trent Huang, Travis S. Humble, Sergei V. Isakov, Evan Jeffrey, Zhang Jiang, Dvir Kafri, Kostyantyn Kechedzhi, Julian Kelly, Paul V. Klimov, Sergey Knysh, Alexander Korotkov, Fedor Kostritsa, David Landhuis, Mike Lindmark, Erik Lucero, Dmitry Lyakh, Salvatore Mandrà, Jarrod R. McClean, Matthew McEwen, Anthony Megrant, Xiao Mi, Kristel Michielsen, Masoud Mohseni, Josh Mutus, Ofer Naaman, Matthew Neeley, Charles Neill, Murphy Yuezhen Niu, Eric Ostby, Andre Petukhov, John C. Platt, Chris Quintana, Eleanor G. Rieffel, Pedram Roushan, Nicholas C. Rubin, Daniel Sank, Kevin J. Satzinger, Vadim Smelyanskiy, Kevin J. Sung, Matthew D. Trevithick, Amit Vainsencher, Benjamin Villalonga, Theodore White, Z. Jamie Yao, Ping Yeh, Adam Zalcman, Hartmut Neven, and John M. Martinis, "Quantum supremacy using a programmable superconducting processor," Nature **574**, 505–510 (2019).

[3] Piotr Szańkowski, Marek Trippenbach, Łukasz Cywiński, and Yehuda B Band, "The dynamics of two entangled qubits exposed to classical noise: role of spatial and temporal noise correlations," Quantum Information Processing **14**, 3367–3397 (2015).

[4] Sandeep Mavadia, Virginia Frey, Stephen Dona Jarrah Sastrawan, and Michael J Biercuk, "Prediction and real-time compensation of qubit decoherence via machine learning," Nature communications **8** (2017).

[5] Riddhi Swaroop Gupta and Michael J Biercuk, "Machine learning for predictive estimation of qubit dynamics subject to dephasing," Physical Review Applied **9**, 064042 (2018).

[6] Swarnadeep Majumder, Leonardo Andreta de Castro, and Kenneth R Brown, "Real-time calibration with spectator qubits," arXiv preprint arXiv:1907.03864 (2019).

[7] Félix Beaudoin, Leigh M Norris, and Lorenza Viola, "Ramsey interferometry in correlated quantum noise environments," Physical Review A **98**, 020102 (2018).

[8] Lukas Postler, Ángel Rivas, Philipp Schindler, Alexander Erhard, Roman Stricker, Daniel Nigg, Thomas Monz, Rainer Blatt, and Markus Müller, "Experimental quantification of spatial correlations in quantum dynamics," arXiv preprint arXiv:1806.08088 (2018).

[9] Sebastian Thrun, Wolfram Burgard, and Dieter Fox, "A probabilistic approach to concurrent mapping and localization for mobile robots," Autonomous Robots **5**, 253–271 (1998).

[10] Sebastian Thrun, "A probabilistic on-line mapping algorithm for teams of mobile robots," The International Journal of Robotics Research **20**, 335–363 (2001).

[11] Sebastian Thrun, Wolfram Burgard, and Dieter Fox, *Probabilistic robotics* (MIT press, 2005).

[12] Alan Bain and Dan Crisan, *Fundamentals of Stochastic Filtering*, Stochastic Modelling and Applied Probability (Springer, 2009).

[13] Dan Crisan, Pierre Del Moral, and Terry Lyons, *Discrete filtering using branching and interacting particle systems* (Université de Toulouse. Laboratoire de Statistique et Probabilités [LSP], 1998).

[14] Dan Crisan and Arnaud Doucet, "A survey of convergence results on particle filtering methods for practitioners," IEEE Transactions on signal processing **50**, 736–746 (2002).

[15] Christophe Andrieu, Arnaud Doucet, and Elena Punskaya, "Sequential monte carlo methods for optimal filtering," in *Sequential Monte Carlo Methods in Practice* (Springer, 2001) pp. 79–95.

[16] P. Szekeres and Cambridge University Press, *A Course in Modern Mathematical Physics: Groups, Hilbert Space and Differential Geometry* (Cambridge University Press, 2004).

[17] T. Li, M. Bolic, and P. M. Djuric, "Resampling methods for particle filtering: Classification, implementation, and strategies," IEEE Signal Processing Magazine **32**, 70–86 (2015).

[18] Drew Bagnell, "Statistical techniques in robotics," (2014).

[19] Simon Godsill and Tim Clapp, "Improvement strategies for monte carlo particle filters," in *Sequential Monte Carlo methods in practice* (Springer, 2001) pp. 139–158.

[20] Péter Karkus, David Hsu, and Wee Sun Lee, "Particle filter networks: End-to-end probabilistic localization from visual observations," arXiv preprint arXiv:1805.08975 **2** (2018).

[21] Arnaud Doucet, Nando De Freitas, and Neil Gordon, "An introduction to sequential monte carlo methods," in *Sequential Monte Carlo methods in practice* (Springer, 2001) pp. 3–14.

[22] Paul Fearnhead and Hans R Künsch, "Particle filters and data assimilation," Annual Review of Statistics and Its Application **5**, 421–449 (2018).

[23] Juho Kokkala and Simo Särkkä, "On the (non-) convergence of particle filters with gaussian importance distributions," IFAC-PapersOnLine **48**, 793–798 (2015).

[24] Hans R Künsch *et al.*, "Recursive monte carlo filters: algorithms and theoretical analysis," The Annals of Statistics **33**, 1983–2021 (2005).

[25] Michael J Evans and Jeffrey S Rosenthal, *Probability and statistics: The science of uncertainty* (Macmillan, 2004).

[26] Len Bos, Stefano De Marchi, Marco Vianello, and Yuan Xu, "Bivariate lagrange interpolation at the padua points: the ideal theory approach," Numerische Mathematik **108**, 43–57 (2007).

[27] Len Bos, Marco Caliari, Stefano De Marchi, Marco Vianello, and Yuan Xu, "Bivariate lagrange interpolation at the padua points: the generating curve approach," Journal of Approximation Theory **143**, 15–25 (2006).

[28] Stefano De Marchi, Wolfgang Erb, and Francesco Marchetti, "Spectral filtering for the reduction of the gibbs phenomenon for polynomial approximation methods on lissajous curves with applications in mpi," Dolomites Research Notes on Approximation **10** (2017).

[29] Olivier Landon-Cardinal, Luke CG Govia, and Aashish A Clerk, "Quantitative tomography for continuous variable quantum systems," Physical review letters **120**, 090501 (2018).

[30] Bernard Widrow, Istvan Kollar, and Ming-Chang Liu, "Statistical theory of quantization," IEEE Transactions on instrumentation and measurement **45**, 353–361 (1996).

[31] Rickard Karlsson and Fredrik Gustafsson, *Filtering and estimation for quantized sensor information*, Tech. Rep. LiTH-ISY-R-2674 (Linköping University, 2005).

[32] Ratthachat Chatpatanasiri, "How to sample from a truncated distribution if you must," Artificial Intelligence Review **31**, 1 (2009).

[33] John Burkardt, "The truncated normal distribution," Department of Scientific Computing Website, Florida State University (2014).

[34] Stanley P Lipshitz, Robert A Wannamaker, and John Vanderkooy, "Quantization and dither: A theoretical survey," Journal of the audio engineering society **40**, 355–375 (1992).

[35] Fredrik Gustafsson and Rickard Karlsson, "Generating dithering noise for maximum likelihood estimation from quantized data," Automatica **49**, 554–560 (2013).

Appendix A: Derivation of $g_1(\lambda_1, Y_t)$

In this section, we provide a full derivation of the likelihood function, $g_1(\lambda_1, Y_t)$ stated in the main text. The derivation proceeds by focusing on the case of a single qubit at an arbitrary location.

Let y_t be a continuous real variable that is linearly associated to the Born probability of observing a qubit in $Y_t = 0$ or $Y_t = 1$ state after a projective, single-qubit measurement at t. The action of a projective measurement is represented by the notation, $\mathcal{Q}(\cdot)$. Here, $\mathcal{Q}(\cdot)$ represents a single 'coin-toss' experiment (single Bernoulli trial) yielding a '0' or a '1', where the bias on the coin is given by the argument of $\mathcal{Q}(\cdot)$. For ease, we require y_t above to be zero mean and require that $y_t+1/2 \in [0,1], \forall t$ is interpreted as a Born probability. The measurement model now reads as:

$$Y_t \equiv \mathcal{Q}(y_t + \frac{1}{2}), \quad y_t \in [-\frac{1}{2}, \frac{1}{2}] \tag{A1}$$

The signal y_t is thus associated with the Born probability $y_t+1/2$ of seeing a qubit in a 0 or a 1 state at the moment a projective measurement is performed.

We assert that the effect of the projective measurement action $\mathcal{Q}(\cdot)$ is to discretize the amplitude of an otherwise continuous-amplitude, abstract signal $y_t + 1/2$ into two allowable values - $Y_t = 0$ or $Y_t = 1$. Suppose one has a sequence of real values $\{y_t + 1/2, t = 1, 2, \ldots\}$ which forms an abstract, unknown classical signal that is continuous in amplitude but discretised in time. The signal $y_t + 1/2$ is said to undergo a binary amplitude quantisation, where upon detection, $y_t+1/2$ can only assume one of the allowed amplitude levels to result in the output Y_t. The process of discretising or quantising the continuous amplitude of a classical real variable into a finite set of allowed ampltitude values is known as 'amplitude quantisation' and the resulting statistical properties are well described in classical literature [30, 31]. A trivial example of this scenario is if the true Born probability, $y_t+1/2$, associated with a qubit is constant in t over many identical projective measurements. A non-trivial example is if the true Born probability drifts slowly in t over many identical projective measurements; where drifts in t are 'slow' relative to the speed of the projective measurement action. One mechanism for correlations in t to arise is due to coupling of a qubit to a correlated (noise) environment [5].

We now proceed to derive the likelihood function $g_1(\lambda_1, Y_t)$ and illustrate concepts from classical amplitude quantisation. Our first step will be to link the measurement procedure in Eq. (A1) to the state space of some (hidden) true state variable X. To establish this link, we assume that the Born probability at t for our qubit depends on some function of a true state variable at each t i.e. $f(X_t)$. We additionally wish to introduce uncertainty in our knowledge of y_t, by adding a zero-mean Gaussian white noise, v_t, yielding a linear model where:

$$y_t := f(X_t) + v_t \tag{A2}$$

In the equation above, both $f(X_t), y_t$ are zero mean random variables that yield valid Born probabilities if their values lie in the interval $[-\frac{1}{2}, \frac{1}{2}]$. This means that the uncertainty given by the term v_t will have restricted values and the form of these restrictions will be detailed below. (If no such restrictions on v_t existed, we would be able to write a Gaussian distribution function for $v_t \sim \mathcal{N}(0, \Sigma_v)$.) A specific example of $f(\cdot)$ is given in the main text in Eq. (10); here, the true state is the relative qubit phase accumulation, F_t in a fixed Ramsey measurement, and $f(F_t) := \frac{1}{2}\cos(F_t)$. In general, however, the choice of the physically relevant state X_t and measurement procedure $f(\cdot)$ is not limited to this specific example.

Given the measurement model above, we assume that a probability mass function over the data, state and noise variables, $Pr(Y_t, f(X_t), v_t)$, exists, and:

$$Pr(Y_t|f(X_t)) := \int dv_t Pr(Y_t|f(X_t), v_t) P(v_t) \tag{A3}$$

The likelihood of obtaining a measurement outcome, $Y_t = d, d \in \{0, 1\}$, is given by Eq. (A3).

For a qubit, we know the first term under the integral of Eq. (A3) as the Born probability:

$$Pr(Y_t = d|f(X_t), v_t) \equiv y_t \left(\delta(d-1) - \delta(d)\right) + \frac{1}{2} \tag{A4}$$

By substitution for values of d, we can verify that the expression above is the binomial distribution for a single trial with success probability equal to $y_t + 1/2$ for $d = 1$ ($1/2 - y_t$ for $d = 0$).

The second term in Eq. (A3) refers to the distribution of errors, v_t. Here, v_t is interpreted as our lack of knowledge of the true, unknown $f(X_t)$, and $Pr(v_t)$ represents this uncertainty as additive white Gaussian noise over the real line. With a physical interpretation that $|y_t| \leq \frac{1}{2}$, we replace $Pr(v_t) \to \Xi(v_t)$, where $\Xi(v_t)$ represents Gaussian random variables over a finite interval, and y_t can retain an interpretation as a probability. We now obtain an expression for $Pr(Y_t = d|f(X_t))$ by marginalising over all

values of v_t using $\Xi(v_t)$:

$$Pr(d|f(X_t)) = \int_{-\infty}^{\infty} dv_t \quad Pr(Y_t = d|f(X_t), v_t) \quad \Xi(v_t) \tag{A5}$$

$$= \int_{-\infty}^{\infty} dv_t \left(y_t \left(\delta(d-1) - \delta(d)\right) + \frac{1}{2} \right) \Xi(v_t) \tag{A6}$$

$$= \left(f(X_t) \left(\delta(d-1) - \delta(d)\right) + \frac{1}{2} \right) \int_{-\infty}^{\infty} dv_t \Xi(v_t) + \left(\delta(d-1) - \delta(d)\right) \int_{-\infty}^{\infty} dv_t \quad v_t \Xi(v_t) \tag{A7}$$

$$= \frac{\rho_0}{2} + \rho_0 f(X_t) \left(\delta(d-1) - \delta(d)\right) + \rho_1 \left(\delta(d-1) - \delta(d)\right) \tag{A8}$$

In the equations above, the scaling factors, ρ_0, ρ_1 are given by:

$$\rho_0 \equiv \int_{-\infty}^{\infty} dv_t \quad \Xi(v_t) \tag{A9}$$

$$\rho_1 \equiv \int_{-\infty}^{\infty} dv_t \quad v_t \Xi(v_t) \tag{A10}$$

In the derivation above, since v_t is serially uncorrelated and independent of $f(X_t)$, it is safe to treat $f(X_t)$ outside the integral with respect to v_t. The marginalisation procedure yields two numbers, ρ_0, ρ_1 whose values do not depend on the outcome under consideration, d, or the conditioning random variable, $f(X_t)$. In the case that $\Xi(v_t) \equiv Pr(v_t)$, then ρ_0 is the normalisation factor and is unity by definition and ρ_1 is the mean, which is zero for zero mean noise.

Next, we design the form of the restricted error distribution $\Xi(v_t)$. Numerous algorithms exist for generated band-limited probability distribution functions either in context of studying quantisation or saturation effects in digital signal processing; or in application of dithering noise, anti-aliasing and noise-shaping techniques [32–35]. We use the approximation that the classical quantisation of a continuous variable via physical measurement is statistically comparable to the addition of a independent uniformly distributed random variable [30]. This means that for a qubit, the full error model is a convolution ($\star$) of $Pr(v_t)$ with a uniform distribution, $\mathcal{U}$, defined over the finite interval (a, b), and we define the error model as:

$$\Xi(v_t) \equiv Pr(v_t) \star \mathcal{U}(a, b) \tag{A11}$$

We compute our error model by always considering a symmetric saturation of v_t, such that $a \equiv -b$:

$$\Xi(v_t) = \frac{1}{4b} \mathrm{erf}\left(\frac{b - v_t}{\sqrt{2\Sigma_v}}\right) + \frac{1}{4b} \mathrm{erf}\left(\frac{b + v_t}{\sqrt{2\Sigma_v}}\right) \tag{A12}$$

Under $a = -b$, the integral for $\rho_1 \equiv 0$. This arises because the two error functions contribute equal areas and the linear dependence on v_t means that these areas will have opposite signs for integral limits which are symmetric around zero.

$$Pr(d|f(X_t)) = \frac{\rho_0}{2} + \rho_0 f(X_t) \left(\delta(d-1) - \delta(d)\right) \tag{A13}$$

One computes the Fisher information contribution to yield:

$$I(f(X_t)) = \frac{\rho_0}{(\frac{1}{2} - f(X_t))(\frac{1}{2} + f(X_t))} \tag{A14}$$

In the equation above, the Fisher contribution is that of a coin flip with the probability $\frac{1}{2} + f(X_t)$. This term is scaled by a factor ρ_0 that depends on our level of uncertainty, Σ_v, and the effect of bandlimited random errors, b.

We now use a physical argument to constrain the choice of b, and the regimes for Σ_v, for which our approximation $\Xi(v_t)$ holds. Consider a single bit flip error namely, that a signal is encoded as a 0, instead of a 1. This is the maximum error that can be resolved, for example, in a Ramsey experiment, corresponding to a maximal π accumulation of relative phase between qubit states 0 and 1 under dephasing noise. Hence the maximal resolvable error is such that $|v_t| \leq v_{bound} \equiv 1$. We define the parameter of the independent uniform distribution parameter, b, with respect to this physical bound on the quantised errors:

$$b \equiv \frac{v_{bound}}{2} = \frac{1}{2} \tag{A15}$$

Given this, $b \gtrapprox 3\sqrt{\Sigma_v}$ minimises saturation effects, else a marginalisation using $\Xi(v_t)$ will throw out information. Lastly, we compute ρ_0 as a function of Σ_v, b using the standard integral formulae:

$$\rho_0 \equiv \frac{1}{2} \frac{1}{2b} \int_{-b}^{b} dv_t \quad \left(\mathrm{erf}\left(\frac{b - v}{\sqrt{2\Sigma_v}}\right) + \mathrm{erf}\left(\frac{b + v}{\sqrt{2\Sigma_v}}\right) \right) \tag{A16}$$

$$= \frac{1}{2b} \sqrt{2\Sigma_v} \int_{0}^{\frac{2b}{\sqrt{2\Sigma_v}}} d\tau \quad \mathrm{erf}(\tau) \tag{A17}$$

$$= \frac{1}{2b} \sqrt{2\Sigma_v} \left(\frac{2b}{\sqrt{2\Sigma_v}} \mathrm{erf}\left(\frac{2b}{\sqrt{2\Sigma_v}}\right) + \frac{e^{-(\frac{2b}{\sqrt{2\Sigma_v}})^2}}{\sqrt{\pi}} - \frac{1}{\sqrt{\pi}} \right) \tag{A18}$$

$$\rho_0 = \mathrm{erf}\left(\frac{2b}{\sqrt{2\Sigma_v}}\right) + \frac{\sqrt{2\Sigma_v}}{2b} \frac{e^{-(\frac{2b}{\sqrt{2\Sigma_v}})^2}}{\sqrt{\pi}} - \frac{1}{2b} \frac{\sqrt{2\Sigma_v}}{\sqrt{\pi}} \tag{A19}$$

In the noiseless case, $\Sigma_v \to 0$, $\rho_0 \to 1$ irrespective of the value b, and we recover a Bernoulli experiment parameterised by the Born probability exactly. In this case, setting the noiseless Born probability, $p := \frac{1}{2} + f(X_t)$ with $\rho_0 \to 1$ in Eq. (A14) gives the desired limiting case of Fisher information for a single classical coin flip with probability heads p, $I(p) := 1/(p(1-p))$.

Since the derivation above applies independently to each qubit at any location, j, we substitute Eq. (A19) into Eq. (A13) and equate with Eqs. (11) and (12) in the main text by setting the measurement model to $f(F_t^{(j)}) := \frac{1}{2}\cos(F_t^{(j)})$.

Appendix B: Convergence of particle approximations in non-linear filtering

In this section, we support Section III of the main text by summarising important results in foundational convergence analysis for particle filters following [12]; and providing new derivations as necessary in the analysis of the NMQA framework.

The first step will be to establish the link between Theorem 1 and Theorem 2 in the main text. Of these, Theorem 1 is a re-statement of Theorem 10.7 in [12]. This theorem applies to a fixed observation path. In contrast, the proof below establishes Theorem 2 part (i) by extending Theorem 1 to conditions of convergence for *random* measures. Here, the randomness of measures in Theorem 2 accrues from two sources (a) the random observation vector and (b) the particle approximation for continuous distributions. We will eventually find that the affect of these random sources is to require an additional condition on Theorem 2. This condition is that for some non-zero constant k_t, the expected value of the product $p_t g_t \geq k_t$.

Theorem 4 (Bain & Crisan, 2009)**.** *Assume that for any $t \geq 0$, there exists a constant k_t such that $p_t g_t \geq k_t$. Then for all $f \in B(\mathcal{S}_X)$:*

$$\lim_{n\to\infty} \mathbb{E}[|\pi_t^n f - \pi_t f|] = 0 \tag{B1}$$

$$\lim_{n\to\infty} \mathbb{E}[|p_t^n f - p_t f|] = 0 \tag{B2}$$

hold if and only if the following hold:

$$\lim_{n\to\infty} \mathbb{E}[|\pi_0^n f - \pi_0 f|] = 0 \tag{B3}$$

$$\lim_{n\to\infty} \mathbb{E}[|p_t^n f - K_{t-1}\pi_{t-1}^n f|] = 0 \tag{B4}$$

$$\lim_{n\to\infty} \mathbb{E}[|\pi_t f - \bar{\pi}_t^n f|] = 0 \tag{B5}$$

where $\bar{\pi}_t^n := g_t * p_t^n$ and $\bar{\pi}_t^n f = p_t^n(fg_t)/p_t^n g_t$.

Proof. Assume Eqs. (B1) and (B2) hold. Then Eq. (B1) $\implies$ Eq. (B3) by setting $t = 0$.

By the triangle inequality, for all $f \in B(\mathbb{S}_X)$:

$$\begin{aligned}
&|p_t^n f - K_{t-1}\pi_{t-1}^n f| \\
\leq &|p_t^n f - K_{t-1}\pi_{t-1} f| + |K_{t-1}\pi_{t-1} f - K_{t-1}\pi_{t-1}^n f| && \text{(B6)}\\
\leq &|p_t^n f - p_t f| + |\pi_{t-1}(K_{t-1} f) - \pi_{t-1}^n(K_{t-1} f)| && \text{(B7)}
\end{aligned}$$

Taking expectations of both sides as $n \to \infty$, the two terms on the right hand side are zero from Eqs. (B1) and (B2) and we recover Eq. (B4).

For Eq. (B5), require that:

$$\mathbb{E}[|\pi_t f - \bar{\pi}_t^n f|] \leq \mathbb{E}[|\pi_t f - \pi_t^n f|] + \mathbb{E}[|\pi_t^n f - \bar{\pi}_t^n f|] \tag{B8}$$

We recover Eq. (B5) if $\mathbb{E}[|\pi_t^n f - \bar{\pi}_t^n f|] = 0$. We defer this until later in the proof below.

In the reverse direction, we obtain Eq. (B2) by applying the triangle inequality, and Eq. (B1) follows by induction from using Eq. (B3) as a starting point and Eqs. (B2) and (B4).

Hence, we need only to show $\mathbb{E}[|\pi_t^n f - \bar{\pi}_t^n f|] = 0$. To see this, we rearrange the expressions as follows:

$$\begin{aligned}
\pi_t f - \bar{\pi}_t^n f =& \frac{p_t(fg_t)}{p_t g_t} - \frac{p_t^n(fg_t)}{p_t^n g_t} && \text{(B9)}\\
=& \frac{p_t(fg_t)}{p_t g_t} - \frac{p_t^n(fg_t)}{p_t^n g_t} \\
&+ \frac{p_t^n(fg_t)}{p_t g_t} - \frac{p_t^n(fg_t)}{p_t g_t} && \text{(B10)}\\
=& \frac{1}{p_t g_t}\left(p_t(fg_t) - p_t^n(fg_t)\right) \\
&+ \frac{p_t^n(fg_t)}{p_t g_t} - \frac{p_t^n(fg_t)}{p_t^n g_t} && \text{(B11)}\\
=& \frac{1}{p_t g_t}\left(p_t(fg_t) - p_t^n(fg_t)\right) \\
&+ \frac{p_t^n(fg_t)}{p_t g_t p_t^n g_t}\left(p_t^n g_t - p_t g_t\right) && \text{(B12)}
\end{aligned}$$

Denote $\frac{p_t^n(fg_t)}{p_t^n g_t} \leq \| f_\infty \|$ and talking the absolute values and expectation on both sides yields:

$$\begin{aligned}
&\mathbb{E}[|\pi_t f - \bar{\pi}_t^n f|] \\
\leq &\mathbb{E}[\frac{1}{p_t g_t}|p_t(fg_t) - p_t^n(fg_t)|] \\
&+ \| f_\infty \| \, \mathbb{E}[\frac{1}{p_t g_t}|p_t^n g_t - p_t g_t|] && \text{(B13)}\\
\leq &\mathbb{E}[\frac{1}{k_t}|p_t(fg_t) - p_t^n(fg_t)|] \\
&+ \| f_\infty \| \, \mathbb{E}[\frac{1}{k_t}|p_t^n g_t - p_t g_t|] && \text{(B14)}\\
\leq &\frac{1}{k_t}\mathbb{E}[|p_t(fg_t) - p_t^n(fg_t)|] \\
&+ \frac{\| f_\infty \|}{k_t}\mathbb{E}[|p_t^n g_t - p_t g_t|] && \text{(B15)}
\end{aligned}$$

Here, we invoked the assumption that there exists a constant k_t such that $p_t g_t > k_t > 0$, such that $\frac{1}{k_t}$ is greater than $\frac{1}{p_t g_t}$, and this allows us to bring k_t outside the expectation value.

In the limit $n \to \infty$, both terms on the right hand side of the last line go to zero by Eq. (B2). □

The derivation above establishes Theorem 2 (i), and hence it establishes that Theorem 1 applies to random measures.

We now provide additional commentary around the requirements for the transition kernel and likelihood function which appear in both Theorem 2 (ii) and (iii) [12]. To obtain 'if and only if' conditions for convergence in expectation using Theorem 1, one needs to additionally argue that in proceeding to step Eq. (B7) that the transition kernel is Feller; and secondly, the limit in Eq. (B15) for both terms on the right-hand side is zero by invoking the continuity and boundedness of the likelihood function. This re-states Corollary 10.10 in [12] (fixed observation vector). In a similar manner, Corollary 10.30 in [12] (random observation vector) should be read as invoking these additional requirements on the likelihood function and transition kernel. These restrictions on the likelihood function and transition kernel enable us to interpret Theorem 2(i) and Theorem 1 as an 'if and only if' statement about convergence in expectation.

Next, the following theorem establishes conditions for almost-sure convergence of p_t^n to p_t and π_t^n to π_t, encapsulated by Theorem 2 (iii).

Theorem 5 (Bain & Crisan, 2009)**.** *Assume that the transition kernel for X is Feller, and the likelihood functions are continuous. Then the sequence p_t^n converges to p_t and π_t^n converges to π_t almost surely, for all $t \geq 0$, if and only if:*

$$\lim_{n\to\infty} \pi_0^n = \pi_0 \quad \mathbb{P} - \text{a.s.} \tag{B16}$$

$$\lim_{n\to\infty} d(p_t^n, K_{t-1}\pi_{t-1}^n) = 0 \quad \mathbb{P} - \text{a.s.} \tag{B17}$$

$$\lim_{n\to\infty} d(\pi_t^n, \bar{\pi}_t^n) = 0 \quad \mathbb{P} - \text{a.s.} \tag{B18}$$

where $d(\cdot,\cdot)$ is any metric that generates a weak topology on the space of finite measures, and $\mathcal{M}$ is a convergence determining set in $C_b(\mathbb{S}_X)$.

Proof. This is a re-statement of Theorem 10.12 (fixed observation vector) and Proposition 15 in [12] (random observation vector); proofs can be found within the reference. □

In the above, the first condition establishes that we start from a good approximation of π_0. The second and third conditions state two requirements which must be met in satisfying the recursion relations for our approximations of the true posterior π_t^n. The second condition states that the empirical predictive distribution and the transition kernel enable us to track the posterior from t to $t+1$ 'closely enough'. The third condition states that empirical distributions before ($\bar{\pi}_t^n$) and after (π_t^n) the branching mechanism in the particle measure do not deviate within each iteration t.

Of interest is the the branching mechanism - while it can be designed arbitrarily, the branching mechanism satisfying Proposition 3 has desirable properties for convergence, including the identities below:

Lemma 6. *The following identities hold for the empirical distributions of a particle filter with the branching mechanism described in Proposition 3.*

$$\mathbb{E}[p_t|\mathcal{G}_{t-1}] = K_{t-1}\pi_{t-1}^n \tag{B19}$$

$$\pi_t^n = \frac{1}{n}\sum_{i=1}^n \xi_t^{(i)} \delta(\bar{x}_t^{(i)}) \tag{B20}$$

Proof. The first identity is obtained by substituting definitions for empirical measures as:

$$p_t^n := \frac{1}{n}\sum_{i=1}^n \delta(\bar{x}_t^{(i)}) = \frac{1}{n}\sum_{i=1}^n \delta(K_{t-1}x_{t-1}^{(i)}) \tag{B21}$$

$$\implies \mathbb{E}[p_t|\mathcal{G}_{t-1}] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}[\delta(K_{t-1}x_{t-1}^{(i)})|\mathcal{G}_{t-1}] \tag{B22}$$

$$= \frac{1}{n}\sum_{i=1}^n K_{t-1}\delta(x_{t-1}^{(i)}) \tag{B23}$$

$$= K_{t-1}\frac{1}{n}\sum_{i=1}^n \delta(x_{t-1}^{(i)}) \tag{B24}$$

$$= K_{t-1}\pi_{t-1}^n \tag{B25}$$

For the second identity, each particle $\bar{x}_t^{(i)}$ replaces itself $\xi_t^{(i)}$ number of times, that is, $\bar{x}_t^{(i)}$ is counted up $\xi_t^{(i)}$ number of times in the re-sampled posterior π_t^n. Hence we can write the resampled posterior as:

$$\pi_t^n \equiv \sum_{i=1}^n \mathbb{P}(\bar{x}_t^{(i)}\text{chosen}) \cdot \delta(\bar{x}_t^{(i)}) \tag{B26}$$

$$= \sum_{i=1}^n \left(\frac{\xi_t^{(i)}}{\sum_{k=1}^n \xi_t^{(k)}}\right)\delta(\bar{x}_t^{(i)}) \tag{B27}$$

$$= \sum_{i=1}^n \left(\frac{\xi_t^{(i)}}{n}\right)\delta(\bar{x}_t^{(i)}) \tag{B28}$$

The last line follows from the previously stated assumption that the number of particles are held constant after each re-sampling step. □

We now link the branching mechanism of the particle filter to the main convergence theorems listed previously.

Theorem 7 (Bain & Crisan, 2009)**.** *Assume that for all t, there exists a constant k_t such that $p_t g_t \geq k_t$ and the covariance matrix of a branching mechanism satisfies $q^t A_t^n q \leq n c_t$. Then for a random observation vector $\pi_t^n \equiv \pi_t^{n,Y_{0:t}}$ and $p_t^n \equiv p_t^{n,Y_{0:t}}$ yield:*

$$\lim_{n\to\infty} \mathbb{E}[|\pi_t^n f - \pi_t f|] = 0 \tag{B29}$$

$$\lim_{n\to\infty} \mathbb{E}[|p_t^n f - p_t f|] = 0 \tag{B30}$$

for all $f \in B(\mathbb{S}_X), t \geq 0$.

Proof. Following the proof structure suggested in [12], we use Theorem 4. First, we observe show convergence for the initial conditions. Since $\delta(x_0^{(i)}) \sim \pi_0$ and

$\pi_0^n(1) = 1, \forall n$, then the dominated convergence theorem for measure valued random variables applies:

$$\lim_{n\to\infty} \pi_0^n = \pi_0 \implies \lim_{n\to\infty} \mathbb{E}[|\pi_0^n f - \pi_0 f|] = 0, \quad \mathbb{P} - \text{a.s..} \tag{B31}$$

This verifies Eq. (B3). For Eq. (B4), we use Eq. (B19), $\mathbb{E}[p_t f|\mathcal{G}_{t-1}] = \pi_{t-1}^n(K_{t-1}f) \equiv \mu$.

$$\begin{aligned}
&\mathbb{E}[(p_t^n f - \pi_{t-1}^n(K_{t-1}f))^2|\mathcal{G}_{t-1}] \\
=&\mathbb{E}[(p_t^n f)^2|\mathcal{G}_{t-1}] - \mu^2 && \text{(B32)}\\
=&\mathbb{E}[(p_t^n f)^2|\mathcal{G}_{t-1}] - \pi_{t-1}^n(K_{t-1}f)^2 && \text{(B33)}\\
=&\frac{1}{n^2}\mathbb{E}[\sum_{i=1}^n\sum_{j=1}^n f(\bar{x}_t^{(i)})f(\bar{x}_t^{(j)})|\mathcal{G}_{t-1}] \\
&- \pi_{t-1}^n(K_{t-1}f)^2 && \text{(B34)}\\
=&\frac{1}{n^2}\sum_{i=1}^n \mathbb{E}[(f(\bar{x}_t^{(i)}))^2|\mathcal{G}_{t-1}] \\
&- \pi_{t-1}^n(K_{t-1}f)^2 && \text{(B35)}\\
=&\frac{1}{n}\pi_{t-1}^n\left(K_{t-1}f^2 - (K_{t-1}f)^2\right) && \text{(B36)}\\
\implies &\mathbb{E}[(p_t^n f - \pi_{t-1}^n(K_{t-1}f))^2] \leq \frac{\parallel f \parallel_\infty^2}{n} && \text{(B37)}
\end{aligned}$$

In going from Eq. (B34) to Eq. (B35), the sums in the first term are brought outside of the expectation value due to independence of samples $\bar{x}_t^{(i)}$ conditional on $\mathcal{G}_{t-1}$. The second term invokes both linearity of expectation values and independence of the predictive samples at t conditional on $\mathcal{G}_{t-1}$. The first term in Eq. (B36) follows by applying the same reasoning to Eq. (B21). Since f is a bounded, Borel measurable function [16], we assume there exists f_∞, a the limiting vector for all t, n such that $\pi_t^n f \leq \parallel f \parallel_\infty, \forall n$ [12]. This assumption is used in the last step.

The limit in Eq. (B4) is implied by the last line since for any random variable, x, $0 \leq \mathbb{V}\{x\} \leq \mathbb{E}[x^2]$. Almost sure convergence of the second moment and variance of x to zero implies $\mathbb{E}[x]^2$ almost surely; letting $x := p_t^n f - \pi_{t-1}^n(K_{t-1}f)$ implies Eq. (B4).

Next, we derive an analogous result for $\mathbb{E}[(\pi_t^n - \bar{\pi}_t^n)^2]$, by re-writing key quantities in terms of n dimensional vectors whose elements represent individual particles:

$$\begin{aligned}
(\pi_t^n - \bar{\pi}_t^n)^2 &:= (W_t f_t)^T(W_t f_t) && \text{(B38)}\\
f_t &:= f(\delta(\bar{x}_t)), && \text{(B39)}\\
\delta(\bar{x}_t) &= [\delta(\bar{x}_t{}^{(1)}), \ldots, \delta(\bar{x}_t{}^{(n)})]^T && \text{(B40)}\\
W_t &:= \frac{\xi_t - n w_t}{n} && \text{(B41)}\\
\xi_t &:= [\xi_t^{(1)}, \ldots, \xi_t^{(n)}] && \text{(B42)}\\
w_t &:= [w_t^{(1)}, \ldots, w_t^{(n)}] && \text{(B43)}
\end{aligned}$$

Since f is a bounded, Borel measurable function, we assume there exists $f_\infty \geq f(\delta(\bar{x}_t))$ i.e. the limiting vector for all t, n.

$$\begin{aligned}
&\mathbb{E}[(\pi_t^n - \bar{\pi}_t^n)^2] \\
&:= \mathbb{E}[(W_t f_t)^T(W_t f_t)] && \text{(B44)}\\
&= \frac{1}{n^2}\mathbb{E}[f_t^T(\xi_t - n w_t)^T(\xi_t - n w_t)f_t] && \text{(B45)}\\
&\leq \frac{1}{n^2}\mathbb{E}[f_\infty^T(\xi_t - n w_t)^T(\xi_t - n w_t)f_\infty] && \text{(B46)}\\
&= \frac{1}{n^2} f_\infty^T \mathbb{E}[(\xi_t - n w_t)^T(\xi_t - n w_t)] f_\infty && \text{(B47)}\\
&= \frac{1}{n^2} f_\infty^T A_t^n f_\infty && \text{(B48)}\\
&= \frac{\parallel f \parallel_\infty^2}{n^2} \frac{f_\infty^T}{\parallel f \parallel_\infty} A_t^n \frac{f_\infty}{\parallel f \parallel_\infty} && \text{(B49)}\\
&= \frac{\parallel f \parallel_\infty^2}{n^2} q^T A_t^n q, \quad q = \frac{f_\infty}{\parallel f \parallel_\infty} && \text{(B50)}\\
&\leq \frac{\parallel f \parallel_\infty^2}{n^2} c_t n = \frac{c_t \parallel f \parallel_\infty^2}{n} && \text{(B51)}\\
&\implies \mathbb{E}[(\pi_t^n - \bar{\pi}_t^n)^2] \leq \frac{c_t \parallel f \parallel_\infty^2}{n} && \text{(B52)}
\end{aligned}$$

In the above, we observe that the definition of W_t invokes Eq. (B20). If Eq. (B4) holds, then Eq. (B5) is true by Eq. (B52) and the proof given under Theorem 4. □

Theorem 7 establishes the link between branching mechanisms in Proposition 3 and Theorem 2 in the main text.

We now focus on branching mechanisms that are multinomial random processes. In this special case, Proposition 3 are satisfied and the constant c_t can be given a value to further specify the rate of convergence. Both of these results can be found in [12] and they are listed for completeness below.

Lemma 8 (Bain & Crisan, 2009)**.** *If the offspring distributions are multinomial, then all properties of the branching mechanism are satisfied.*

Proof. Let $\xi_t^{(i)}$ be a multinomially distributed where the probability of choosing particle $\bar{x}_t^{(i)}$ is proportional to its weight $w_t^{(i)}$

$$\mathbb{P}(\xi_t^{(i)} = m^{(i)}) := \frac{n!}{\prod_{i=1}^k m^{(i)}!}\prod_{i=1}^k (w_t^{(i)})^{m^{(i)}} \tag{B53}$$

Then by the properties of the multinomial distribution, we obtain expectation values of means and covariances matrix that satisfy the assumptions of the required

branching mechanisms:

$$\mathbb{E}[\xi_t^{(i)}|\bar{\mathcal{G}}_t] = nw_t^{(i)} \tag{B54}$$

$$A_t^n := \mathbb{E}[(\xi_t - nw_t)^T(\xi_t - nw_t)] \tag{B55}$$

$$(A_t^n)_{i,j} = \begin{cases} nw_t^{(i)}(1-w_t^{(i)}), & i=j \\ -nw_t^{(i)}w_t^{(j)}, & i\neq j \end{cases} \tag{B56}$$

$$\xi_t := [\xi_t^{(1)}, \ldots, \xi_t^{(n)}] \tag{B57}$$

$$w_t := [w_t^{(1)}, \ldots, w_t^{(n)}] \tag{B58}$$

$$\implies q^T A_t^n q_t := \sum_{i=1}^n \sum_{j=1}^n q_i q_j (A_t^n)_{i,j} \tag{B59}$$

$$= \sum_{i=1}^n q_i q_i n w_t^{(i)}(1-w_t^{(i)}) - \sum_{i\neq j}^n \sum_{j=1}^n n q_i q_j w_t^{(i)} w_t^{(j)} \tag{B60}$$

$$= \sum_{i=1}^n q_i q_i n w_t^{(i)} - \sum_{i=1}^n \sum_{j=1}^n n q_i q_j w_t^{(i)} w_t^{(j)} \tag{B61}$$

$$= n\sum_{i=1}^n (q_i)^2 w_t^{(i)} - n(\sum_{i=1}^n q_i w_t^{(i)})^2 \tag{B62}$$

$$\leq n\sum_{i=1}^n (q_i)^2 w_t^{(i)} \tag{B63}$$

$$\leq n\sum_{i=1}^n w_t^{(i)}, \quad |q_i| \leq 1 \forall i \tag{B64}$$

$$= n \tag{B65}$$

$$\implies c_t = 1 \quad \forall t \tag{B66}$$

□

The results thus far establish Theorem 2 (ii).

Corollary 9. *If the off-spring distributions are multi-nomial and under the conditions of Theorem 5 for the random observation vector, then*

$$\lim_{n\to\infty} p_t^n = p_t, \quad \mathbb{P}-\text{a.s.} \tag{B67}$$

$$\lim_{n\to\infty} \pi_t^n = \pi_t, \quad \mathbb{P}-\text{a.s.} \tag{B68}$$

Proof. This is a partial restatement of Corollary 10.31 in [12]. It follows from Theorem 5 and Theorem 7. □

The corollary above additionally establishes almost-sure convergence for Theorem 2 part (iii).

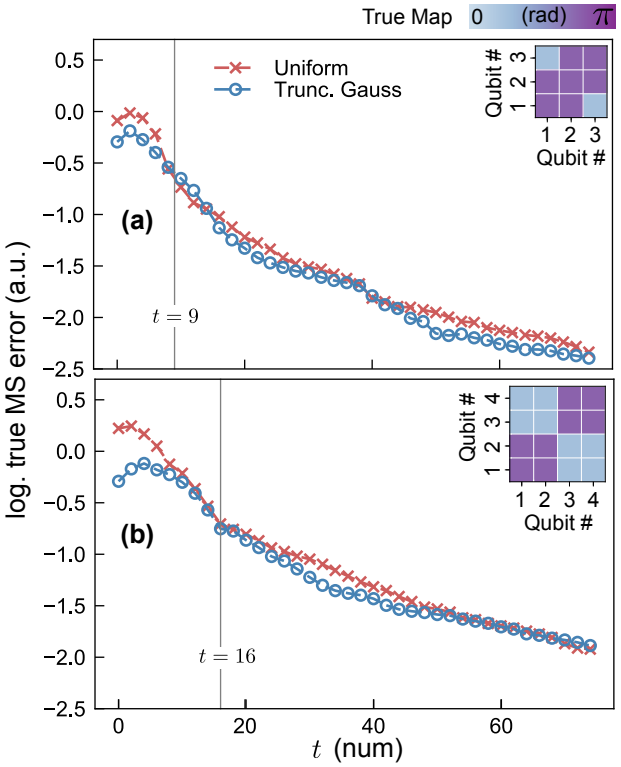

FIG. 5. NMQA performance by choice of 'Uniform' v. 'Trunc. Gaussian' expansion strategy for system size $d = 9, 16$. 2D array with a square field depicted as heat-maps of qubit phases (insets of (a)-(b)), with low and high field values at 0.25π and 0.75π radians respectively. (a)-(b) main panels are semi-log plots where the y-axis is the log of the expected mean square map reconstruction error per qubit, over 50 runs; x-axis represents the number of iterations, t. Each iteration corresponds to one physical measurement. Uniform (Trunc. Gaussian) expansion strategy shown as red crosses (blue circles). Vertical grey lines mark t for which number of physical measurements equals system size, $t = d$. Tuned parameters $(\Sigma_v, \Sigma_F, \lambda_1, \lambda_2)$: Uniform: (a) $(7.1e^{-7}, 0.05, 0.93, 0.68)$, $n_\alpha = 9$; (b) $(4.2e^{-3}, 2.6e^{-4}, 0.88, 0.72)$, $n_\alpha = 30$. Trunc. Gaussian: (a) $(6.3e^{-7}, 7.9e^{-7}, 0.95, 0.84)$, $n_\alpha = 3$; (b) $(4.2e^{-3}, 2.6e^{-4}, 0.93, 0.68)$, $n_\alpha = 30$. For all data, $n_\beta = \frac{2}{3}n_\alpha$.

Proposition 10. *Assume Markov X_t with identity dynamics; bounded, strictly positive continuous likelihood functions, and a constant particle number from t to $t+1$. In NMQA, the first resampling of $\bar{\pi}_t^{n=n_\alpha n_\beta}$ and the second re-sampling of $\bar{\pi}_t^{n=n_\alpha}$ can be viewed as a one-step multi-nomial re-sampling process within each t.*

Proof. To see this, consider the re-sampling mechanism at iteration t. Let N be the total number of samples from a multi-nomial distribution. Let the set $\{j = 1, 2, \ldots, n_\alpha n_\beta\}$ denote the full set of particles for both α and β layers, where the probability of success $w_t^{(j)}$ is given by g_t for the j-th particle.

Let the set of labels $\{j = 1, 2, \ldots, n_\alpha n_\beta\}$ be partitioned into non-empty set of $\{A_i, i = 1, 2, \ldots, \kappa\}$ categories, where $\kappa < n_\alpha n_\beta$. Then the total number of

particles is conserved over this re-grouping:

$$N = \sum_{j=1}^{n_\alpha n_\beta} m^{(j)} = \sum_{i=1}^{\kappa} h^{(i)} = \sum_{i=1}^{\kappa} \sum_{j \in A_i} m^{(j)} \tag{B69}$$

$$h^{(i)} := \sum_{j \in A_i} m^{(j)} \tag{B70}$$

Meanwhile, since A_i represents union of original categories, the success probability, $\bar{w}_t^{(i)}$, of seeing $h^{(i)}$ counts in the i-th regrouped category is the sum over the probabilities of the original categories, $w_t^{(j)}$, for all $j \in A_i$ i.e.

$$\bar{w}_t^{(i)} = \sum_{j \in A_i} w_t^{(j)} \tag{B71}$$

Then for N constant, and by the multi-nomial theorem, the random variables $H_t^{(i)} = h^{(i)}$, where $H_t^{(i)} = \sum_{j \in A_i} \xi_t^{(j)}$ are multinomially distributed with success probabilities $\bar{w}_t^{(i)}$. Setting $N = n_\alpha$ for all re-sampling steps means that the branching mechanism in NMQA remains a multi-nomial process from the beginning to the end of each t. □

The result above establishes the claim in Section III that the branching mechanism for NMQA is a multi-nomial random process, with $c_t = 1$ in Theorem 7.

Appendix C: Supporting Numerical Analysis

We will present additional numerical evidence to support the conclusions of the main text. Our analysis focuses on the error scaling behaviour for NMQA with system size.

In panels (a)-(b) of Fig. 5, we change the system size for a 2D Square Field from $d = 9, 16$, with true maps displayed as heat-maps of qubit phase in the right inset. Vertical lines mark $t = d$. True expected mean square error per qubit over 50 runs is depicted for Uniform (Trunc. Gaussian) in red crosses (blue circles) as the number of iterations t increase. These error traces show two distinctive regimes of behaviour for $t < d$ (sparse data) with both rising and decreasing regions of error. For $t > d$, a different rate of decrease in error is observed; with the change of scaling behaviour more pronounced for large d. The case for $t = 25$ is provided in the main text.

In Fig. 6, we plot the error scaling factor for particle number, Δ, against number of iterations, t, for the two system sizes $d = 9, 16$. The behaviour of Δ also depicts features at $t \approx d$ suggesting two different regimes of behaviour apply, i.e. $t < d$ (sparse data) and $t \gg d$ (high data limit). The data for Uniform (Trunc. Gaussian) is given in red crosses (blue circles). Our results agree with our theoretical expectations set by Theorem 2: in the high t regime, $\Delta > 0$ for Uniform and the condition $\Delta \in [-1, 0)$ is satisfied for Trunc. Gaussian.

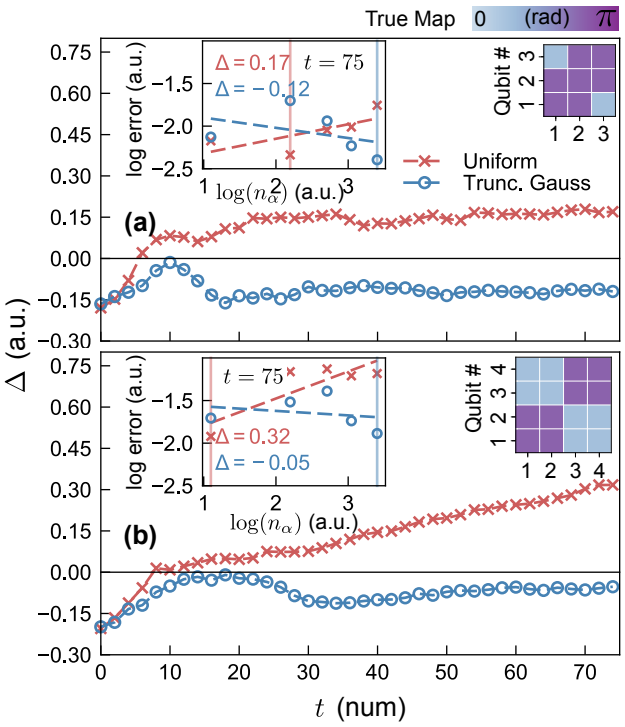

FIG. 6. Error scaling behaviour for Uniform and Trunc. Gaussian. Rows represent increasing qubit system size $d = 9, 16$ for 2D Square Field (right insets); with high and low qubit phase values of $0.25\pi, 0.75\pi$ radians depicted on heat-maps. (a)-(b) main panels depict Δ against t for tuned NMQA parameters. $\Delta > 0$ for Uniform; $\Delta \in [-1, 0)$ for Trunc. Gaussian for $t \gg d$ agrees with Theorem 2. Data for Uniform (crosses) and Trunc. Gaussian (open circles). Left insets depict the log of the expected mean square map reconstruction error per qubit over 50 runs against the log of n_α number of α-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 $t = 75$. Δ is the gradient of the line of best fit (dashed lines). Vertical colored lines mark tuned particle configuration plotted in Fig. 5. Tuned parameters $(\Sigma_v, \Sigma_F, \lambda_1, \lambda_2)$ for Uniform: (a) $(7.1e^{-7}, 0.05, 0.93, 0.68)$; (b) $(4.2e^{-3}, 2.6e^{-4}, 0.88, 0.72)$. Trunc. Gaussian: (a) $(6.3e^{-7}, 7.9e^{-7}, 0.95, 0.84)$; (b) $(4.2e^{-3}, 2.6e^{-4}, 0.93, 0.68)$.

In Fig. 7, we provide scaling behaviour information for the case $\lambda_1 = \lambda_2 = 0$ for all cases discussed in the main text and this appendix. Here, we have turned off the sharing mechanism in NMQA by setting λ_1, λ_2 to zero; where this choice leads to larger true expected mean square error per qubit than picking non-zero λ_1, λ_2. Nevertheless, despite the sub-optimal choice of setting $\lambda_1 = \lambda_2 = 0$, this is not a difficult regime to analyse theoretically. From Section III in the main text, we expect Theorem 2 to hold under Trunc. Gaussian for $\lambda_1 = \lambda_2 = 0$. Numerically, we confirm that $\Delta \in [-1, 0)$ is indeed satisfied for Trunc. Gaussian for $t > d$ in a variety of physical 2D configurations in Fig. 7(a)-(d). The expectation that $\Delta > 0$ for Uniform for $t > d$ is additionally satisfied. Similar results hold in 1D.

Collectively, these numerical results provide additional support for the conclusions of the main text.

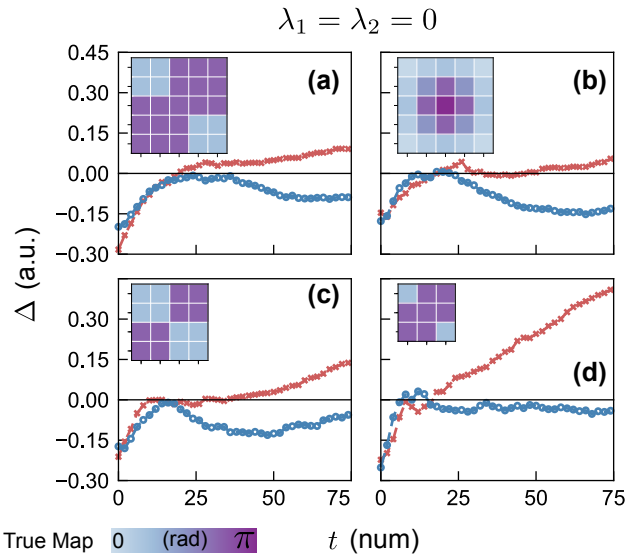

FIG. 7. Error scaling behaviour for $\lambda_1 = \lambda_2 = 0$. Panels represent different physical configurations (left insets): (a), (c), (d) 2D Square field for $d = 25, 16, 9$ respectively; (b) 2D Gaussian field for $d = 25$. (a)-(d) main panels depict Δ against t for tuned NMQA parameters. Data for Uniform (crosses) and Trunc. Gaussian (open circles); $\Delta > 0$ for Uniform; $\Delta \in [-1, 0)$ for Trunc. Gaussian for $t \gg d$ agrees with Theorem 2. Tuned variance parameters (Σ_v, Σ_F) for Uniform: (a) $(7.1e^{-7}, 0.047)$; (b) $(5.9e^{-9}, 0.096)$; (c) $(4.2e^{-3}, 2.6e^{-4})$; (d) $(7.1e^{-7}, 0.047)$. Trunc. Gaussian: (a) $(8.9e^{-7}, 1.9e^{-9})$; (b) $(0.77, 4.6e^{-6})$; (c) $(4.2e^{-3}, 2.6e^{-4})$; (d) $(6.3e^{-7}, 7.9e^{-7})$. $n_\alpha = 3, 9, 15, 21, 30; n_\beta = \frac{2}{3} n_\alpha$. High and low true field values of $0.25\pi, 0.75\pi$ radians.