



PAPER

Flexible resources for quantum metrology

OPEN ACCESS

RECEIVED
1 March 2017REVISED
20 April 2017ACCEPTED FOR PUBLICATION
5 May 2017PUBLISHED
30 June 2017

Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/4.0/).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

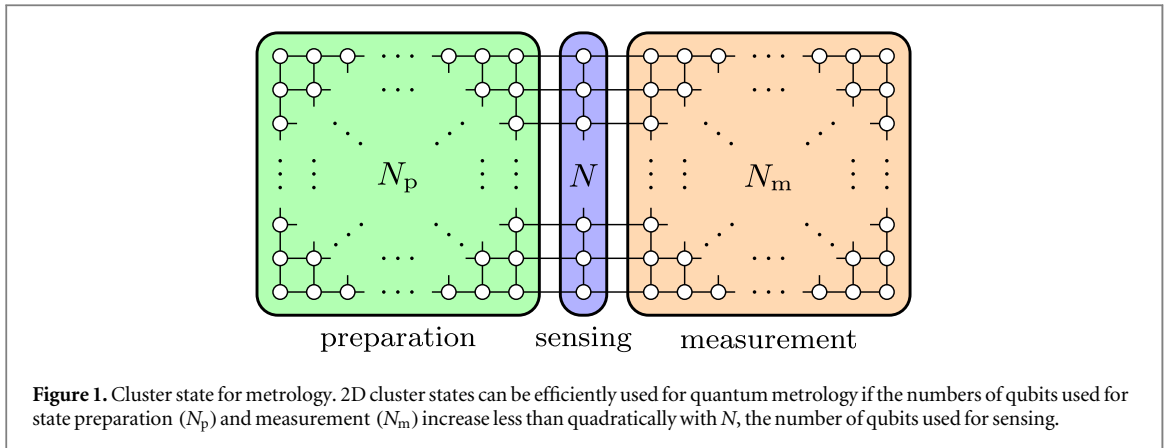
Nicolai Friis^{1,2} , Davide Orsucci², Michalis Skotiniotis^{2,3} , Pavel Sekatski², Vedran Dunjko² ,
Hans J Briegel^{2,4} and Wolfgang Dür²¹ Institute for Quantum Optics and Quantum Information, Austrian Academy of Sciences, Boltzmanngasse 3, A-1090 Vienna, Austria² Institute for Theoretical Physics, University of Innsbruck, Technikerstraße 21a, A-6020 Innsbruck, Austria³ Física Teòrica: Informació i Fenòmens Quàntics, Departament de Física, Universitat Autònoma de Barcelona, E-08193 Bellaterra, Spain⁴ Department of Philosophy, University of Konstanz, 78457 Konstanz, GermanyE-mail: nicolai.friis@univie.ac.at**Keywords:** quantum metrology, quantum computers, cluster states, measurement-based quantum computation**Abstract**

Quantum metrology offers a quadratic advantage over classical approaches to parameter estimation problems by utilising entanglement and nonclassicality. However, the hurdle of actually implementing the necessary quantum probe states and measurements, which vary drastically for different metrological scenarios, is usually not taken into account. We show that for a wide range of tasks in metrology, 2D cluster states (a particular family of states useful for measurement-based quantum computation) can serve as flexible resources that allow one to efficiently prepare any required state for sensing, and perform appropriate (entangled) measurements using only single qubit operations. Crucially, the overhead in the number of qubits is less than quadratic, thus preserving the quantum scaling advantage. This is ensured by using a compression to a logarithmically sized space that contains all relevant information for sensing. We specifically demonstrate how our method can be used to obtain optimal scaling for phase and frequency estimation in local estimation problems, as well as for the Bayesian equivalents with Gaussian priors of varying widths. Furthermore, we show that in the paradigmatic case of local phase estimation 1D cluster states are sufficient for optimal state preparation and measurement.

1. Introduction

Quantum metrology is positioned at the forefront of modern quantum sciences, spearheading the development of future quantum technologies. By utilising the power of quantum mechanics to gain advantages over previously known techniques in practical tasks such as parameter estimation [1–4], state discrimination [5], or hypothesis testing [6], quantum-enhanced measurement procedures have already led to breakthrough discoveries [7, 8]. Moreover, nonclassical effects can be harnessed to enhance the precision of determining quantities of interest, including magnetic fields [9, 10], forces [11, 12], phases [13, 14], or frequencies [15–17]. For many different applications, the quantum advantage manifests as a quadratic scaling gap in terms of the relevant resources [18–21], e.g., the number of sensing systems, with respect to the best classical approaches. However, to achieve this so-called Heisenberg scaling, different tasks require different resource states as well as different (potentially non-local) measurements, which have to be separately determined for any specific case, rendering the design of a universally applicable, optimal sensing device difficult. Moreover, this still leaves open the important (and often ignored) question of how the desired states and measurements can be implemented efficiently.

Here we report on the design of a flexible device that allows one to obtain a quantum scaling advantage for a large class of different metrological problems by using only a specific entangled state and single-qubit operations. We show that a 2D cluster state [22, 23]—a particular entangled state associated with a rectangular lattice that can be prepared by commuting, nearest-neighbour (NN) interactions among qubits on the lattice—allows achieving Heisenberg scaling for an important group of paradigmatic metrology problems. This includes the sensing of local observables such as magnetic fields [9, 10], as well as the estimation of phases [18, 19],



frequencies [15–17], and certain interaction strengths [24]. Crucially, we show that this can be done both in the local (frequentist) approach with arbitrarily many repetitions, and in the (single-shot) Bayesian approach for arbitrary cost functions and priors (see, e.g., [25]), including flat [19–21] and Gaussian priors with varying width [26, 27]. The key difference between these estimation problems lies in the incorporation of *a priori* available knowledge about the estimated parameter. In local estimation, no quantification of prior knowledge is required in principle, but it is often assumed that fluctuations around a well-known value of the parameter are being estimated in order to make use of the quantum Fisher information (QFI) as a relevant figure of merit. In Bayesian estimation, the initial information is encoded in a prior probability distribution that is updated according to Bayes’ law after each individual measurement.

The optimal probe state for these different problems vary strongly, ranging from Greenberger–Horne–Zeilinger (GHZ) states in the case of local phase estimation, to certain superpositions of states with different Hamming weights (e.g., with sine-shaped profiles for the coefficients [19]) for Bayesian phase estimation (for flat priors). Moreover, also the corresponding optimal measurements are vastly different, including simple local measurements for GHZ states, but also complicated, entangled measurements on all qubits [28, 29], e.g., discrete Fourier basis measurements for Bayesian estimation with flat priors [19]. In particular, some states and measurements may be significantly more difficult to realise than others.

The 2D cluster state allows one to deal with all of these problems. On the one hand, the fact that it is a universal resource for measurement-based quantum computation (MBQC) [30, 31] trivially enables arbitrary state preparation and measurements on a subset of the qubits in the cluster, provided the latter is large enough. On the other hand, MBQC provides a simple, unifying framework in which state preparation and measurements can be assigned an unambiguous resource cost in terms of the overall number of qubits in the cluster⁵, as illustrated in figure 1. To guarantee a quantum scaling advantage for metrological applications, the probe preparation and measurements must be efficiently executable. That is, any metrological scaling advantage is lost if the size of the cluster required for a given estimation strategy with an N -qubit probe grows as N^2 or stronger, in which case it becomes favourable to use all qubits in the cluster as individual, classical probes instead.

We show that the preparation of optimal probe states and corresponding suitable measurements for local as well as Bayesian phase and frequency estimation can indeed be carried out efficiently using 2D cluster states. For the local scenario, we explicitly construct the preparation and measurement strategy achieving optimality. For the Bayesian scenario, we present a construction that can generate all optimal probe states with a linear overhead in N . We then introduce a compression procedure that can be implemented on a 2D cluster with $O(N \log^2 N)$ qubits, which enables one to efficiently perform measurements even when the circuit descriptions of the corresponding unitaries are of exponential size in the number of qubits of the compressed space. These constructions allow achieving Heisenberg scaling for phase and frequency estimation scenarios using the 2D cluster in a flexible manner. Crucially, this flexibility holds the potential for yielding (nearly) optimal scaling performance for a variety of estimation problems, and hence goes beyond the capabilities of architectures dedicated to specific individual tasks [33]. To further illustrate these general results, we discuss a particular choice of probe states and measurements that can be efficiently implemented in our framework, for which Heisenberg scaling can be achieved for Gaussian priors of varying widths.

This paper is structured as follows. In section 2 we first discuss the basic structure of parameter estimation problems and the general form of all optimal probe states. We then argue that 2D cluster states provide flexible resources to achieve Heisenberg scaling in phase and frequency estimation problems by using an efficient

⁵ Note that, in a different context, cluster states have previously also been used for specific metrology tasks directly (i.e., using all qubits for sensing) [32].

compression to the subspace of the optimal probes. In section 3 we then show how Heisenberg scaling can be achieved in Bayesian phase (and frequency) estimation, before demonstrating in section 4 how the necessary probe states can be prepared in a measurement-based architecture consisting of $O(N)$ qubits. Finally, we introduce the explicit construction of the efficient compression algorithm required for the measurements in section 5. At last, we discuss our findings and their implications in section 6, including generalisation to the estimation of quantities other than phases and frequencies.

2. 2D Cluster states as universal resources for phase and frequency estimation

2.1. Parameter estimation problems

In typical parameter estimation procedures, one wishes to determine an unknown parameter θ that is not directly measurable. To this end, a probe state described by a density operator ρ_o is prepared, which undergoes a dynamical evolution governed by θ , encoding the parameter in the resulting state $\rho(\theta)$. The evolution can in principle be an arbitrary quantum channel but we are here mainly interested in pure states $\rho_o = |\psi\rangle\langle\psi|$ and unitary channels, where

$$\rho(\theta) = U_\theta |\psi\rangle\langle\psi| U_\theta^\dagger, \quad (1)$$

for a unitary $U_\theta = \exp(-i\theta H)$ generated by the Hamiltonian⁶ $H = H^\dagger$. For example, in phase (and frequency) estimation, one considers a local Hamiltonian for N qubits, i.e.,

$$H = \sum_{i=1}^N H_i \quad (2)$$

and H_i acts nontrivially only on the i th qubit. Typically, one has

$$H_i \equiv \frac{1}{2} Z \quad \forall i, \quad (3)$$

where Z is the usual Pauli operator, but other local Hamiltonians can be brought to this form by local unitaries. After the encoding, a measurement of the probe state $\rho(\theta)$ is performed, which can be represented by a positive-operator valued measure (POVM), i.e., a set $\{E_m\}$ of positive semi-definite operators $E_m \geq 0$ satisfying $\sum_m E_m = \mathbb{1}$, where $\mathbb{1}$ is the identity operator. For an introduction to POVM measurements see, e.g., [34, p 90] or [35].

From the measurement outcomes, labelled m , an estimate of the parameter in question can be obtained. The precise nature of the estimator depends on the type of estimation scenario, distinguishing, for example, between local and Bayesian estimation mentioned previously. All these scenarios have in common that the precision of the estimation (as quantified by some figure of merit, e.g., the mean-square error (MSE)) improves with the number N of probe systems. For classical strategies based on product states, this increase is at most linear in N , which is referred to as the standard quantum limit (or shot noise scaling). However, using approaches based on the optimal quantum mechanical probes the improvement in this figure of merit can be quadratic in N , i.e., achieving (optimal) Heisenberg scaling. For reviews of parameter estimation techniques and quantum metrology we direct the reader to, e.g., [1, 3, 4] or the [appendix](#).

In local phase (and frequency) estimation one typically considers many repetitions of the same measurement that provide an estimate, whose variance one is interested in minimising using the available resources. In this scenario, the optimal N -qubit probe state is a GHZ state

$$|\psi_{\text{GHZ}}\rangle = \frac{1}{\sqrt{2}} (|0\rangle^{\otimes N} + |1\rangle^{\otimes N}), \quad (4)$$

and the accompanying optimal measurements are local X measurements. This can be determined via the QFI, the relevant figure of merit for local estimation, as we explain in more detail in [appendix A.1](#). In Bayesian parameter estimation (see, e.g., [4, 36] or [appendix A.2](#)), the situation is somewhat different. Here one quantifies the initial knowledge (or belief) about the parameter by a prior probability distribution that is updated after each single measurement. In this case, a figure of merit is the average variance of the updated distribution. In the Bayesian estimation scenario, the optimal probes and measurements depend on the shape of the prior and the cost function used. For instance, for phase estimation with flat priors (i.e., no prior knowledge), the optimal probe state achieving Heisenberg scaling is given by

⁶ We work in units where $\hbar = 1$. In addition, we adopt the usual convention of Hamiltonian estimation where the eigenvalues of H (and hence θ) are taken to be dimensionless. For example, for frequency estimation one then has $\theta = \omega t$, where the time t is assumed to be known precisely.

$$|\psi_{\text{opt}}\rangle = \sum_{n=0}^N \psi_n |n\rangle, \quad (5)$$

where $|n\rangle$ are eigenstates of H corresponding to its $N + 1$ different eigenvalues, and the coefficients ψ_n have a sinusoidal profile (see, e.g., [19]), i.e.,

$$\psi_n = \sqrt{\frac{2}{N+2}} \sin\left(\frac{(n+1)\pi}{N+2}\right). \quad (6)$$

Although different from the optimal measurement, we find that for the state in equation (5) a projective measurement in the basis obtained via the quantum Fourier transform (QFT) of the basis $\{|n\rangle\}$ allows for Heisenberg scaling for Bayesian phase and frequency estimation with Gaussian priors of varying widths, as we discuss in section 3, as well as in appendices A.4 and A.5.

The crucial observation required to extend the applicability of this approach to arbitrary priors (and cost functions) lies in noticing that in N -qubit phase (and frequency) estimation scenarios of any kind, H only has $N + 1$ different eigenvalues. For each of these values, only one representative eigenstate needs to be selected. Moreover, within the subspaces corresponding to fixed eigenvalues one may choose those eigenstates that can be prepared most efficiently. Instead of the typical Dicke states that are symmetric with respect to the exchange of the qubits, we therefore employ eigenstates corresponding to a unary encoding of n , i.e.,

$$|n\rangle_{\text{un}} = |1\rangle^{\otimes n} |0\rangle^{\otimes N-n}. \quad (7)$$

All optimal probe states can hence be chosen to be of the form of equation (5) with $|n\rangle \equiv |n\rangle_{\text{un}}$ for some choice of the coefficients ψ_n . Most importantly, all of these probe states have support in an $(N + 1)$ -dimensional subspace of the 2^N -dimensional overall Hilbert space.

Therefore, the problem of optimal state preparation and measurements for N qubits can be translated to that of $\lambda := \lceil \log(N + 1) \rceil$ qubits (where the logarithm is understood to be base 2), provided that one can efficiently and coherently convert the unary encoding $|n\rangle_{\text{un}}$ to a binary encoding in λ qubits. More precisely, one can initially prepare a state of λ qubits and convert it (efficiently) to the desired N -qubit state for sensing (using at least $N - \lambda$ auxiliary qubits). After the parameter has been encoded, one performs the reverse procedure before carrying out the final measurement on λ qubits. In section 5 we present a quantum circuit of size $O(N \log^2 N)$ (and its MBQC representation) achieving exactly such a unary-to-binary compression. On the logarithmically small space of these λ qubits the probe state preparation and measurement can then be carried out even with exponential overhead in λ while maintaining Heisenberg scaling.

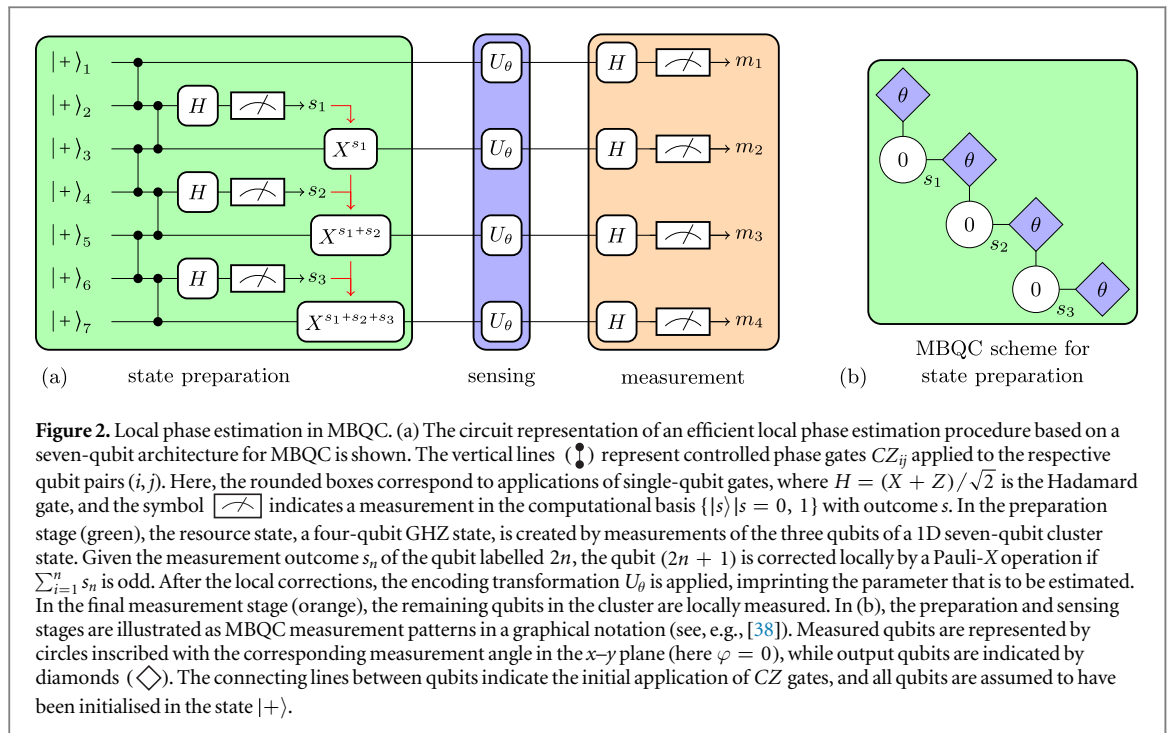
2.2. Parameter estimation in MBQC architectures

The premise for taking advantage of the quadratic scaling gap in resources (here, the number of qubits) between the quantum strategy described in the previous section and the best classical strategy is that the required probe states and measurements can be implemented efficiently. Here, we will take efficiency to mean that the overhead in the number of qubits used for the execution of the quantum strategy, including preparation and measurement, must grow less strongly than N^2 . To illustrate this requirement, consider a situation where an array of qubits is provided and one is given the task of using the array most efficiently for the estimation of a parameter. For instance, an array of spins (which may otherwise be used for quantum computation or quantum simulation) could be exposed to a magnetic field with fixed direction but unknown strength for this purpose. If one has the ability to prepare arbitrary quantum states of these (spin) qubits, then one may initialise a GHZ state for local phase estimation, or the corresponding optimal state for Bayesian phase estimation (or any other estimation problem for that matter). However, as we have seen in the previous section, states and measurements that offer advantages for different metrological problems are in general quite distinct, and the conversion from one to the other may involve arbitrarily long sequences of entangling operations. The preparation and measurement hence comes at a cost that we wish to quantify.

An approach that allows for preparing arbitrary quantum states and performing any measurements on them, while naturally including a resource count for these tasks is MBQC. In this paradigm, introduced in [30], an array of qubits is initialised in a particular (entangled) quantum state, typically a so-called cluster state [22]. A cluster state is a type of graph state, i.e., it can be represented by a graph (a set of vertices v_i and edges e_{ij} connecting the vertices). Each vertex represents a qubit initialised in the state $|+\rangle$, and controlled phase gates CZ, given by

$$CZ_{ij} = |0\rangle\langle 0|_i \otimes \mathbb{1}_j + |1\rangle\langle 1|_i \otimes Z_j = CZ_{ji}, \quad (8)$$

are applied to each pair of qubits connected by an edge. For simplicity, we will here only consider 2D cluster states where the underlying graph is a regular, rectangular lattice, but in principle, also other graph states [37] could be considered for our purposes. By applying only single-qubit gates and carrying out local measurements on a subset of all qubits in a 2D cluster, arbitrary unitary operations can be implemented on the remaining qubits



[31]. Performing a unitary transformation in the circuit model of quantum computation hence translates to a sequence of measurement angles for single-qubit measurements in the cluster. For a more detailed introduction to MBQC see [38, 39], or appendix A.6.

In other words, a number of the initial qubits can be sacrificed to obtain a probe state of fewer qubits, which is more suitable for a given metrological task at hand. Note that using the unmodified cluster state as a probe state itself does not provide a scaling advantage with respect to classical strategies, i.e., its QFI is $O(N)$. Similarly, additional qubits can be used to implement arbitrary measurements by performing appropriate unitaries followed by computational basis measurements. Here, one needs to ensure that only the part of the cluster used to prepare the probe state is subjected to the transformation encoding the parameter. This can be achieved, e.g., by appropriately timed Pauli- X operations on the qubits used for the measurement at the middle and at the end of the interaction period. For spins this corresponds to the general practice of refocusing of the magnetisation, i.e., a spin echo.

Crucially, the overall number of qubits required for the preparation and measurement of this N -qubit probe state must grow less than quadratically with N to maintain a potential metrological scaling advantage. This is possible, for instance, for local phase estimation, where the optimal measurement strategy can be carried out with $2N - 1$ qubits in a 1D cluster state as shown in figure 2. As we will show in the following, such efficient constructions also exist for Bayesian phase (and frequency) estimation problems. In section 4, we demonstrate that all probe states (including the optimal ones) of the form of equation (5) can be efficiently prepared from a 2D cluster state using only local operations. In section 5 we then present the unary-to-binary compression requiring $O(N \log^2 N)$ qubits of the cluster to reduce the problem of implementing optimal measurements to the subspace of $\lambda := \lceil \log(N + 1) \rceil$ qubits. On this subspace, projective measurements in any basis can be carried out efficiently, provided that the unitary transformation relating it to a computational-basis measurement requires no more than $O(2^\lambda)$ (nearest neighbour) gates. This is the case, for instance, for the QFT measurement, which performs optimally for flat priors [19] and achieves Heisenberg scaling for Gaussian priors of varying widths as we will show next.

3. Quantum advantage in Bayesian estimation

We now briefly discuss the Bayesian phase estimation scenario, more details on which can be found in appendix A.2, and show that the combination of sine states and QFT measurements can achieve Heisenberg scaling. In Bayesian parameter estimation, the initial knowledge about the parameter is encoded in a prior probability distribution $p(\theta)$. When a measurement with POVM elements $\{E_m\}$ is performed on the parameter-encoded state $\rho(\theta)$, the conditional probability of obtaining the outcome labelled m is

$$p(m|\theta) = \text{tr}(E_m \rho(\theta)). \quad (9)$$

To obtain the unconditional probability for the same outcome, these values are weighed according to one's prior belief, i.e.,

$$p(m) = \int d\theta p(\theta) p(m|\theta). \quad (10)$$

The information obtained in a measurement with outcome m is then used to update this belief via Bayes' law, obtaining the posterior distribution $p(\theta|m)$ given by

$$p(\theta|m) = \frac{p(m|\theta)p(\theta)}{p(m)}. \quad (11)$$

In turn, the posterior distribution provides an estimate $\hat{\theta}(m)$ for the parameter via

$$\hat{\theta}(m) = \int d\theta p(\theta|m)\theta. \quad (12)$$

As a figure of merit for this estimation procedure one then quantifies the width of the posterior by a suitable measure $V_{\text{post}}^{(m)}$ and averages over all possible outcomes, such that

$$\bar{V}_{\text{post}} = \sum_m p(m) V_{\text{post}}^{(m)}. \quad (13)$$

For instance, when the parameter in question has support over all of \mathbb{R} (e.g., for frequency estimation, see appendix A.5), one may use the MSE

$$V_{\text{post}}^{(m)} = V[p(\theta|m)] = \int d\theta p(\theta|m)(\theta - \hat{\theta}(m))^2. \quad (14)$$

Here, we want to focus on phase estimation, i.e., the case where the parameter has support on the interval $[-\pi, \pi] \subset \mathbb{R}$. When the prior is appropriately narrow, one may still use the MSE, which allows the use of some simple techniques (e.g., a Bayesian version of the Cramér–Rao (CR) inequality, see the appendix A.2.2 and [25]) for the comparison with classical strategies. Nonetheless, wrapped distributions and covariant measures of their width are in general more suitable for phase estimation. As an example, one can consider the wrapped Gaussian distribution of the form

$$p(\theta) = \frac{1}{\sqrt{2\pi}\sigma} \sum_{q=-\infty}^{\infty} e^{-\frac{(\theta-\theta_0+2\pi q)^2}{2\sigma^2}}, \quad (15)$$

where $q \in \mathbb{Z}$, and the mean angle is

$$\arg(\langle e^{i\theta} \rangle_{p(\theta)}) = \int_{-\pi}^{\pi} d\theta p(\theta) = \theta_0. \quad (16)$$

The non-negative parameter σ can be identified with the circular standard deviation

$$S = \sqrt{\ln(1/|\langle e^{i\theta} \rangle_{p(\theta)}|^2)} = \sigma, \quad (17)$$

corresponding to the width of the underlying Gaussian distribution. However, for our purposes, it is more useful to quantify the width of this wrapped distribution by the Holevo phase variance [40] V_ϕ , given by

$$V_\phi = |\langle e^{i\theta} \rangle_{p(\theta)}|^{-2} - 1 = e^{\sigma^2} - 1. \quad (18)$$

Likewise, we will quantify the width of the posterior by

$$V_{\phi,\text{post}}^{(m)} = |\langle e^{i\theta} \rangle_{p(\theta|m)}|^{-2} - 1. \quad (19)$$

For the probe states of equation (5) with the sinusoidal profile of equation (6), and the QFT measurement represented by the basis $\{|e_k\rangle\}$, where

$$|e_k\rangle = \frac{1}{\sqrt{N+1}} \sum_{n=0}^N e^{in\frac{2\pi k}{N+1}} |n\rangle_{\text{un}}, \quad (20)$$

we then calculate the average phase variance $\bar{V}_{\phi,\text{post}} = \sum_m p(m) V_{\phi,\text{post}}^{(m)}$. The results for various values of σ and for up to 100 qubits are shown in figure 3. The numerical results indicate that for all widths of the priors the example quantum strategy exhibits Heisenberg scaling. In appendix A.4 we discuss the performance of this measurement strategy in more detail and give a comparison with the performance of classical strategies, which can be shown to exhibit shot noise scaling.

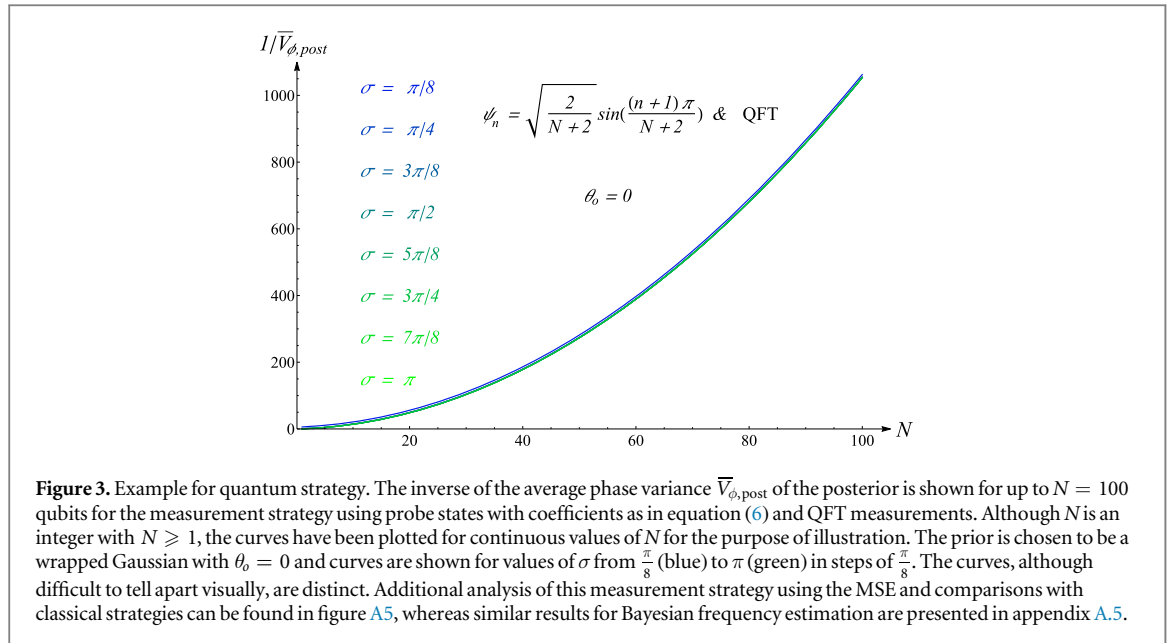


Figure 3. Example for quantum strategy. The inverse of the average phase variance $\bar{V}_{\phi, \text{post}}$ of the posterior is shown for up to $N = 100$ qubits for the measurement strategy using probe states with coefficients as in equation (6) and QFT measurements. Although N is an integer with $N \geq 1$, the curves have been plotted for continuous values of N for the purpose of illustration. The prior is chosen to be a wrapped Gaussian with $\theta_0 = 0$ and curves are shown for values of σ from $\frac{\pi}{8}$ (blue) to π (green) in steps of $\frac{\pi}{8}$. The curves, although difficult to tell apart visually, are distinct. Additional analysis of this measurement strategy using the MSE and comparisons with classical strategies can be found in figure A5, whereas similar results for Bayesian frequency estimation are presented in appendix A.5.

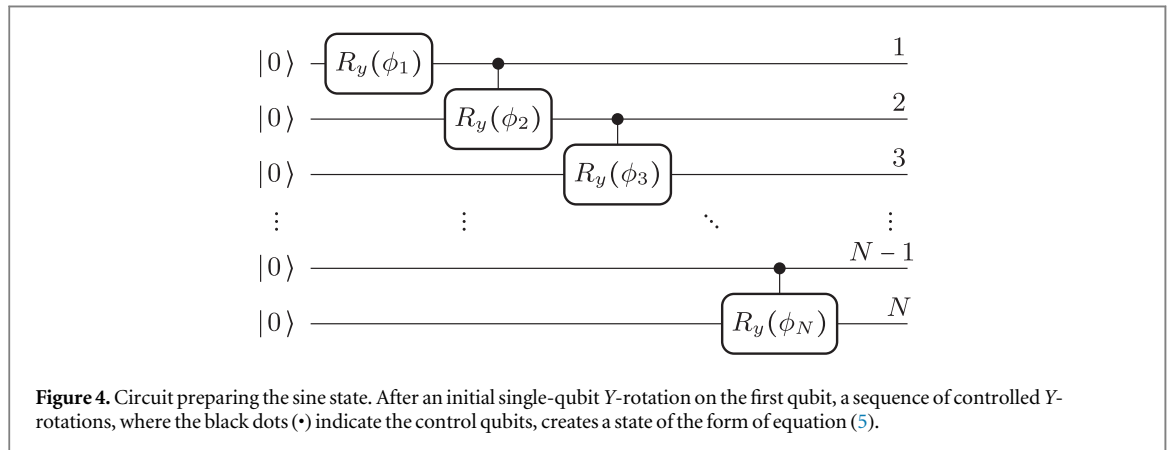


Figure 4. Circuit preparing the sine state. After an initial single-qubit Y -rotation on the first qubit, a sequence of controlled Y -rotations, where the black dots (\bullet) indicate the control qubits, creates a state of the form of equation (5).

4. Efficient preparation of probe states

In this section we present a method that allows for the efficient preparation of the probe state of equation (5), which immediately generalises to any state in the subspace of optimal probes spanned by $\{|n\rangle_{\text{un}}\}_{n=0, \dots, N}$. This method relies on the simple observation that in the bit-string $(u_1 u_2 u_3 \dots u_N)$ representing the state

$$|n\rangle_{\text{un}} = |1\rangle^{\otimes n} |0\rangle^{\otimes N-n} = |u_1\rangle |u_2\rangle \dots |u_N\rangle, \tag{21}$$

i.e., where $u_k \in \{0, 1\}$ and $n = \sum_k u_k$, the n entries $u_1, u_2, \dots, u_n = 1$ are always to the left of the entries $u_{n+1}, \dots, u_N = 0$. In other words, the k th qubit can only be in the state $|1\rangle$, if all of the $k - 1$ qubits before are also in the state $|1\rangle$.

Focussing on the sine state of equation (6) as an example, note that the coefficients are all real and positive. Initialising all qubits in the state $|0\rangle$, the circuit preparing the sine state must hence be a cascade of N (controlled) single-qubit Y -rotations $CR_y(\phi_i)$, whose angles $\{\phi_i\}_{i=1, \dots, N}$ determine the weights ψ_n , see figure 4. This becomes apparent when inspecting the single-qubit Pauli- Y rotations

$$R_y(\phi) = \exp\left(i\frac{\phi}{2}Y\right) = \begin{pmatrix} \cos\left(\frac{\phi}{2}\right) & \sin\left(\frac{\phi}{2}\right) \\ -\sin\left(\frac{\phi}{2}\right) & \cos\left(\frac{\phi}{2}\right) \end{pmatrix}. \tag{22}$$

The action of the circuit in figure 4 then transforms the k th qubit to the state $\cos\left(\frac{\phi_k}{2}\right)|0\rangle + \sin\left(\frac{\phi_k}{2}\right)|1\rangle$ if the $(k - 1)$ th qubit is in the state $|1\rangle$. All together, these N rotations are parametrised by angles $\phi_n \in [0, \pi/4)$, such that both the sine and the cosine in the above expression are non-negative. It is straightforward to verify that the output of the circuit is the state of equation (5) with amplitudes

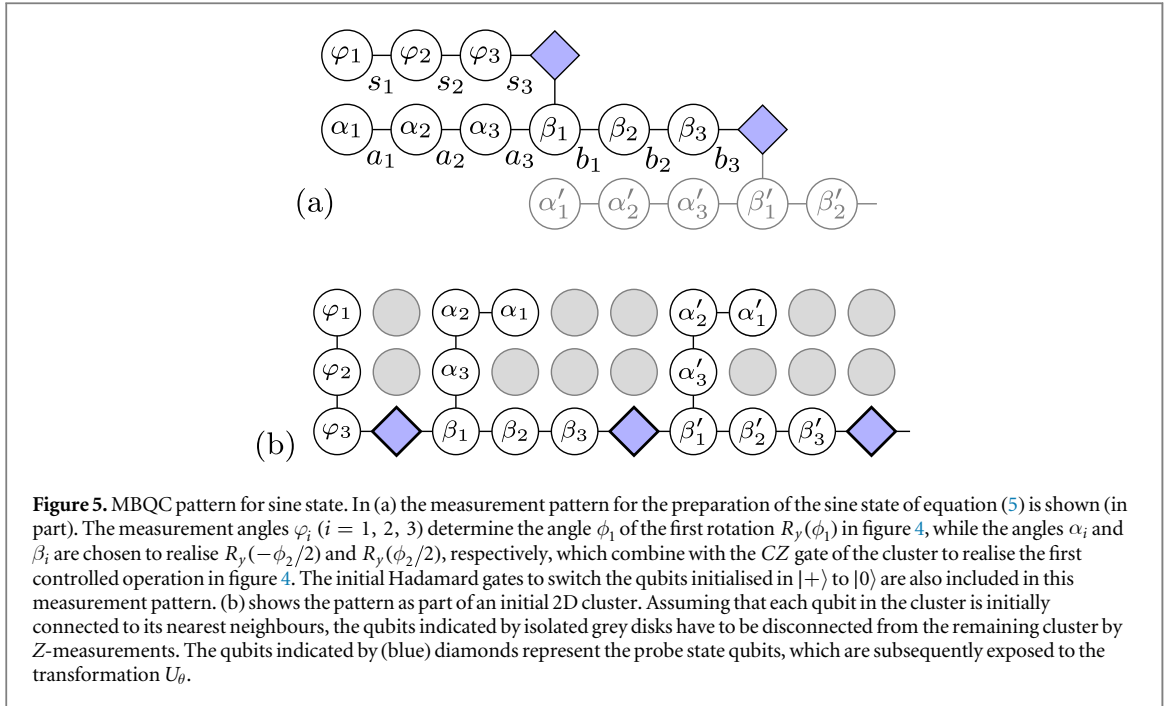


Figure 5. MBQC pattern for sine state. In (a) the measurement pattern for the preparation of the sine state of equation (5) is shown (in part). The measurement angles φ_i ($i = 1, 2, 3$) determine the angle ϕ_1 of the first rotation $R_y(\phi_1)$ in figure 4, while the angles α_i and β_i are chosen to realise $R_y(-\phi_2/2)$ and $R_y(\phi_2/2)$, respectively, which combine with the CZ gate of the cluster to realise the first controlled operation in figure 4. The initial Hadamard gates to switch the qubits initialised in $|+\rangle$ to $|0\rangle$ are also included in this measurement pattern. (b) shows the pattern as part of an initial 2D cluster. Assuming that each qubit in the cluster is initially connected to its nearest neighbours, the qubits indicated by isolated grey disks have to be disconnected from the remaining cluster by Z-measurements. The qubits indicated by (blue) diamonds represent the probe state qubits, which are subsequently exposed to the transformation U_θ .

$$\psi_n = \begin{cases} \cos\left(\frac{\phi_{n+1}}{2}\right) \prod_{k=1}^n \sin\left(\frac{\phi_k}{2}\right) & \forall n \in \{0, 1, \dots, N-1\} \\ \prod_{k=1}^N \sin\left(\frac{\phi_k}{2}\right) & \text{for } n = N \end{cases}. \quad (23)$$

Note that ψ_0 uniquely determines ϕ_1 and that each of the ψ_n depends only on $\{\phi_k\}_{k=1}^{n+1}$. This allows inverting equation (23) and expressing the angles ϕ_n as

$$\phi_n = \begin{cases} 2 \arccos(\psi_0) & \text{for } n = 1 \\ 2 \arccos\left(\frac{\psi_{n-1}}{\sqrt{1 - \sum_{k=0}^{n-2} \psi_k^2}}\right) & \text{for } n \in \{2, 3, \dots, N\} \end{cases}, \quad (24)$$

which allows reconstructing the rotation angles for any real, non-negative choice of $\{\psi_n\}$.

Having found the circuit shown in figure 4, the only difficulty is to arrange the required measurements such that the overall preparation procedure can be embedded efficiently in a rectangular 2D structure, which is shown in appendix A.6.2. We hence arrive at the MBQC measurement pattern depicted in figure 5, which generates the sine state of equation (5) with weights as in equation (6). It requires a square 2D cluster of (at most) $3 \times (4N - 2)$ qubits to prepare an N -qubit probe state. Crucially, the number of qubits in the cluster increases only linearly with the size of the probe. Moreover, any other probe state in the subspace spanned by the vectors $\{|n\rangle_{\text{un}}\}_{n=0, \dots, N}$ can be prepared with the same efficiency in a similar way by replacing the Pauli-Y rotations by other single-qubit unitaries.

Next, we will show in section 5 how a large class of useful measurements of the encoded probe states (including the QFT measurement) can be carried out efficiently.

5. Efficient unary-to-binary compression

Finally, we turn to the implementation of the measurements required to achieve Heisenberg scaling. In principle, the optimal measurement for a given prior and cost function may be an arbitrarily complicated measurement in an entangled basis of N -qubit states, for example, a projective measurement in the QFT basis (see, e.g., [34, chapter 5] or [42, 43]).

Fortunately, closer inspection reveals that we do not require arbitrary measurements on N but only on $\lambda := \lceil \log(N + 1) \rceil$ qubits, where the logarithm is understood to be to base 2. This is the case because all encoded information about the phase is stored within the $(N + 1)$ -dimensional subspace spanned by the vectors $\{|n\rangle_{\text{un}}\}_{n=0, \dots, N}$. All optimal measurements can hence be restricted to this subspace. To exploit this observation,

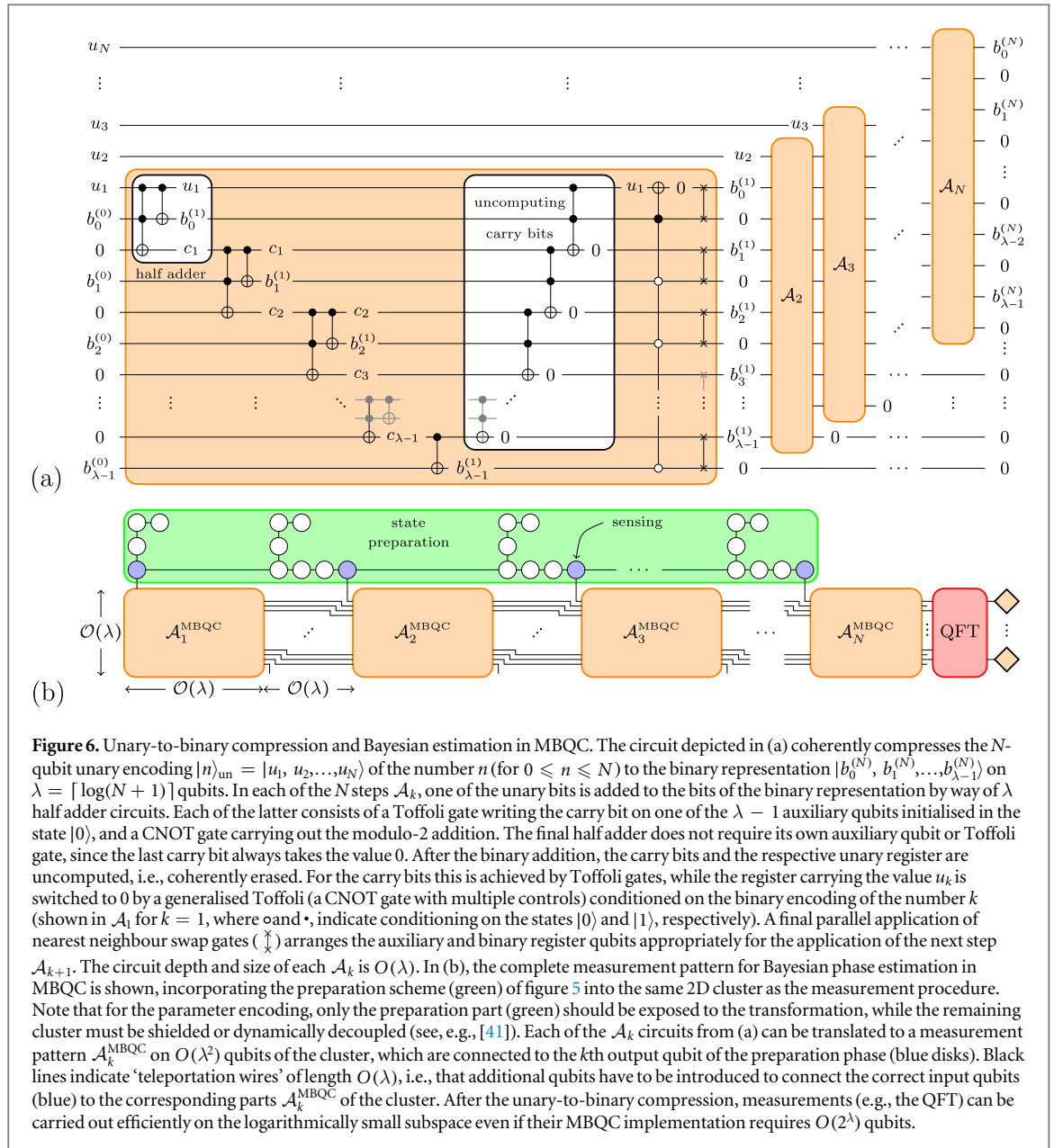


Figure 6. Unary-to-binary compression and Bayesian estimation in MBQC. The circuit depicted in (a) coherently compresses the N -qubit unary encoding $|n\rangle_{\text{un}} = |u_1, u_2, \dots, u_N\rangle$ of the number n (for $0 \leq n \leq N$) to the binary representation $|b_0^{(N)}, b_1^{(N)}, \dots, b_{\lambda-1}^{(N)}\rangle$ on $\lambda = \lceil \log(N+1) \rceil$ qubits. In each of the N steps \mathcal{A}_k , one of the unary bits is added to the bits of the binary representation by way of λ half adder circuits. Each of the latter consists of a Toffoli gate writing the carry bit on one of the $\lambda - 1$ auxiliary qubits initialised in the state $|0\rangle$, and a CNOT gate carrying out the modulo-2 addition. The final half adder does not require its own auxiliary qubit or Toffoli gate, since the last carry bit always takes the value 0. After the binary addition, the carry bits and the respective unary register are uncomputed, i.e., coherently erased. For the carry bits this is achieved by Toffoli gates, while the register carrying the value u_k is switched to 0 by a generalised Toffoli (a CNOT gate with multiple controls) conditioned on the binary encoding of the number k (shown in \mathcal{A}_1 for $k = 1$, where \circ and \bullet indicate conditioning on the states $|0\rangle$ and $|1\rangle$, respectively). A final parallel application of nearest neighbour swap gates (\times) arranges the auxiliary and binary register qubits appropriately for the application of the next step \mathcal{A}_{k+1} . The circuit depth and size of each \mathcal{A}_k is $O(\lambda)$. In (b), the complete measurement pattern for Bayesian phase estimation in MBQC is shown, incorporating the preparation scheme (green) of figure 5 into the same 2D cluster as the measurement procedure. Note that for the parameter encoding, only the preparation part (green) should be exposed to the transformation, while the remaining cluster must be shielded or dynamically decoupled (see, e.g., [41]). Each of the \mathcal{A}_k circuits from (a) can be translated to a measurement pattern $\mathcal{A}_k^{\text{MBQC}}$ on $O(\lambda^2)$ qubits of the cluster, which are connected to the k th output qubit of the preparation phase (blue disks). Black lines indicate ‘teleportation wires’ of length $O(\lambda)$, i.e., that additional qubits have to be introduced to connect the correct input qubits (blue) to the corresponding parts $\mathcal{A}_k^{\text{MBQC}}$ of the cluster. After the unary-to-binary compression, measurements (e.g., the QFT) can be carried out efficiently on the logarithmically small subspace even if their MBQC implementation requires $O(2^\lambda)$ qubits.

we will now present an efficient algorithm that coherently compresses the information encoded in the probe state on the 2^N -dimensional Hilbert space of N qubits to the exponentially smaller space of $\lambda := \lceil \log(N+1) \rceil$ qubits.

The principle of operation of this N -step compression algorithm, shown in figure 6 (a), is to switch from the unary encoding of the number n in the state

$$|n\rangle_{\text{un}} = |1\rangle^{\otimes n} |0\rangle^{\otimes N-n} = |u_1, u_2, \dots, u_N\rangle \quad (25)$$

to a binary encoding of the same number via a unitary transformation and extend the result to superpositions of different states $|n\rangle_{\text{un}}$ by linearity. The unary-to-binary conversion is achieved by successive binary addition of each bit in the string (u_1, u_2, \dots, u_N) to the bit string of an auxiliary register of length λ initially representing the number 0. The corresponding qubits are initialised in the state

$$|b_{\lambda-1}^{(0)}, b_{\lambda-2}^{(0)}, \dots, b_1^{(0)}, b_0^{(0)}\rangle = |0\rangle^{\otimes \lambda}. \quad (26)$$

In the k th step of the procedure, the bit u_k is added to the binary representation $(b_{\lambda-1}^{(k-1)}, \dots, b_0^{(k-1)})$ of the number $n^{(k-1)} = \sum_{i=0}^{\lambda-1} b_i^{(k-1)} 2^i$, where $n^{(k)} = k$ for $0 \leq k \leq n$ and $n^{(k)} = n$ for $n < k \leq N$.

The binary addition of u_k to the least significant digit $b_0^{(k-1)}$ of $n^{(k-1)}$ is performed by a half adder circuit, see figure 6(a). It, in turn consists of a CNOT gate for the modulo-2 addition, producing the original value u_k and the new binary digit $b_0^{(k)} = u_k \oplus b_0^{(k-1)}$, where \oplus denotes addition modulo 2. The CNOT is preceded by a Toffoli gate whose target is an additional auxiliary qubit which stores the carry bit (see, e.g., [44–46] for quantum

arithmetic operations). This carry bit is then added to the next binary digit $b_1^{(k-1)}$ by another half adder. The procedure carries on until reaching the final binary digit $b_{\lambda-1}^{(k-1)}$, where the half adder can be replaced by a simple CNOT gate, since the register size was chosen such that the final carry bit is always 0.

Subsequently, the qubits corresponding to the carry bits and u_k have to be disentangled from the qubits carrying the binary encoding. For the carry bits, this is achieved by another cascade of Toffoli gates (see figure 6(a)), since the carry bit can only have the value 1, if both of the previously added bits have the value 1 as well. To coherently erase u_k , note that the binary string $(b_{\lambda-1}^{(k)}, \dots, b_0^{(k)})$ encodes the number k only if $u_k = 1$. We can hence flip the corresponding qubit conditioned on the binary encoding of k using a generalised Toffoli gate. Using the already existing ancillas (which have previously been returned to the state $|0\rangle$), this multi-controlled CNOT gate can be realised in a standard construction using $\lambda - 1$ NN SWAP gates, preceding and following an array of $2(\lambda - 1)$ Toffolis on three adjacent qubits along with a single CNOT [34, p 184]. Conditioning on states $|0\rangle$ rather than $|1\rangle$ requires at most 2λ additional single-qubit X gates. Having disentangled all other qubits from the λ qubits storing the binary encoding, we perform another λ NN SWAPS in anticipation of inputting the next unary digit u_{k+1} .

Taking into account that each Toffoli or NN SWAP gate can be realised with a constant overhead in NN CNOT and single-qubit gates, we find that the circuit for \mathcal{A}_k requires at most $O(\lambda)$ NN CNOT and single-qubit gates. The entire unary-to-binary compression algorithm consists of N such elements, resulting in a circuit size of $O(N \log N)$ on an input of length $O(\log N)$, which can hence be realised with at most $O(N \log^2 N)$ qubits in MBQC, see figure 6(b).

On the logarithmically sized (in N) output, any measurement can then be performed efficiently as long as the corresponding unitary on $\lambda := \lceil \log(N + 1) \rceil$ qubits requires no more than 2^λ NN gates. While this does not cover all possible unitaries (e.g., the construction discussed in [34 p193] requires $O(\lambda^2 2^\lambda)$ two-qubit and single-qubit gates), some particularly useful unitaries may be much less costly. For instance, an implementation of the QFT on a λ -qubit linear NN architecture presented in [43] has circuit size $O(\lambda \log \lambda)$ and depth $O(\lambda)$, meaning an overhead of only $O(\lambda^2)$ qubits (depth times input length) in a measurement-based setting.

6. Discussion

In summary, we have shown that 2D architectures for MBQC provide flexible resources for quantum-enhanced metrology tasks. That is, an initial array of qubits prepared in a 2D cluster state and local operations are used to achieve Heisenberg scaling for phase and frequency estimation in both the local (frequentist) and the Bayesian approach to parameter estimation. In the Bayesian scenario, the preparation procedure presented can be applied to execute strategies with optimal states for arbitrary priors and cost functions. This flexibility allows outperforming other approaches where a fixed probe state (e.g., an array of differently sized GHZ states) is used for different task without adaption to the specific problem at hand. The efficient compression algorithm further allows to perform measurements with up to exponential circuit sizes. This includes the QFT measurement that is optimal for flat priors, provides Heisenberg scaling for Gaussian priors of varying widths, and is expected to perform similarly well also for other priors under certain regularity conditions.

In principle, our results can be generalised also to scenarios beyond phase and frequency estimation. For all local Hamiltonians that are not proportional to Z , appropriate local corrections can be applied on the sensing qubits before and after the encoding such that the overall transformation commutes with the controlled phase gates used to create the cluster. For instance, when $H = \frac{1}{2}X$, Hadamard gates before and after U_θ produce an encoding transformation that commutes with CZ and can hence be applied after the entire cluster for sensing and measurements has been prepared. Moreover, when the corresponding states and measurements giving Heisenberg scaling are known, a similar method can also be employed for nonlocal interaction Hamiltonians, provided that they are proportional to a product of Pauli operators, or linear combinations of products of only one type of Pauli operators. For example, for parameter estimation with Ising-type couplings of the form $H = \sum_{i,j} c_{ij} X_i \otimes X_j$, GHZ states and local measurements achieve Heisenberg scaling [24], which can hence be efficiently implemented in our scheme. Nonetheless, many interesting questions regarding the applicability to general dynamics and scaling beyond the Heisenberg limit [47–51] remain.

Our results are of practical significance since they suggest that a single platform, 2D cluster states, can be flexible enough for a plethora of precision-enhanced parameter estimation tasks. In addition, this platform could in principle also be part of an integrated device, where a parameter estimation strategy is used to learn about, e.g., stray fields or the particular form of noise processes. For this purpose, part of the 2D cluster state can be used for sensing, while the remaining qubits are used to perform MBQC. The gathered information from the parameter estimation can then be used to improve the performance of the computation: by learning stray fields, one can compensate for systematic errors. By learning the particular shape of a noise process, one can adapt to an optimised error correction code, thereby reducing the overhead for fault-tolerant implementations.

At the same time, this connection between computational and metrological resources provides interesting insights. The advantage in metrology is provided by the entanglement of the cluster state, i.e., the CZ gates applied to neighbouring pairs of qubits, which ensures the improved performance with respect to an array of unentangled, individual qubits. At the same time, it is known that metrological advantages can, but need not arise solely from entanglement [52–54]. For example, nonclassicality in terms of squeezing can lead to Heisenberg scaling in precision [55, 56] without any entanglement when the average energy is considered as the resource. This work hence also contributes to the discussion of the required physical resources for parameter estimation [57], and the relationship between computational power and metrology [58].

Finally, open questions remain regarding the role of noise [59–61], especially in connection with adaptive approaches to computation and error-correction involving metrology [62–67]. Although noise is known to be problematic in the limit of infinitely many qubits since it is known to restrict to a linear scaling of precision, i.e., $\mathcal{I} \leq \kappa N$ for some constant κ , the approach presented here holds the promise of significantly outperforming classical strategies for finite system sizes. Indeed, this follows from the observation that the constant κ strongly depends on the strength and type of the noise [60, 64] and can be arbitrarily large if the noise is weak enough. Meanwhile, the overhead needed for preparation and measurement of the optimal state does not depend on the noise, leaving room for an arbitrarily large advantage of our scheme over classical strategies for any fixed N . In addition, techniques that deal with errors and maintain a metrological advantage are known (see, e.g., [62–64]) and may be applicable here. We leave such extensions for future work, along with the explicit determination of optimal [68, 69] and ‘pretty good’ states [70] for specific metrological tasks in our framework, where recent algorithmic approaches [71] may prove to be useful.

Acknowledgments

We are grateful to Jan Kołodyński and Markus Tiersch for valuable discussions and comments. This work was supported by the Austrian Science Fund (FWF) through Grants No. SFB FoQuS F4012 and No. P28000-N27 and the START project Y879-N27, as well as DK ALM:W1259, the Templeton World Charity Foundation Grant No. TWCF0078/AB46, the Swiss National Science Foundation Grant No. P300P2_167749, the Spanish MINECO through Grant No. FIS2013-40627-P, the Generalitat de Catalunya CIRIT contract 2014-SGR966 and by TherMiQ (Grant Agreement 618074).

Appendix

A.1. Local parameter estimation

In this appendix, we give a detailed description of the local parameter estimation scenario and show how Heisenberg scaling can be achieved using a GHZ state and local measurements.

A.1.1. The local estimation scenario. We consider a typical parameter estimation scenario, where θ , the quantity of interest, is encoded in a density operator $\rho(\theta)$ by a dynamical (unitary) transformation $U_\theta = e^{-i\theta H}$, i.e.,

$$\rho(\theta) = U_\theta \rho(0) U_\theta^\dagger. \quad (\text{A.1})$$

We then perform a measurement with POVM elements $\{E_m\}$ which yields an outcome m . The (conditional) probability of obtaining the measurement outcome m (given that the parameter has the value θ) is then

$$p(m|\theta) = \text{tr}(E_m \rho(\theta)). \quad (\text{A.2})$$

To each measurement outcome m , an estimator $\hat{\theta}(m)$ assigns a corresponding estimate for the value of θ . The estimator is called *unbiased* if it assigns the value θ on average, that is, if the expected value of the estimator satisfies

$$\langle \hat{\theta}(m) \rangle = \sum_m p(m|\theta) \hat{\theta}(m) = \theta. \quad (\text{A.3})$$

This requirement ensures the *accuracy* of the measurement procedure, but not its *precision*, which is determined by the variance $V[\hat{\theta}(m)]$ of the estimator. We use the MSE given by

$$V[\hat{\theta}(m)] = \sum_m p(m|\theta) (\hat{\theta}(m) - \theta)^2, \quad (\text{A.4})$$

and $\sigma = \sqrt{V[\hat{\theta}(m)]}$ is the associated standard deviation. Unfortunately, it is often the case that a given estimator offers high precision only within a small range of the parameter θ , but not globally, as we shall discuss for a simple example in appendix A.1.4. Such estimators are hence useful *locally*, i.e., for estimating small fluctuations of the parameter around some known value. In such local estimation scenarios, accuracy is

guaranteed even when unbiasedness as specified in equation (A.3) is required to hold only in the vicinity of this value.

To increase the precision, the procedure consisting of preparation, encoding, and measurement may be repeated a number of times, say ν , providing estimates $\theta^{(i)}$ ($i = 1, \dots, \nu$), from which the mean value

$$\bar{\theta}_\nu = \frac{1}{\nu} \sum_{i=1}^{\nu} \theta^{(i)} \quad (\text{A.5})$$

and the associated MSE

$$V_\nu[\theta^{(i)}] = \frac{1}{\nu} \sum_{i=1}^{\nu} (\theta^{(i)} - \bar{\theta}_\nu)^2 \quad (\text{A.6})$$

can be calculated. As ν increases, the mean and variance computed from the measurement data converge to the expected value $\hat{\theta}(m)$ of the estimates and the expected value of the corresponding variance, $V[\hat{\theta}(m)]$, respectively. Trusting that the results of the individual runs are independent and identically distributed, the variance of the distribution of mean values with ν samples decreases linearly with ν . The overall expected precision associated to the result $\bar{\theta}_\nu$ is hence quantified by the standard error of the mean, given by $\sigma_\nu = \sigma/\sqrt{\nu} = \sqrt{V[\hat{\theta}(m)]/\nu}$. In other words, the precision increases with the number of runs, but the options for choosing a probe state, measurement, and estimator still leave room for improvement.

It is here that measurement strategies using genuine quantum features such as entanglement and nonclassicality can provide advantages with respect to classical strategies. To determine the potential gain and to allow comparisons with the best classical protocol it is useful to eliminate the choice of estimator, and consider the important CR bound, before discussing an example estimation scenario in appendix A.1.4.

A.1.2. The CR bound. For any unbiased estimator the variance $V[\hat{\theta}(m)]$ can be shown (see, e.g., [72–74] or appendix A.1.3) to satisfy the CR inequality

$$V[\hat{\theta}(m)] \geq \frac{1}{I(\rho(\theta))}, \quad (\text{A.7})$$

where $I(\rho(\theta))$ is the *Fisher information* (FI) given by

$$I(\rho(\theta)) = \sum_m p(m|\theta) \left(\frac{\partial}{\partial \theta} \log p(m|\theta) \right)^2 = \sum_m \frac{(\text{tr}[E_m \dot{\rho}(\theta)])^2}{\text{tr}(E_m \rho(\theta))}. \quad (\text{A.8})$$

Here it is noteworthy that, on the one hand, the FI does not depend on the choice of the estimator (as long as it is unbiased), and one can hence determine a lower bound for the variance based solely on the initial state and the chosen measurement. On the other hand, the FI typically depends on the value of the parameter and an unbiased estimator for which the CR inequality globally becomes an equality may not exist for all values. However, estimators can be found for which the bound is tight locally, and globally in the asymptotic limit of $\nu \rightarrow \infty$, see, e.g., [75].

One may then further ask, what the optimal measurement strategy is for a given probe state and parameter encoding. The maximisation of the FI over all possible POVMs then yields (see, e.g., [76]) the QFI $\mathcal{I}(\rho(\theta))$, given by

$$\mathcal{I}(\rho(\theta)) = 2 \text{tr}(\hat{S}_\theta \dot{\rho}(\theta)), \quad (\text{A.9})$$

where the operator $\hat{S}_\theta \equiv \hat{S}[\rho(\theta)]$, called the *symmetric logarithmic derivative* (SLD) is implicitly given by the relation

$$\hat{S}_\theta \rho(\theta) + \rho(\theta) \hat{S}_\theta = \dot{\rho}(\theta), \quad (\text{A.10})$$

and where the dot indicates the partial derivative with respect to θ , i.e., $\dot{\rho} = \frac{\partial}{\partial \theta} \rho$. The corresponding quantum CR bound is hence simply $V[\hat{\theta}(m)] \geq 1/\mathcal{I}(\rho(\theta))$. The optimal measurement for which the FI and the QFI coincide is a projective measurement in the eigenbasis of the SLD \hat{S}_θ [76].

For unitary encodings that we consider here, see equation (A.1), the QFI is independent of the value of the parameter, $\mathcal{I}(\rho(\theta)) = \mathcal{I}(\rho(\theta')) \forall \theta, \theta'$. To see this, simply note that in such a case $\rho(\theta') = U_{\theta'-\theta} \rho(\theta) U_{\theta'-\theta}^\dagger$ and

$$\dot{\rho}(\theta) = i[\rho(\theta), H]. \quad (\text{A.11})$$

Therefore the derivative appearing in the QFI in equation (A.9) is just $\dot{\rho}(\theta') = U_{\theta'-\theta} \dot{\rho}(\theta) U_{\theta'-\theta}^\dagger$. Using equation (A.10) one then finds that the SLDs are related in the same way, i.e., $\hat{S}_{\theta'} = U_{\theta'-\theta} \hat{S}_\theta U_{\theta'-\theta}^\dagger$. Cyclically permuting the unitary operators in the trace then gives the result, $\mathcal{I}(\rho(\theta)) = \mathcal{I}(\rho(\theta'))$.

If we additionally restrict to pure probe states $|\psi\rangle$ as before, the QFI takes the simple form (see, e.g., [4])

$$\mathcal{I}(|\psi\rangle) = 4(\langle \dot{\psi}_\theta | \dot{\psi}_\theta \rangle - |\langle \dot{\psi}_\theta | \psi_\theta \rangle|^2), \quad (\text{A.12})$$

where $|\psi_\theta\rangle = U_\theta|\psi\rangle$ is the encoded state and the dot indicates a partial derivative with respect to θ . Since $U_\theta = e^{-i\theta H}$, a simple computation then reveals that the QFI for such scenarios is proportional to the variance of the Hamiltonian generating the dynamics, i.e.,

$$\mathcal{I}(|\psi\rangle) = 4 (\langle H^2 \rangle_{\psi_\theta} - \langle H \rangle_{\psi_\theta}^2) = 4(\Delta H)^2 \quad (\text{A.13})$$

and the SLD coincides with $\dot{\rho}(\theta)$. The QFI is hence maximal for pure states that maximise the variance of H , see, e.g., [4, 77].

Let us now consider an estimation scenario where a probe state of N qubits is subject to a local transformation, i.e., where the Hamiltonian is of the form $H = \sum_{i=1, \dots, N} H_i$ and H_i acts nontrivially only on the i th qubit. For simplicity, we assume that each qubit undergoes the same local transformation, $H_i = H_j \equiv H_1 \forall i, j$, and that the local Hamiltonian has eigenvalues⁷ $\pm \frac{1}{2}$ with the corresponding eigenstates denoted by $|0\rangle$ and $|1\rangle$. We may further align our reference frame such that $H_1 = S_z = \frac{1}{2}Z$, where S_i is the spin- $\frac{1}{2}$ angular momentum operator for direction $i = x, y, z$ and X, Y, Z denote the usual Pauli operators. A comment on estimation scenarios for other Hamiltonians can be found in section 6 of the main text, but here we are restricting our discussion to phase estimation scenarios where $U_\theta = \bigotimes_{n=1}^N U_\theta^{(n)}$, such that $U_\theta^{(n)} = \exp(-i\theta S_z^{(n)})$ acts only on the n th qubit. For ease of notation, we will drop the superscript (n) in the following when referring to single-qubit operations and there is no risk of confusion.

If the probe state is classical, i.e., a product state of the form $|\psi_1\rangle \otimes |\psi_2\rangle \otimes \dots \otimes |\psi_N\rangle$, then the QFI becomes maximal when the local single-qubit probe states are all chosen to be $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$, maximising the variance of $H_1 = S_z$. From equation (A.13) it then follows immediately that the largest possible value of the QFI for a classical probe of N qubits is

$$\mathcal{I}(|+\rangle^{\otimes N}) = 4N (\langle H_1^2 \rangle - \langle H_1 \rangle^2) = N. \quad (\text{A.14})$$

The corresponding SLD is easily found to be $\hat{S}_\theta = (\cos \theta S_y - \sin \theta S_x)^{\otimes N}$, i.e., the optimal measurement is realised by single-qubit projective measurements in the basis $U_{\theta+\pi/2}|\pm\rangle$, where $|\pm\rangle$ are the eigenstates of $S_x = \frac{1}{2}X$. It hence becomes obvious that classical measurement strategies can (at most) decrease the variance linearly with the number of qubits. This scaling behaviour is referred to as the *standard quantum limit*. As we shall discuss next, a different scaling behaviour can be achieved for quantum probes.

A.1.3. A proof of the CR bound. In this section we present a proof of the CR bound of equation (A.7) following [73, 74]. For an unbiased estimator $\hat{\theta}(m)$, we can write the unbiasedness condition of equation (A.3) as

$$\sum_m p(m|\theta) (\hat{\theta}(m) - \theta) = 0, \quad (\text{A.15})$$

where we have used that the conditional probability is normalised, i.e., $\sum_m p(m|\theta) = 1$. Differentiating the condition of equation (A.15) with respect to the parameter we have

$$\sum_m \frac{\partial p(m|\theta)}{\partial \theta} (\hat{\theta}(m) - \theta) - \sum_m p(m|\theta) = 0, \quad (\text{A.16})$$

which we can rewrite as

$$\sum_m (\hat{\theta}(m) - \theta) p(m|\theta) \frac{\partial}{\partial \theta} \log p(m|\theta) = 1. \quad (\text{A.17})$$

Then, we define the quantities $x := \sqrt{p(m|\theta)} (\hat{\theta}(m) - \theta)$ and

$$y := \sqrt{p(m|\theta)} \frac{\partial}{\partial \theta} \log p(m|\theta), \quad (\text{A.18})$$

and use the Cauchy–Schwarz inequality

$$\left| \sum x y \right|^2 \leq \sum |x|^2 \sum |y|^2 \quad (\text{A.19})$$

to arrive at

$$\sum_m p(m|\theta) (\hat{\theta}(m) - \theta)^2 \times \sum_n p(n|\theta) \left(\frac{\partial}{\partial \theta} \log p(n|\theta) \right)^2 \geq 1. \quad (\text{A.20})$$

⁷ Any deviation from this assumption enters the problem as a constant factor multiplying the parameter θ , and can hence be absorbed into θ .

The first factor on the left-hand side of equation (A.20) is just the variance

$$V[\hat{\theta}(m)] = \sum_m p(m|\theta)(\hat{\theta}(m) - \theta)^2. \quad (\text{A.21})$$

Dividing by the second factor, which we identify with the Fisher information

$$I(\rho(\theta)) = \sum_m p(m|\theta) \left(\frac{\partial}{\partial \theta} \log p(m|\theta) \right)^2, \quad (\text{A.22})$$

we finally obtain the CR inequality

$$V[\hat{\theta}(m)] \geq \frac{1}{I(\rho(\theta))}. \quad (\text{A.23})$$

A.1.4. Heisenberg scaling in local metrology. Let us now revisit the local phase estimation scenario for an entangled state, for instance, the N -qubit GHZ state, given by

$$|\psi_{\text{GHZ}}\rangle = \frac{1}{\sqrt{2}}(|0\rangle^{\otimes N} + |1\rangle^{\otimes N}). \quad (\text{A.24})$$

A quick calculation of the QFI of equation (A.12) for this state provides the result $\mathcal{I}(|\psi_{\text{GHZ}}\rangle) = N^2$. The precision may hence quadratically increase with the number of qubits. This optimal scaling behaviour is usually called the *Heisenberg limit*. To see how one can practically achieve Heisenberg scaling, let us consider a simple parity measurement, that is, a projective measurement with outcomes $m = +1$ (even) and $m = -1$ (odd), and associated POVM elements

$$E_{\text{even}} = \sum_{n \text{ even}} E_n \quad \text{and} \quad E_{\text{odd}} = \sum_{n \text{ odd}} E_n, \quad (\text{A.25})$$

where E_n projects into the subspace where n qubits are in the state $|-\rangle$ and $(N - n)$ qubits are in the state $|+\rangle$. Denoting the single-qubit projectors as $P_{\pm} = |\pm\rangle\langle\pm|$, we can write

$$E_n = \sum_i \pi_i (P_-^{\otimes n} P_+^{\otimes N-n}), \quad (\text{A.26})$$

where the sum is over all $\binom{N}{n}$ permutations π_i . One then straightforwardly finds

$$\text{tr}(E_n \rho(\theta)) = \frac{1}{2^N} \binom{N}{n} (1 + (-1)^n \cos(N\theta)), \quad (\text{A.27})$$

which in turn yields the conditional probabilities

$$p(+|\theta) = \text{tr}(E_{\text{even}} \rho(\theta)) = \cos^2\left(\frac{N\theta}{2}\right), \quad (\text{A.28a})$$

$$p(-|\theta) = \text{tr}(E_{\text{odd}} \rho(\theta)) = \sin^2\left(\frac{N\theta}{2}\right). \quad (\text{A.28b})$$

Using the definition in equation (A.8) one can then verify that this measurement is optimal, i.e., the FI and QFI coincide, $I(|\psi_{\text{GHZ}}\rangle) = \mathcal{I}(|\psi_{\text{GHZ}}\rangle) = N^2$.

We then only need to find a suitable estimator. We can construct such an estimator from the expected value of the associated observable M , which has the spectral decomposition $M = E_{\text{even}} - E_{\text{odd}} = X^{\otimes N}$, such that $\langle M \rangle = \cos(N\theta)$. Crucially, note that the required measurements are just local X -measurements, the results of which are multiplied to obtain the overall measurement result in each run, i.e., $m = m_1 m_2 \dots m_N$. For $\theta \in [0, \frac{\pi}{N}]$ we then assign the estimator

$$\hat{\theta}(m) = \frac{1}{N} \arccos(m) = \frac{1-m}{2N} \pi = \begin{cases} 0 & \text{if } m = +1 \\ \frac{\pi}{N} & \text{if } m = -1 \end{cases}. \quad (\text{A.29})$$

Computing the mean and variance for this estimator one finds,

$$\langle \hat{\theta}(m) \rangle = \frac{\pi}{N} \sin^2\left(\frac{N\theta}{2}\right), \quad (\text{A.30a})$$

$$V[\hat{\theta}(m)] = \theta^2 + \sin^2\left(\frac{N\theta}{2}\right) \left(\frac{\pi^2}{N^2} - \frac{2\pi\theta}{N} \right). \quad (\text{A.30b})$$

The estimator is only unbiased for $\theta = \frac{\pi}{2N}$, but in this case the variance admits Heisenberg scaling and takes the value $\frac{\pi^2}{4N^2}$. However, one can do better than this by averaging over the outcomes before assigning an estimate, rather than averaging the individual estimates. Practically speaking, one can view this as estimating $m(\theta) = \cos(N\theta)$ followed by a simple reparametrization using $\theta(m) = \arccos(m)/N$. This estimator is unbiased by definition, since $m(\theta) = \langle M \rangle$ and one finds the variance

$$V[m(\theta)] = \sum_m p(m|m(\theta))(m - \langle M \rangle)^2 = \sin^2(N\theta). \quad (\text{A.31})$$

Propagating the error through the reparameterization then yields

$$V[\theta(m)] = V[m(\theta)] \left(\frac{\partial \theta}{\partial m} \right)^2 = \frac{1}{N^2}. \quad (\text{A.32})$$

One can hence get a quadratic scaling advantage for local phase estimation using an N -qubit GHZ state and local measurements. By extension via error propagation, Heisenberg scaling is also maintained for frequency estimation by reparameterizing $\theta = \omega t$ for any fixed interrogation time t . As shown in figure 2 in the main text, the preparation of an N -qubit GHZ state can be realised using a $(2N - 1)$ -qubit 1D cluster state, which hence constitutes a resource for local phase and frequency estimation at the Heisenberg limit.

A.2. Bayesian parameter estimation

In appendix A.2.1, we first review the basic structure of Bayesian parameter estimation problems. We then discuss an inequality that serves as a Bayesian analogue of the CR bound in appendix A.2.2 and present a simple proof in appendix A.2.3, before highlighting an interesting connection between Bayesian estimation and noisy local estimation in appendix A.2.4. Finally, we investigate the limitations of the MSE cost function for Bayesian estimation in appendix A.2.5.

A.2.1. The Bayesian estimation scenario. Much like in the local estimation scenario discussed in appendix A.1, the Bayesian scenario considers the estimation of a parameter θ that has been encoded onto a quantum state $\rho(\theta)$ by performing a measurement given by some POVM $\{E_m\}$. As before, the conditional probability to obtain the outcome m given that the parameter has the value θ is

$$p(m|\theta) = \text{tr}(E_m \rho(\theta)). \quad (\text{A.33})$$

However, where the local estimation scenario requires only that the parameter be close to values for which an unbiased estimator is available, the Bayesian estimation scenario captures all previously held belief about θ in a probability distribution referred to as the *prior* $p(\theta)$. Performing a single measurement, the probability to obtain the outcome m is then simply

$$p(m) = \int d\theta p(\theta) p(m|\theta) = \text{tr}(E_m \Gamma), \quad (\text{A.34})$$

where we have defined the quantity

$$\Gamma = \int d\theta p(\theta) \rho(\theta), \quad (\text{A.35})$$

following the notation of [36]. Given some outcome m , we then want to provide an estimate $\hat{\theta}(m)$ for the value of the parameter. To this end, note that Bayes' law lets us determine $p(\theta|m)$, the probability that the parameter had the value θ given the outcome m , as

$$p(\theta|m) = \frac{p(m|\theta)p(\theta)}{p(m)}. \quad (\text{A.36})$$

As an estimate we then simply average the possible values of θ weighted with the corresponding probabilities $p(\theta|m)$, i.e.,

$$\hat{\theta}(m) = \int d\theta p(\theta|m) \theta = \int d\theta \frac{p(m|\theta)p(\theta)}{p(m)} \theta = \frac{\text{tr}(E_m \eta)}{\text{tr}(E_m \Gamma)}, \quad (\text{A.37})$$

where we have inserted from equations (A.36) and (A.34), and defined the new quantity [36]

$$\eta = \int d\theta p(\theta) \rho(\theta) \theta. \quad (\text{A.38})$$

Thus, the estimate for θ given that the outcome m was observed, depends on the prior $p(\theta)$ and the encoding of the parameter in the state $\rho(\theta)$ via the quantities Γ and η from equations (A.35) and (A.38), respectively, and on the chosen POVM $\{E_m\}$. Note that the estimator used for the Bayesian estimation scenario need not be unbiased in the sense of equation (A.3). Instead, on average, we now expect the estimator to assign the same mean value as the prior, i.e.,

$$\bar{\theta} = \sum_m p(m) \hat{\theta}(m) = \sum_m \text{tr}(E_m \eta) = \text{tr}(\eta) = \int d\theta p(\theta) \theta. \quad (\text{A.39})$$

As a figure of merit for the precision of the estimate, we then wish to quantify how close $\hat{\theta}(m)$ is to θ according to our updated belief. We are hence interested in the variance $V_{\text{post}}^{(m)}$ of the *posterior* $p(\theta|m)$ given the outcome m . Using the MSE approach as in equation (A.4), but now with the posterior instead of the conditional

probability $p(m|\theta)$, we write

$$\begin{aligned} V_{\text{post}}^{(m)} &= V[p(\theta|m)] = \int d\theta p(\theta|m)(\theta - \hat{\theta}(m))^2 \\ &= \frac{1}{p(m)} \left[\text{tr} \left(E_m \int d\theta p(\theta) \rho(\theta) \theta^2 \right) - \frac{(\text{tr}(E_m \eta))^2}{\text{tr}(E_m \Gamma)} \right], \end{aligned} \quad (\text{A.40})$$

where we have used (A.36) and (A.37). In general, the width of the posterior may decrease or increase with respect to the width of the prior, depending on the measurement outcome. It is therefore more useful to average over different outcomes and define

$$\bar{V}_{\text{post}} = \sum_m p(m) V_{\text{post}}^{(m)} = \int d\theta p(\theta) \theta^2 - \sum_m \frac{(\text{tr}(E_m \eta))^2}{\text{tr}(E_m \Gamma)}. \quad (\text{A.41})$$

Here, a comment on the choice of \bar{V}_{post} as a figure of merit for the average increase in the knowledge is in order. For parameters (and priors) that have support on the entirety of \mathbb{R} , the MSE is certainly a useful choice. However, when estimating parameters with bounded support other quantifiers of the width of the posterior may be more appropriate. For instance, for phase estimation one may consider the Holevo phase variance as discussed in section 3. We will nonetheless consider the MSE in the following. This has several reasons. First, the MSE can still be useful for phase estimation when the priors are suitably narrow (see appendix A.2.5) and it allows to establish some simple bounds (see appendix A.2.2) for the optimal classical estimation strategies, as we shall explain in appendix A.3.2. Second, the MSE is of course useful for frequency estimation problems (see appendix A.5), where the parameter range is not bounded. We hence allow the parameter to take values $\theta \in [-\infty, \infty]$ for the remainder of this work.

As a simple example, consider a Gaussian prior of width $\sigma > 0$ centred at $\theta = \theta_0$, that is,

$$p(\theta) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(\theta - \theta_0)^2}{2\sigma^2}}, \quad (\text{A.42})$$

with $\bar{\theta} = \theta_0$ and $V[p(\theta)] = \int d\theta p(\theta)(\theta - \theta_0)^2 = \sigma^2$. The first term on the right-hand side of equation (A.41) then evaluates to

$$\int d\theta p(\theta) \theta^2 = \sigma^2 + \theta_0^2, \quad (\text{A.43})$$

while the remaining term

$$\theta_0^2 \leq \sum_m \frac{(\text{tr}(E_m \eta))^2}{\text{tr}(E_m \Gamma)} < \sigma^2 + \theta_0^2 \quad (\text{A.44})$$

determines the average decrease in width of the posterior with respect to the prior.

A.2.2. A Bayesian CR bound. The average variance of the posterior can be bounded from below using the *van Trees inequality* (see, e.g., [74, 78] or appendix A.2.3)

$$\bar{V}_{\text{post}} \geq \frac{1}{I(p(\theta)) + \bar{I}(\rho(\theta))}, \quad (\text{A.45})$$

which can be viewed as a type of CR bound for Bayesian estimation, where

$$I(p(\theta)) = \int d\theta p(\theta) \left(\frac{\partial}{\partial \theta} \log p(\theta) \right)^2, \quad (\text{A.46})$$

is the classical Fisher information of the prior and

$$\bar{I}(\rho(\theta)) = \int d\theta p(\theta) I(\rho(\theta)) = \int d\theta p(\theta) \sum_m p(m|\theta) \left(\frac{\partial}{\partial \theta} \log p(m|\theta) \right)^2.$$

is the averaged (over the unknown parameter θ) FI associated to the state $\rho(\theta)$ and the POVM $\{E_m\}$ as specified in equation (A.8).

Since the QFI $\mathcal{I}(\rho(\theta))$ arises as a maximisation of the FI $I(\rho(\theta))$ over all POVMs, we have $\mathcal{I}(\rho(\theta)) \geq I(\rho(\theta))$. If, as before for the local case, we consider the parameter to be encoded by a unitary transformation of the form of equation (A.1), the QFI is independent of θ , as we have shown in appendix A.1.2. This allows us to bound the average FI by the QFI, i.e.,

$$\bar{I}(\rho(\theta)) \leq \int d\theta p(\theta) \mathcal{I}(\rho(\theta)) = \mathcal{I}(\rho(\theta)), \quad (\text{A.47})$$

and consequently we can modify the van Trees inequality to

$$\bar{V}_{\text{post}} \geq \frac{1}{I(p(\theta)) + \mathcal{I}(\rho(\theta))}. \quad (\text{A.48})$$

In contrast to the (quantum) CR inequality (A.7), the bounds in (A.45) and (A.48) are generally not tight, so they do not allow us to conclude that a measurement strategy exists such that $1/\bar{V}_{\text{post}}$ grows quadratically with N . And while it can indeed be shown [25] that Heisenberg scaling is asymptotically achievable for arbitrary priors in the Bayesian regime we require an explicit description of the involved states and measurements to determine whether these can be efficiently implemented.

Nonetheless, a simple consequence of the van Trees inequality pertains to the classical scaling behaviour. Recall from equation (A.14) that the maximal value of the QFI for product states is proportional to N . This implies that \bar{V}_{post} decreases at most linearly with N for classical strategies, i.e., $1/\bar{V}_{\text{post}} \leq N + I(p(\theta))$, where $I(p(\theta))$ is a constant independent of N . For instance, for the Gaussian prior of equation (A.42), which we want to focus on in the following, we have $I(p(\theta)) = 1/\sigma^2$.

At this point, two comments on the choice of Gaussian priors are in order. First, note that there exists an interesting connection between Bayesian estimation with Gaussian priors and local estimation subject to parallel, Gaussian noise [26]. As is outlined in appendix A.2.4, this connection provides an alternative way of computing the variance \bar{V}_{post} via the (quantum) Fisher information of the probe state after a noisy channel. Here, we do not explicitly consider the problem of noisy metrology in more detail, but we refer the interested reader to [59, 60].

The second comment concerns the fact that the probability distribution of equation (A.42) has support on the entire real line, whereas for phase estimation, θ only takes values in an interval of length 2π . In addition, the use of the MSE means that differences between estimates and parameter values larger than π are disproportionately penalised. Intuitively it is clear that this becomes an issue when the width of the Gaussian prior becomes comparable with (half of) the length of the interval for θ . In appendix A.2.5 this problem is discussed in more detail.

For sufficiently narrow priors the MSE is hence still a useful cost function for the variance and (non-wrapped) Gaussians can be employed instead of the more complicated wrapped Gaussians to simplify calculations. Moreover, the use of the MSE (rather than some circular statistics equivalent or covariant cost function, see [4]) as a measure for the precision of the estimate allows us to remain within the framework of [36]. It also permits us to apply the Bayesian CR bound of inequation (A.48), which provides a straightforward comparison with classical strategies, as we shall discuss in appendix A.3.2. Finally, note that these considerations arise for the phase estimation problem discussed in this section, but are no cause for concern in the frequency estimation paradigm, which is presented in appendix A.5.

A.2.3. A proof of the Bayesian CR bound. We now want to present an explicit proof that the average variance \bar{V}_{post} of the posterior $p(\theta|m)$ can be bounded from below by the *van Trees inequality* [78], which is the Bayesian equivalent of the CR bound, given by

$$\bar{V}_{\text{post}} \geq \frac{1}{I(p(\theta)) + \bar{I}(\rho(\theta))}, \quad (\text{A.49})$$

where $I(p(\theta))$ is the classical Fisher information of the prior, given by

$$I(p(\theta)) = \int d\theta p(\theta) \left(\frac{\partial}{\partial \theta} \log p(\theta) \right)^2, \quad (\text{A.50})$$

and $\bar{I}(\rho(\theta)) = \bar{I}(\rho(\theta), \{E_m\})$ is the Fisher information associated to the state $\rho(\theta)$ and the POVM $\{E_m\}$, averaged over the (unknown) parameter θ . That is, it is given by

$$\bar{I}(\rho(\theta)) = \int d\theta p(\theta) I(\rho(\theta)) = \int d\theta p(\theta) \sum_m p(m|\theta) \left(\frac{\partial}{\partial \theta} \log p(m|\theta) \right)^2. \quad (\text{A.51})$$

In the frequency estimation scenario, the parameter θ is typically allowed to take on any value in \mathbb{R} , but the prior is assumed to have compact support, such that $p(\pm\infty) = 0$. In the phase estimation scenario, on the other hand, the parameter can take values in the interval $[a, a + 2\pi]$ for some $a \in \mathbb{R}$ and w.l.o.g. one may pick $a = 0$. In this case, one may assume that the probability densities are either wrapped, e.g., the prior satisfies $p(\theta) = p(\theta \bmod 2\pi)$ and θ is to be understood as $\theta \bmod 2\pi$. Alternatively, one can also treat θ to be any real number, and require that the prior be sufficiently narrow. In the latter scenario, one can still use the MSE approach for the variance, but care needs to be taken with the initial width of the prior, as discussed in appendix A.2.5. With this in mind, we now discuss a proof of equation (A.49). First, note that

$$\int_a^b d\theta \hat{\theta}(m) \frac{\partial}{\partial \theta} (p(\theta)p(m|\theta)) = \hat{\theta}(m)[p(\theta)p(m|\theta)]_a^b = 0, \quad (\text{A.52})$$

due to the assumptions above for $(a, b) = (0, 2\pi)$ or $(a, b) = (-\infty, +\infty)$, respectively. Similarly, integration by parts immediately lets us evaluate the integral

$$\int_a^b d\theta \theta \frac{\partial}{\partial \theta} (p(\theta)p(m|\theta)) = - \int_a^b d\theta p(\theta)p(m|\theta), \quad (\text{A.53})$$

where we have eliminated the boundary term using the previous assumptions. Using Bayes' law (see equation (A.36)) and the preliminary results of equations (A.52) and (A.53), we can then calculate

$$\int d\theta (\hat{\theta}(m) - \theta) \frac{\partial}{\partial \theta} (p(\theta)p(m|\theta)) = \int d\theta p(\theta)p(m|\theta) = p(m) \int d\theta p(\theta|m) = p(m), \quad (\text{A.54})$$

since $p(\theta|m)$ is normalised. When we sum over the possible measurement outcomes, we must hence just get

$$\sum_m \int d\theta (\hat{\theta}(m) - \theta) \frac{\partial}{\partial \theta} (p(\theta)p(m|\theta)) = \sum_m p(m) = 1. \quad (\text{A.55})$$

On the other hand, we can rewrite parts of the integrand as

$$\frac{\partial}{\partial \theta} (p(\theta)p(m|\theta)) = p(\theta)p(m|\theta) \frac{\partial}{\partial \theta} \log(p(\theta)p(m|\theta)) \quad (\text{A.56})$$

and use the Cauchy–Schwarz inequality from equation (A.19) with $x = (\hat{\theta}(m) - \theta) \sqrt{p(\theta)p(m|\theta)}$ and

$$y = \sqrt{p(\theta)p(m|\theta)} \frac{\partial}{\partial \theta} \log[p(\theta)p(m|\theta)] \quad (\text{A.57})$$

to arrive at the inequality

$$1 \leq \sum_m \int d\theta p(\theta)p(m|\theta) (\hat{\theta}(m) - \theta)^2 \times \sum_n \int d\theta' p(\theta')p(n|\theta') \left(\frac{\partial}{\partial \theta'} \log[p(\theta')p(n|\theta')] \right)^2. \quad (\text{A.58})$$

The first factor on the right-hand side of equation (A.58) is just \bar{V}_{post} from equation (A.41). The second factor can be split into three terms by squaring

$$\frac{\partial}{\partial \theta} \log[p(\theta)p(m|\theta)] = \frac{\partial}{\partial \theta} \log p(\theta) + \frac{\partial}{\partial \theta} \log p(m|\theta). \quad (\text{A.59})$$

Summing over the normalised conditional probability $p(m|\theta)$, the first term gives the classical Fisher information for the prior, i.e.,

$$\int d\theta p(\theta) \left(\frac{\partial}{\partial \theta} \log p(\theta) \right)^2 = I(p(\theta)), \quad (\text{A.60})$$

while the term containing the square of $\frac{\partial}{\partial \theta} \log p(m|\theta)$ gives $\bar{I}(\rho(\theta))$ as defined in equation (A.51). The remaining cross term is of the form

$$\begin{aligned} \sum_m \int d\theta p(\theta)p(m|\theta) \left(\frac{\partial}{\partial \theta} \log p(\theta) \right) \left(\frac{\partial}{\partial \theta} \log p(m|\theta) \right) &= \sum_m \int d\theta \left(\frac{\partial}{\partial \theta} p(\theta) \right) \left(\frac{\partial}{\partial \theta} p(m|\theta) \right) \\ &= \int d\theta \left(\frac{\partial}{\partial \theta} p(\theta) \right) \left(\frac{\partial}{\partial \theta} \sum_m p(m|\theta) \right) = 0, \end{aligned} \quad (\text{A.61})$$

which vanishes since the sum over $p(m|\theta)$ is independent of θ , i.e.,

$$\sum_m p(m|\theta) = \sum_m \text{tr}(E_m \rho(\theta)) = \text{tr}(\rho) = 1. \quad (\text{A.62})$$

Dividing both sides of the inequality in (A.58) by the sum of the nonzero terms of the second factor, we arrive at the Bayesian CR bound

$$\bar{V}_{\text{post}} \geq \frac{1}{I(p(\theta)) + \bar{I}(\rho(\theta))}. \quad (\text{A.63})$$

A.2.4. Relating noisy local estimation with Bayesian estimation for Gaussian priors. In this appendix we discuss an interesting connection between noisy local estimation and Bayesian estimation for Gaussian priors. We hence consider a local estimation scenario as in section A.1.1, where ‘parallel’ noise is present on top of the unitary encoding of equation (A.1). That is, the noise is generated by the same Hamiltonian as the encoding of the parameter but distributed according to some probability distribution $\tilde{p}(\theta)$. The state encoding the parameter is then given by

$$\tilde{\rho}(\theta) = U_\theta \tilde{\rho}(0) U_\theta^\dagger, \quad (\text{A.64})$$

where the noise can be understood as part of preparing the initial state

$$\tilde{\rho}(0) = \int d\theta' \tilde{p}(\theta') U_{\theta'} |\psi\rangle\langle\psi| U_{\theta'}^\dagger \quad (\text{A.65})$$

starting from some pure state $|\psi\rangle$. We further assume that the noise has a Gaussian profile centred around zero, that is, the noise distribution is

$$\tilde{p}(\theta) = p(\theta + \theta_o), \quad (\text{A.66})$$

where θ_o is the mean of the Gaussian prior $p(\theta)$ of equation (A.42). We can now see how the encoded state of this noisy local scenario corresponds to the quantity Γ from equation (A.35) in the Bayesian scenario, i.e.,

$$\begin{aligned} \Gamma &= \int d\theta p(\theta) \rho(\theta) = \int d\theta p(\theta) U_\theta |\psi\rangle\langle\psi| U_\theta^\dagger \\ &= \int d\theta' p(\theta' + \theta_o) U_{\theta'+\theta_o} |\psi\rangle\langle\psi| U_{\theta'+\theta_o}^\dagger = \tilde{\rho}(\theta_o), \end{aligned} \quad (\text{A.67})$$

where we have substituted $\theta' = \theta - \theta_o$. To establish a similar connection for η from equation (A.38), we make use of the fact that the prior (and the noise distribution in the local scenario) are Gaussian, such that

$$\frac{d}{d\theta} p(\theta) = \dot{p}(\theta) = -\frac{\theta - \theta_o}{\sigma^2} p(\theta). \quad (\text{A.68})$$

With this, we find

$$\begin{aligned} \eta &= \int d\theta \theta p(\theta) \rho(\theta) = \theta_o \int d\theta p(\theta) \rho(\theta) - \sigma^2 \int d\theta \dot{p}(\theta) \rho(\theta) \\ &= \theta_o \Gamma + i\sigma^2 \int d\theta p(\theta) [\rho(\theta), H] = \theta_o \Gamma + \sigma^2 \dot{\tilde{\rho}}(\theta_o), \end{aligned} \quad (\text{A.69})$$

where the dot indicates a partial derivative w.r.t. θ . Reinserting the expressions for Γ and η into equation (A.41), the last term gives

$$\sum_m \frac{(\text{tr}(E_m \eta))^2}{\text{tr}(E_m \Gamma)} = \sigma^4 \sum_m \frac{(\text{tr}[E_m \dot{\tilde{\rho}}(\theta_o)])^2}{\text{tr}(E_m \tilde{\rho}(\theta_o))} + \theta_o^2. \quad (\text{A.70})$$

The first term on the right-hand side of equation (A.70) can easily be recognised as the Fisher information $I(\tilde{\rho}(\theta_o))$ from equation (A.8) for the POVM $\{E_m\}$ in the local scenario with parallel Gaussian noise. Combining this result with equations (A.41) and (A.43), we find

$$\bar{V}_{\text{post}} = \sigma^2 - \sigma^4 I(\tilde{\rho}(\theta_o)). \quad (\text{A.71})$$

Since the variance of the initial Gaussian prior is just $V[p(\theta)] = \sigma^2$, one arrives at the conclusion that the average decrease in variance in the Bayesian scenario with Gaussian prior, $\Delta V = V[p(\theta)] - \bar{V}_{\text{post}}$ for any given POVM is proportional to the Fisher information at $\theta = \theta_o$ for the same POVM in the local scenario with parallel Gaussian noise, $\Delta V = \sigma^4 I(\tilde{\rho}(\theta_o))$. In particular, for the optimal POVM one obtains the QFI, which is independent of the value of the parameter for the unitary encoding with parallel noise, and hence

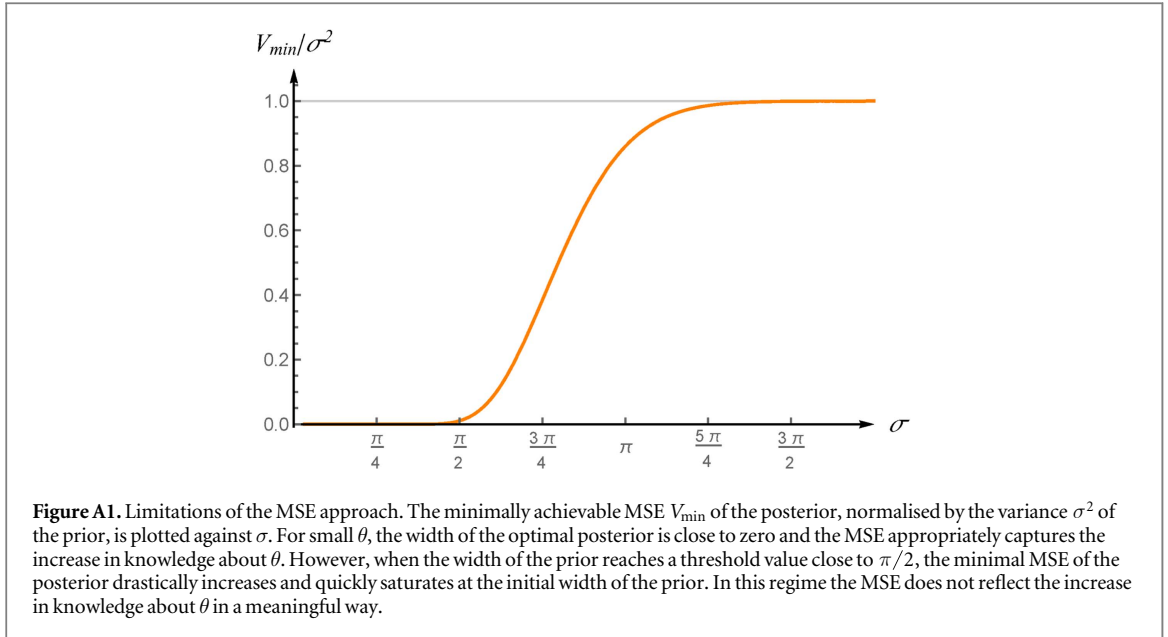
$$\Delta V_{\text{opt}} = \sigma^4 \mathcal{I}(\tilde{\rho}). \quad (\text{A.72})$$

This result immediately informs us about an important property of the Bayesian scenario. Since $\mathcal{I}(\tilde{\rho})$ is the QFI in a scenario with parallel noise that can be viewed as dephasing, one cannot expect Heisenberg scaling of $\mathcal{I}(\tilde{\rho})$, i.e., that $\mathcal{I}(\tilde{\rho})$ increases quadratically with N , see [59, 60]. Instead, it is clear that $\mathcal{I}(\tilde{\rho}) \leq 1/\sigma^2$ since $\bar{V}_{\text{post}} \geq 0$. On the other hand, one expects that $\mathcal{I}(\tilde{\rho})$ approaches the bound $1/\sigma^2$ from below as N increases. As suggested in [61], it is reasonable to assume that

$$\mathcal{I}(\tilde{\rho}) = \frac{1}{\sigma^2} - \frac{K}{N^\alpha} \quad (\text{A.73})$$

for some positive constant K and some power $\alpha \geq 1$, such that $\bar{V}_{\text{post}} = K\sigma^4/N^\alpha$. Therefore, a scaling advantage of a quantum strategy with respect to a classical strategy is obtained if one finds an (efficiently preparable) state and POVM such that $\alpha > 1$.

A.2.5. Limitations of the MSE approach. Here, we aim to discuss the limitations of applicability of the MSE cost function for Bayesian phase estimation, i.e., for a scenario where the parameter θ is encoded by a unitary $U_\theta = e^{-i\theta H}$, with $H = \frac{1}{2}Z$ for each qubit. Since the difference between the two eigenvalues of H is 1, it is immediately apparent that values of θ that differ by 2π cannot be distinguished in such a scenario. This periodicity is not accurately reflected in the use of the MSE, since estimates that differ by integer multiples of 2π are unduly penalised. In a local estimation scenario where small fluctuations around a fixed value of the parameter are being estimated, this is not an issue. Similarly, this is of no concern for Bayesian estimation when



the prior is sufficiently localised, but can become an issue for larger values of σ (where we focus on Gaussian priors as in equation (A.42)). We are therefore interested in quantifying for which values of σ the approach using the MSE cost function becomes problematic.

We will take a pragmatic point of view and consider the MSE approach as useful, if this post-processing of the measurement data provides an increase in knowledge in the sense of an average decrease of the width of the posterior $p(\theta|m)$. We therefore ask, what the minimal MSE of the posterior can be in principle, given a fixed Gaussian prior of width σ . When obtaining a measurement outcome m , the corresponding estimate may in principle only be understood modulo 2π . In other words, if no prior knowledge is available, and one were to trust the estimate of the parameter unconditionally, the posterior would be a ‘comb’ of Dirac delta functions $\delta(\theta - 2\pi k)$ for all values k such that $\theta - 2\pi k$ lies within the allowed range of parameters. For an unrestricted range, $\theta \in \mathbb{R}$, we hence have infinitely many side-peaks at distances $2\pi k$ for $k \in \mathbb{Z}$. If we take into account the prior information, some of these peaks are suppressed by its shape, e.g., as $\exp\left(-\frac{\theta^2}{2\sigma^2}\right)$ for a Gaussian prior. The optimally reachable posterior is then given by

$$p_{\text{opt}}(\theta) = \mathcal{N} \sum_{k \in \mathbb{Z}} e^{-\frac{\theta^2}{2\sigma^2}} \delta(\theta - 2\pi k), \quad (\text{A.74})$$

where the normalisation is given by $\mathcal{N}^{-1} = \sum_{k \in \mathbb{Z}} \exp\left(-\frac{(2\pi k)^2}{2\sigma^2}\right)$. The MSE of this distribution is

$$V_{\min} = \mathcal{N} \sum_{k \in \mathbb{Z}} e^{-\frac{(2\pi k)^2}{2\sigma^2}} (2\pi k)^2. \quad (\text{A.75})$$

As illustrated in figure A1, the MSE of this optimal posterior strongly increases from around $\sigma \approx \pi/2$, and from around $5\pi/4$ the width stays virtually constant as compared to the MSE of the prior. Of course this does not mean that the measurement process does not provide information about the parameter. Clearly, knowing the value of θ exactly modulo 2π is more useful than a uniform prior. However, the MSE simply fails to capture this distinction. We hence have to keep this limited applicability of the approach using (non-wrapped, Gaussian) priors and the MSE cost function in mind. Specifically, we restrict our analysis to Gaussian priors of widths smaller or equal than 1.

A.3. Classical Bayesian estimation strategies

After introducing the quantities of interest for Bayesian parameter estimation in the previous appendix, we now want to illustrate these techniques for classical Bayesian estimation. This provides the opportunity to establish a direct comparison with the results obtained for a strategy exploiting quantum features that we will present in appendix A.4.

We consider a strategy to be classical, if no quantum correlations are used for the state preparation or measurement, which corresponds to the choice of product states for N qubits along with single-qubit measurements. The Bayesian approach allows updating the estimation strategy based on the outcomes of previous measurements. Consequently, a parallel strategy of N individual probes that are prepared and measured in the same way may not be optimal even among the classical measurement schemes. At the same

time, the explicit evaluation of a sequential measurement strategy with intermediate updates is computationally extremely demanding. To give a fair representation of the performance of classical strategies we hence consider a bound for the sequential measurement scheme in appendix A.3.2, and compute the average variance explicitly for the optimal parallel strategy in appendix A.3.3. In preparation for these scenarios, we begin with the single-qubit Bayesian estimation problem in appendix A.3.1

A.3.1. Single-qubit measurements. For the scenario that we consider here, i.e., Gaussian priors as in equation (A.42) and unitary parameter encodings as in equation (A.1), the optimal single-qubit measurement strategy for Bayesian estimation is similar to that of the local scenario. That is, the probe state is chosen to be $|+\rangle$, i.e., a uniform superposition of the eigenstates of H . The optimal accompanying measurement is a projective measurement with POVM elements

$$\tilde{E}_{\pm} = U_{\theta_o + \pi/2} |\pm\rangle\langle\pm| U_{\theta_o + \pi/2}^{\dagger}, \quad (\text{A.76})$$

which corresponds to a measurement in a direction on the equatorial plane of the Bloch sphere that is orthogonal to the direction obtained by rotating $|\pm\rangle$ by the expected value θ_o of the prior. This can be seen by noting that probe states and measurement directions can be restricted to the equatorial plane, followed by an optimisation over the angle defining their relative orientation. For this combination of state and measurement, the conditional probabilities to obtain the outcomes ‘+’ or ‘-’ are

$$p(\pm|\theta) = \text{tr}(\tilde{E}_{\pm}\rho(\theta)) = \frac{1}{2}(1 \pm \sin(\theta - \theta_o)), \quad (\text{A.77})$$

such that $p(\pm|\theta_o) = 1/2$. We further compute

$$\text{tr}(\tilde{E}_{\pm}\Gamma) = p(m = \pm) = \frac{1}{2}, \quad (\text{A.78a})$$

$$\text{tr}(\tilde{E}_{\pm}\eta) = \frac{1}{2}(\theta_o \pm \sigma^2 e^{-\sigma^2/2}). \quad (\text{A.78b})$$

The corresponding estimates are then easily found by inserting into equation (A.37), yielding

$$\theta_{m=\pm} = \theta_o \pm \sigma^2 e^{-\sigma^2/2}. \quad (\text{A.79})$$

Together with equations (A.41) and (A.43) we then find the average variance of the posterior

$$\bar{V}_{\text{post}} = \sigma^2(1 - \sigma^2 e^{-\sigma^2}). \quad (\text{A.80})$$

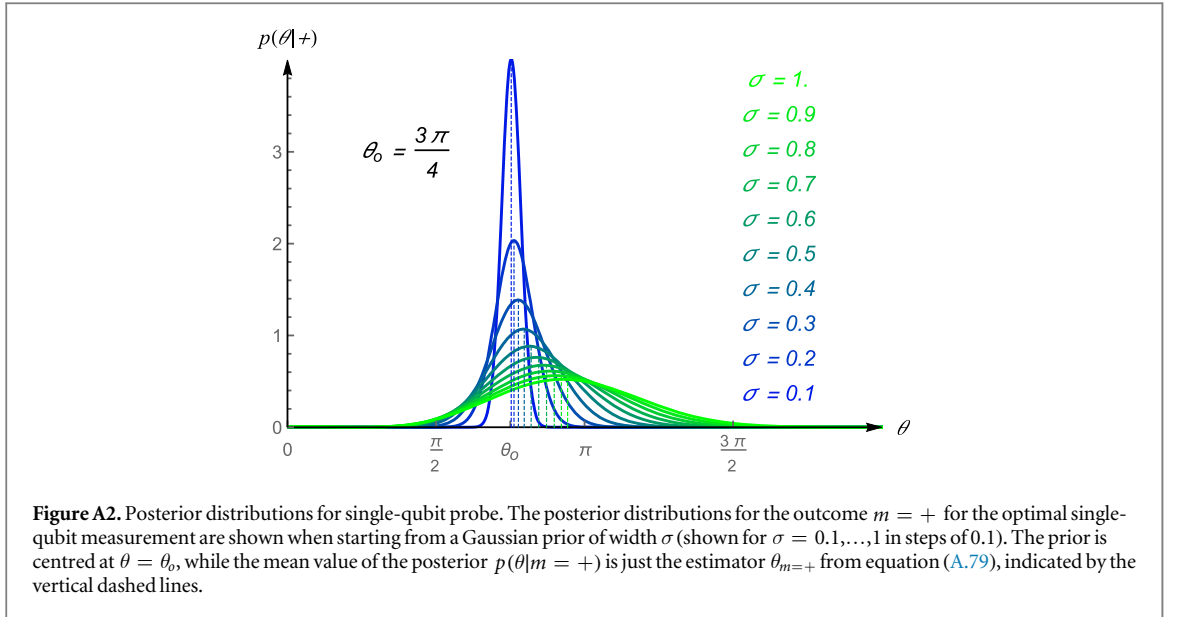
Since $0 < \sigma^2 e^{-\sigma^2} < 1$ for finite, nonzero σ , the variance decreases on average, $\bar{V}_{\text{post}} < \sigma^2$, but it becomes apparent that the decrease in width quantified by $\Delta V := \sigma^2 - \bar{V}_{\text{post}}$ has a maximum for $\sigma = 2$. This signifies that the MSE approach using Gaussian priors is not useful for priors of large width when considering phase estimation (see appendix A.2.5 for a discussion of this issue). It is also interesting to note that the posterior distributions conditional on the outcomes $m = \pm$ are given by

$$p(\theta|m = \pm) = \frac{1}{\sqrt{2\pi}\sigma}(1 \pm \sin(\theta - \theta_o))e^{-\frac{(\theta - \theta_o)^2}{2\sigma^2}}. \quad (\text{A.81})$$

Unlike the prior, the posterior distributions illustrated in figure A2 are no longer Gaussian, and they are not symmetric around their mean values $\theta = \theta_{m=\pm}$.

A.3.2. Bound for multi-qubit measurements. We are now interested in making statements about the optimal strategy for Bayesian estimation using a sequence of N consecutive single-qubit probes. Unfortunately, the posterior even after one measurement is no longer Gaussian (or symmetric). Therefore, determining the optimal single-qubit measurements and updating the prior becomes problematic for large numbers of measurements. This may not be an issue in an actual measurement, where each qubit gives a single outcome based on which the next measurement is chosen. However, we are interested in the variance of the posterior averaged over all possible sequential measurement outcomes, the set of which grows exponentially. Having 2^N potentially different posterior distributions makes such an approach computationally infeasible.

We shall therefore refrain from obtaining the exact expression for the optimal expected variance \bar{V}_{post} after N sequential single-qubit measurements with updated directions. Instead, we construct a bound based on the Bayesian CR inequality (A.48). We note that the updating procedure can be entirely thought of as part of the choice of measurement direction, while the probe state $|+\rangle$ remains the same throughout. Further recall that the QFI entails an optimisation over all possible measurements including correlated measurements that can depend on previous outcomes. A lower bound for \bar{V}_{post} in the classical case is hence obtained from the QFI for the state $|+\rangle^{\otimes N}$, which we have previously determined in equation (A.14) to be given by $\mathcal{I}(|+\rangle^{\otimes N}) = N$. Inserting into inequation (A.48) we arrive at the bound



$$\bar{V}_{\text{post}} \geq \frac{\sigma^2}{1 + N\sigma^2}, \quad (\text{A.82})$$

where we have used that $I(p(\theta)) = 1/\sigma^2$. Any classical strategy, may it consist of parallel or sequential measurements, must give an expected variance larger than this bound. This result also extends to the (asymptotic) behaviour of the Holevo phase variance V_ϕ of equation (18) since V_ϕ reduces to the MSE as $\sigma \rightarrow 0$ (see, e.g., [79, p 7]). Consequently, the Holevo phase variance of any successful sequential measurement strategy will approach the behaviour of the MSE. The faster (in terms of the number of measurements) the strategy decreases the phase variance, the sooner one will enter a regime where the bound of inequation (A.63) applies. Moreover, the bound in inequation (A.82) is not tight and might significantly overestimate the performance of classical strategies since the optimisation in the QFI also includes entangled measurements. We therefore complement this bound by an investigation into the optimal parallel strategy in section A.3.3.

A.3.3. Optimal parallel strategy. Having obtained the previous lower bound for \bar{V}_{post} for the optimal classical strategy, one may wonder, how close a practical classical strategy may come to this bound. To address this question, we now consider the optimal classical, parallel strategy for Gaussian priors. That is, we compute \bar{V}_{post} in the case where N qubits are identically prepared and measured (i.e., without intermediate updates) with the optimal single-qubit strategy based on the prior information (see section A.3.1). The probe state is hence $|+\rangle^{\otimes N}$ and for each qubit we perform the POVM with elements \tilde{E}_\pm as in equation (A.76). Since the state is invariant under the exchange of qubits, it is irrelevant, which of the N qubits give results ‘+’, and which give results ‘-’, we note that there are only $N + 1$ qualitatively different measurement outcomes.

We label these outcomes by $m = 0, 1, \dots, N$, which we take to be the number of outcomes ‘-’. In other words, for the given state this measurement is equivalent to the POVM with element E_m from equation (A.26). The conditional probability to obtain the outcome m , given that the parameter takes the value θ is then

$$p(m|\theta) = p(+|\theta)^{N-m} p(-|\theta)^m \binom{N}{m}. \quad (\text{A.83})$$

We then insert for $p(\pm|\theta)$ from equation (A.77) and find

$$\begin{aligned} \text{tr}(E_m \Gamma) &= \binom{N}{m} \int d\theta p(\theta) p(+|\theta)^{N-m} p(-|\theta)^m \\ &= \frac{1}{2^N} \binom{N}{m} \sum_{k=0}^{N-m} \sum_{k'=0}^m \binom{N-m}{k} \binom{m}{k'} (-1)^{k'} I_{k+k'}, \end{aligned} \quad (\text{A.84})$$

where the quantity $I_{k+k'}$ is given by

$$I_n = \frac{1}{\sqrt{2\pi}\sigma} \int d\theta e^{-\frac{\theta^2}{2\sigma^2}} \sin^n \theta \quad (\text{A.85})$$

and powers of the sine function arise from the binomial expansion of $(1 \pm \sin(\theta - \theta_0))$ followed by a substitution $\theta - \theta_0 \rightarrow \theta$. Now it is easy to see that the integral I_n vanishes for odd n . To solve the integral in equation (A.85) for even n , we use the trigonometric identity (which holds only for even n)

$$\sin^n \theta = \frac{1}{2^n} \binom{n}{\frac{n}{2}} + \frac{2}{2^n} \sum_{l=0}^{\frac{n}{2}-1} (-1)^{\frac{n}{2}-l} \binom{n}{l} \cos([n-2l]\theta), \quad (\text{A.86})$$

as well as the integral formula

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} d\theta e^{-\frac{\theta^2}{2\sigma^2}} \cos(a\theta) = e^{-a^2\sigma^2/2}. \quad (\text{A.87})$$

Combining equations (A.84) to (A.87) we obtain

$$\begin{aligned} \text{tr}(E_m \Gamma) &= \sum_{k=0}^{N-m} \sum_{k'=0}^m \frac{((-1)^{k'} + (-1)^k) N!}{2^{N+k+k'+1} k! k'! (N-m-k)! (m-k)!} \\ &\times \left\{ \frac{(k+k')!}{\left[\left(\frac{k+k'}{2}\right)!\right]^2} + \sum_{l=0}^{\frac{k+k'-2}{2}} \frac{(-1)^{\frac{k+k'-2l}{2}} e^{-\frac{(k+k'-2l)^2\sigma^2}{2}} (k+k')!}{l!(k+k'-l)!} \right\} \end{aligned} \quad (\text{A.88})$$

For the Gaussian prior we then need to further compute

$$\bar{V}_{\text{post}} = \sigma^2 + \theta_o^2 - \sum_m \frac{(\text{tr}(E_m \eta))^2}{\text{tr}(E_m \Gamma)} = \sigma^2 - \sum_m \frac{\gamma_m^2}{\text{tr}(E_m \Gamma)}, \quad (\text{A.89})$$

where

$$\gamma_m = \binom{N}{m} \int d\theta \theta p(\theta) p(+|\theta)^{N-m} p(-|\theta)^m = \frac{1}{2^N} \binom{N}{m} \sum_{k=0}^{N-m} \sum_{k'=0}^m \binom{N-m}{k} \binom{m}{k'} (-1)^{k'} J_{k+k'}. \quad (\text{A.90})$$

Here we have a different integral, $J_{k+k'}$, which is of the form

$$J_n = \frac{1}{\sqrt{2\pi}\sigma} \int d\theta \theta e^{-\frac{\theta^2}{2\sigma^2}} (\sin \theta)^n. \quad (\text{A.91})$$

Since $\theta \exp(-\frac{\theta^2}{2\sigma^2}) = -\sigma^2 \frac{\partial}{\partial \theta} \exp(-\frac{\theta^2}{2\sigma^2})$ we can easily integrate by parts and write

$$J_n = \frac{n\sigma}{\sqrt{2\pi}} \int d\theta e^{-\frac{\theta^2}{2\sigma^2}} (\sin \theta)^{n-1} \cos \theta. \quad (\text{A.92})$$

When $k+k' = n$ is even, the integral vanishes. When $k+k' = n$ is odd, on the other hand, then $(n-1)$ is even and we can use the trigonometric identity from equation (A.86) along with the formula

$$\cos \theta \cos([n-1-2l]\theta) = \frac{1}{2} \cos([n-2-2l]\theta) + \frac{1}{2} \cos([n-2l]\theta). \quad (\text{A.93})$$

This, together with the Gaussian integral of equation (A.87) allows us to rewrite J_n as

$$J_n = \frac{n\sigma^2 [1 - (-1)^n]}{2^n} \left\{ e^{-\frac{\sigma^2}{2}} \binom{n-1}{\frac{n-1}{2}} + \sum_{l=0}^{\frac{n-3}{2}} (-1)^{\frac{n-2l-1}{2}} \binom{n-1}{l} \left(e^{-\frac{(n-2l-2)^2\sigma^2}{2}} + e^{-\frac{(n-2l)^2\sigma^2}{2}} \right) \right\}. \quad (\text{A.94})$$

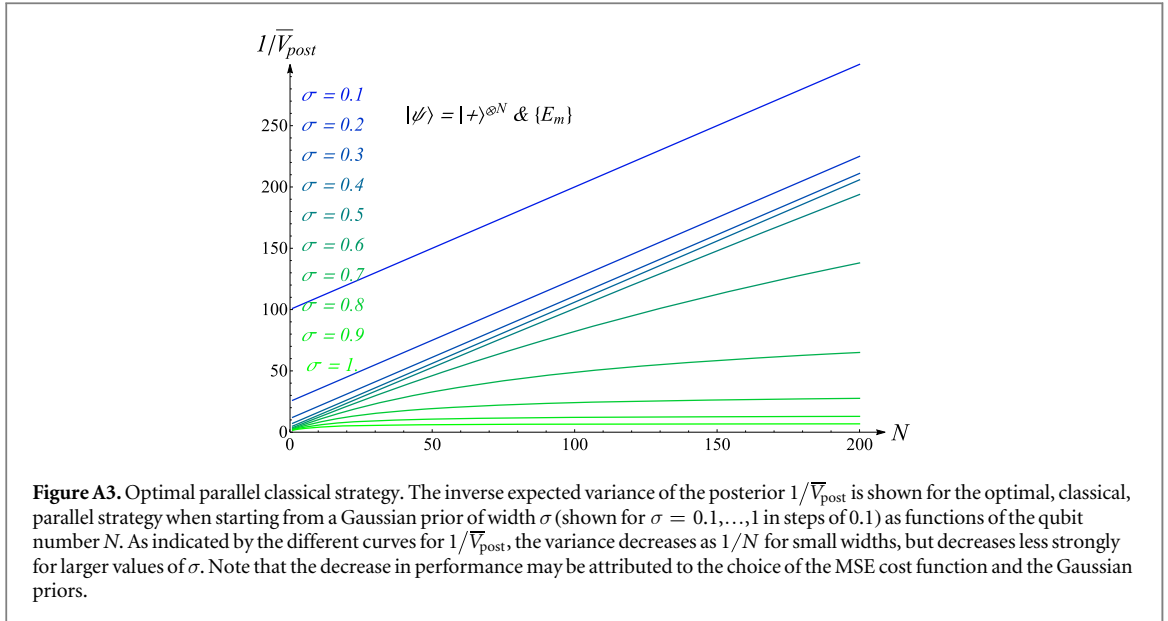
With this, the average variance of the posterior of equation (A.89) can be computed, which we have done for up to $N = 200$ qubits. The results, depicted in figure A3, show that \bar{V}_{post} decreases at most as $1/N$, as expected. Moreover, the data suggests that the parallel classical strategy is close the bound of inequation (A.82) when the width of the prior is much smaller than $\pi/2$. For instance, when $\sigma = 0.1$, the relative deviation of the data for \bar{V}_{post} from the bound $\bar{V}_{\text{min}} = \sigma^2/(1+N\sigma^2)$ as quantified by $\Delta\bar{V} := (\bar{V}_{\text{post}} - \bar{V}_{\text{min}})/\bar{V}_{\text{min}}$ is $\Delta\bar{V} < 5 \times 10^{-7}$ for $N = 1$ and $\Delta\bar{V} < 1.65 \times 10^{-5}$ for $N = 90$. For $\sigma = 0.5$ the corresponding deviations are already at $\Delta\bar{V} \approx 6.6 \times 10^{-3}$ and 2.9×10^{-2} for $N = 1$ and $N = 90$, respectively, and for larger values of σ the deviation of $1/\bar{V}_{\text{post}}$ from a function increasing linearly with N is already clearly visible in figure A3.

Having thoroughly investigated the performance of classical estimation strategies in Bayesian scenarios, we will next turn to strategies involving genuine quantum features.

A.4. Quantum advantage in Bayesian estimation

With respect to the local estimation scenario, Bayesian estimation is made considerably more complicated by the in principle arbitrary shape of the prior. Consequently, results on optimality are scarcely available apart from very special cases such as phase estimation for flat priors [19], for which an optimal (albeit with respect to a different cost function for the variance) pair of probe state and measurement have been determined. Here, we will discuss a slightly modified version of the scheme of [19] as an example and show that it can lead to a scaling advantage also for other choices of priors (and cost functions).

The probe state in question is a superposition of N -qubit computational basis states, where one representative $|n\rangle_{\text{un}} = |1\rangle^{\otimes n} |0\rangle^{\otimes N-n}$ is selected for each Hamming weight, i.e., from each subspace with a fixed



number of qubits in the state $|1\rangle$. That is, $|n\rangle_{\text{un}}$ is a unary encoding of the integer n . For flat priors (and using the Holevo phase variance [40] instead of the MSE of equation (A.40)), the optimal probe state $|\psi_{\text{sine}}\rangle$ is of the form

$$|\psi_{\text{sine}}\rangle = \sum_{n=0}^N \psi_n |n\rangle_{\text{un}}, \quad (\text{A.95})$$

where the coefficients are chosen with a sinusoidal profile (see, e.g., [19]), i.e.,

$$\psi_n = \sqrt{\frac{2}{N+2}} \sin\left(\frac{(n+1)\pi}{N+2}\right). \quad (\text{A.96})$$

For the sake of illustration, we will study the performance of this particular state that we will refer to as the *sine state* for the MSE and Gaussian priors of finite width. Nonetheless, it is crucial to note that the optimal probe state for phase estimation with any prior (and variance) must be of the form of equation (A.95) for some choice of coefficients. This is due to the fact that $|\psi_{\text{sine}}\rangle$ already contains one representative eigenvector of U_θ (and H) for each of its different eigenvalues. Adding any other components outside of the span of $\{|n\rangle_{\text{un}}\}_{n=0, \dots, N}$ would hence not provide any more information about the phase θ . After the unitary dynamics U_θ , the probe state is thus of the form

$$U_\theta |\psi_{\text{sine}}\rangle = e^{-iN\theta/2} \sum_{n=0}^N \psi_n e^{in\theta} |n\rangle_{\text{un}}. \quad (\text{A.97})$$

Also note that the probe state we have chosen is not symmetric with respect to the exchange of the different qubits. However, relinquishing this symmetry requirement allows us to operate in an $(N+1)$ -dimensional subspace of the total Hilbert space of dimension 2^N , which will prove to be crucial for the efficient implementation of the estimation scheme in MBQC.

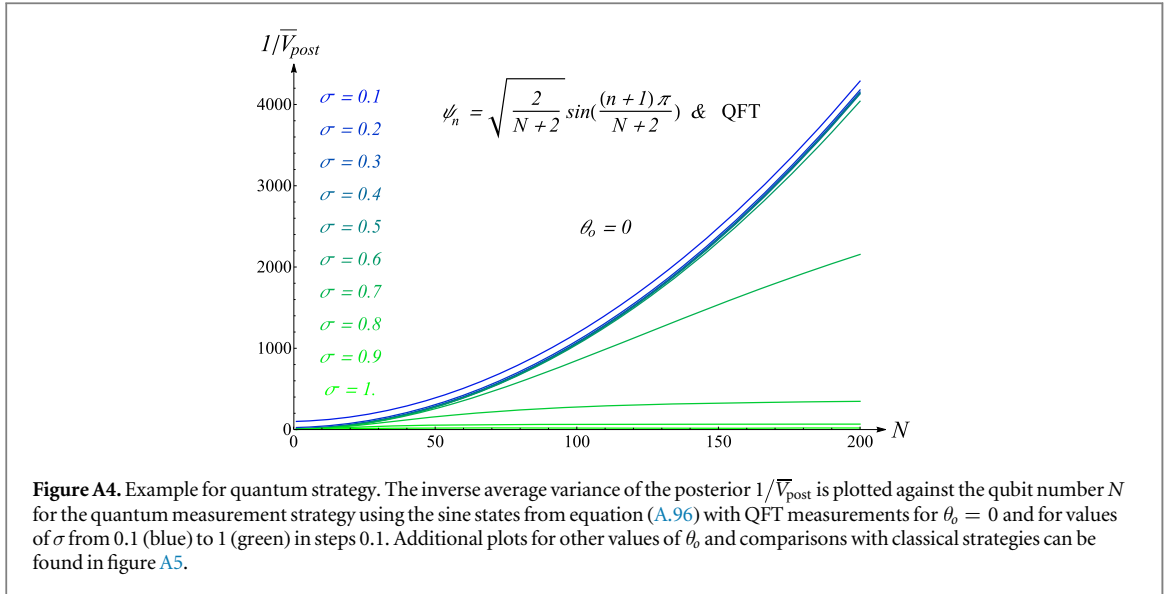
As a measurement strategy for our example, we will consider a QFT in the subspace spanned by the vectors $|n\rangle_{\text{un}}$, followed by computational basis measurements. This measurement can be represented by a POVM with elements $E_k = |e_k\rangle\langle e_k|$ for $k = 0, 1, 2, \dots, N$ and $E_{N+1} = \mathbb{1} - \sum_{k=0, \dots, N} E_k$, where

$$|e_k\rangle = \frac{1}{\sqrt{N+1}} \sum_{n=0}^N e^{in\frac{2\pi k}{N+1}} |n\rangle_{\text{un}}. \quad (\text{A.98})$$

Practically, we can ignore the POVM element E_{N+1} , as the corresponding outcome never occurs for the chosen probe state (in the absence of noise). With this, we are now in a position to compute \bar{V}_{post} from equation (A.41) where we again assume a Gaussian prior as in equation (A.42). We hence need to calculate

$$\bar{V}_{\text{post}} = \sigma^2 + \theta_o^2 - \sum_k \frac{(\text{tr}(E_k \eta))^2}{\text{tr}(E_k \Gamma)}. \quad (\text{A.99})$$

To rewrite this quantity, it is useful to first determine $\text{tr}(E_k \rho(\theta))$ where $\rho(\theta) = U_\theta |\psi_{\text{sine}}\rangle\langle \psi_{\text{sine}}| U_\theta^\dagger$, for which we obtain



$$\text{tr}(E_k \rho(\theta)) = \sum_{m,n=0}^N \frac{\psi_m^* \psi_n}{N+1} e^{i(m-n)\left(\frac{2\pi k}{N+1} - \theta\right)}. \quad (\text{A.100})$$

With this, we quickly find

$$\text{tr}(E_k \Gamma) = \sum_{m,n=0}^N \frac{\psi_m^* \psi_n}{N+1} e^{i(m-n)\left(\frac{2\pi k}{N+1} - \theta_0\right)} e^{-\frac{(m-n)^2 \sigma^2}{2}}, \quad (\text{A.101})$$

$$\text{tr}(E_k \eta) = \sum_{m,n=0}^N \frac{\psi_m^* \psi_n}{N+1} (\theta_0 - i(m-n)\sigma^2) \times e^{i(m-n)\left(\frac{2\pi k}{N+1} - \theta_0\right)} e^{-\frac{(m-n)^2 \sigma^2}{2}}. \quad (\text{A.102})$$

Inserting equations (A.101) and (A.102) back into (A.99) and making use of $\sum_k E_k = \mathbf{1}$, and the normalisation of $\rho(\theta)$ and $p(\theta)$, as well as the identity $\sum_k \exp\left(i(m-n)\frac{2\pi k}{N+1}\right) = (N+1)\delta_{mn}$, one arrives at

$$\bar{V}_{\text{post}} = \sigma^2 - \sum_k \frac{\gamma_k^2}{\text{tr}(E_k \Gamma)}, \quad (\text{A.103})$$

where γ_k is given by

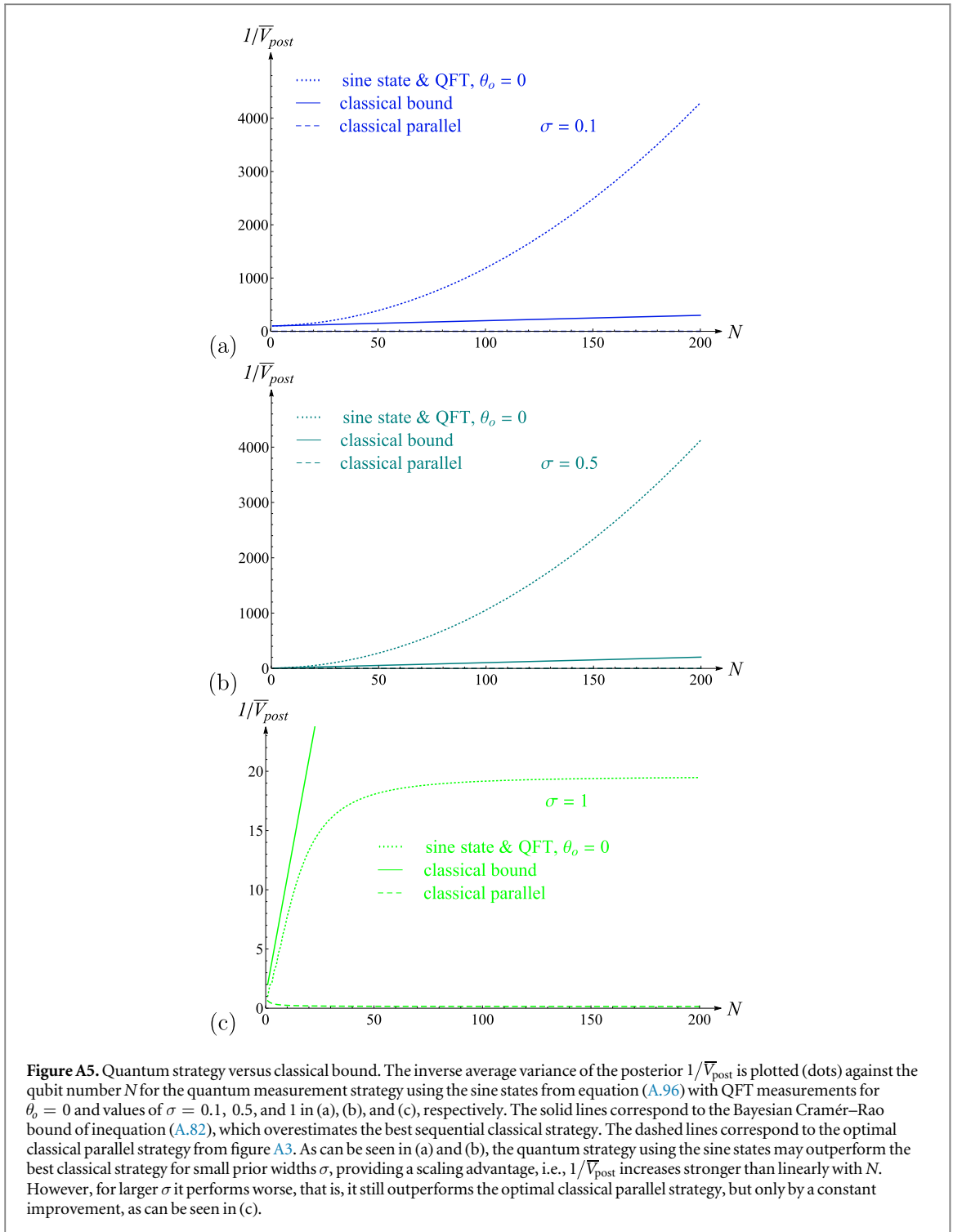
$$\gamma_k = \frac{i\sigma^2}{N+1} \sum_{m,n=0}^N \psi_m^* \psi_n (m-n) e^{i(m-n)\left(\frac{2\pi k}{N+1} - \theta_0\right)} e^{-\frac{(m-n)^2 \sigma^2}{2}}. \quad (\text{A.104})$$

Finally inserting equations (A.100) and (A.104) into the formula for \bar{V}_{post} in equation (A.103), the average variance of the posterior for the sine state and the QFT measurement can be evaluated numerically. The results for up to $N = 200$ qubits and for the prior centred at $\theta_0 = 0$ are shown in figure A4.

The plots in figure A4 indicate that for narrow priors (e.g. for $\sigma = 0.1, \dots, 0.5$) the example quantum strategy exhibits a quadratic scaling gap with respect to all classical measurements schemes, meaning that the variance in the quantum strategy decreases more strongly with N than classically possible. As discussed in [25], this is possible for all priors under certain regularity assumptions, but the explicit form of the optimal states and measurements is generally not known. Indeed, we cannot conclude that the strategy that we discuss here is optimal, but (at least) for narrow Gaussian priors ($\sigma \leq 0.5$) we find that it directly outperforms even the (overly optimistic) bound on classical strategies from inequation (A.82) already for $N = 6$ qubits. For broader priors, we can not report a scaling advantage for this example, but this is to be expected using the MSE. However, recall that the measurement strategy we discuss here is known to be optimal in the case of flat priors for an appropriately chosen cost function [19], and our results are hence complimentary in the sense that we provide numerical evidence for optimal scaling in a regime of narrow priors. Additional plots for direct comparison with the classical bounds can be found in figure A5.

A.5. Bayesian frequency estimation

In this appendix we investigate on Bayesian frequency estimation, i.e., the case where the parameter to be estimated is the angular frequency, ω , rather than the phase θ , i.e., such that $\theta = \omega t$. The key difference of frequency estimation compared to phase estimation is that in the former we have the freedom to optimise over the interrogation time t . We shall do this for some of the states and measurement previously considered for



phase estimation. Specifically, for the optimal classical parallel measurement strategy and for the quantum strategy using the sine states and QFT measurements from equations (A.96) and (A.98), respectively.

More precisely, the dynamical evolution of each qubit is described by the unitary transformation $U(\omega t) = e^{-i\omega t Z/2}$, and our prior information about ω is given by the normal distribution

$$p(\omega) = \sqrt{\frac{1}{2\pi\Delta^2}} e^{-\frac{\omega^2}{2\Delta^2}}, \quad (\text{A.105})$$

where we have assumed without loss of generality that the mean frequency is centred at $\omega_0 = 0$. With the initial state as in equation (A.95) the matrix elements of the operators Γ and η of equations (A.35) and (A.38), respectively, read

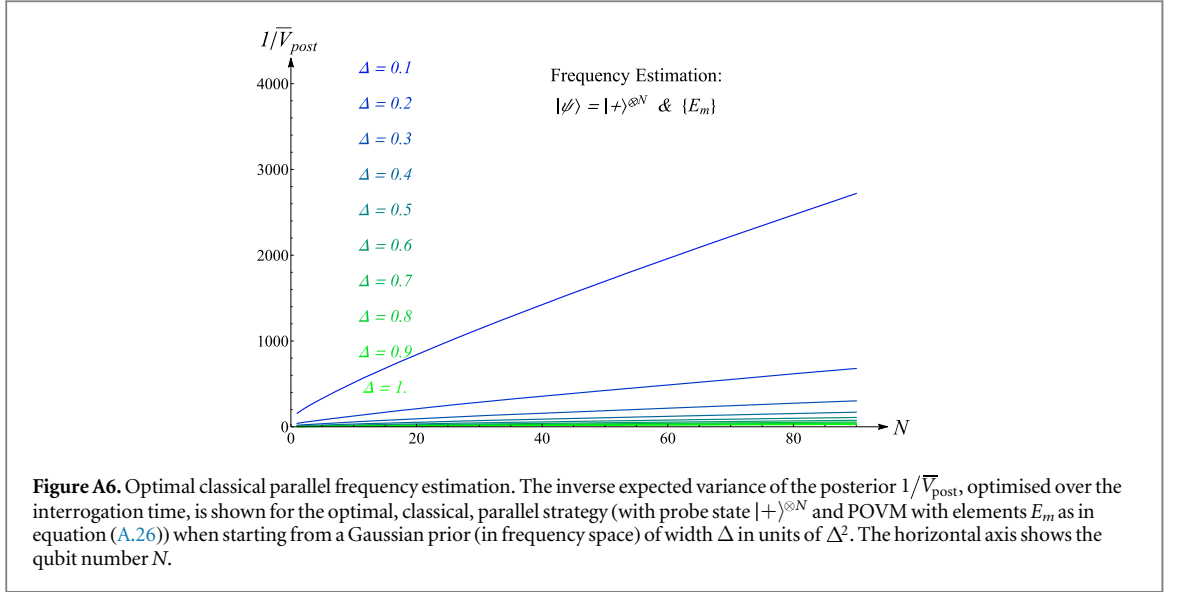


Figure A6. Optimal classical parallel frequency estimation. The inverse expected variance of the posterior $1/\overline{V}_{\text{post}}$, optimised over the interrogation time, is shown for the optimal, classical, parallel strategy (with probe state $|+\rangle^{\otimes N}$ and POVM with elements E_m as in equation (A.26)) when starting from a Gaussian prior (in frequency space) of width Δ in units of Δ^2 . The horizontal axis shows the qubit number N .

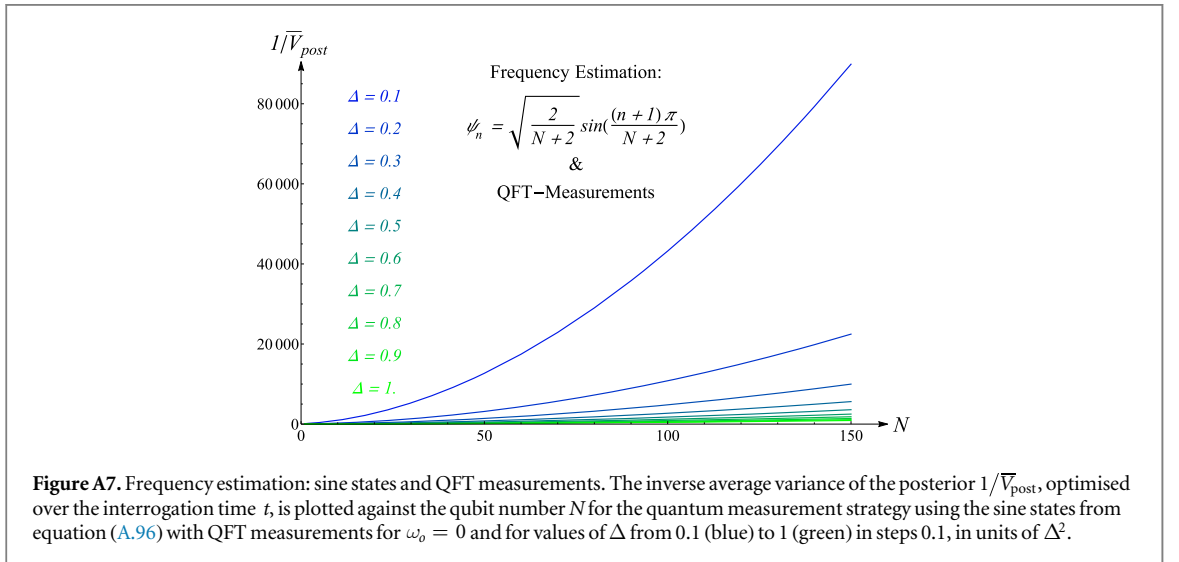


Figure A7. Frequency estimation: sine states and QFT measurements. The inverse average variance of the posterior $1/\overline{V}_{\text{post}}$, optimised over the interrogation time t , is plotted against the qubit number N for the quantum measurement strategy using the sine states from equation (A.96) with QFT measurements for $\omega_0 = 0$ and for values of Δ from 0.1 (blue) to 1 (green) in steps 0.1, in units of Δ^2 .

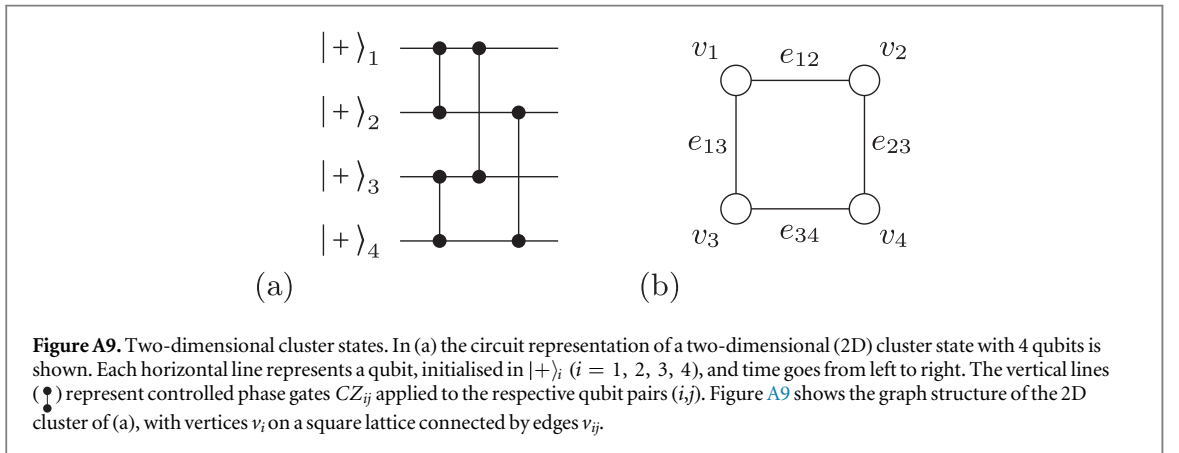
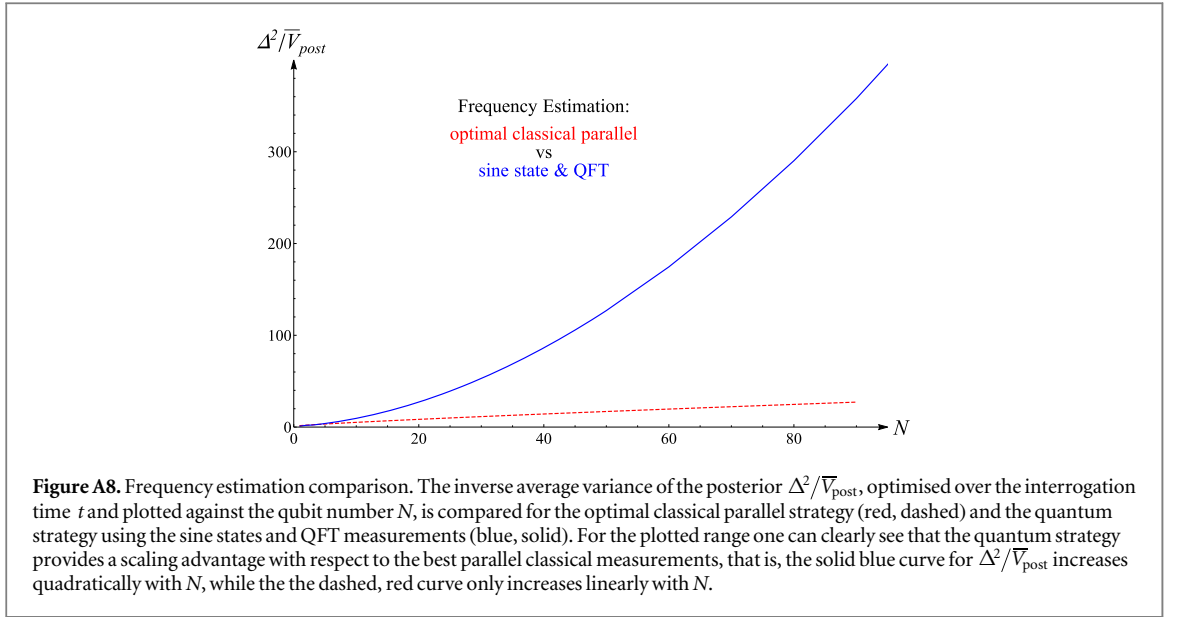
$$\Gamma = \sum_{n,m=0}^N \psi_n \psi_m^* e^{-\frac{(m-n)^2 \tau^2}{2}} |n\rangle \langle m|, \quad (\text{A.106})$$

$$\eta = -i\tau\Delta \sum_{n,m=0}^N (m-n) \psi_n \psi_m^* e^{-\frac{(m-n)^2 \tau^2}{2}} |n\rangle \langle m|, \quad (\text{A.107})$$

where we have defined the dimensionless parameter $\tau \equiv t\Delta$. The final average variance is again given by equation (A.99). However, due to the dependence of the Fisher information on τ , we need to optimise the average final variance over all τ . The results for the optimal, classical parallel strategy (see appendix A.3.3) and for the quantum strategy using the sine states and QFT measurements (see appendix A.4) are plotted in figures A6 and A7, respectively, and a comparison is shown in figure A8.

A.6. Measurement-based quantum computation

A.6.1. Basics of MBQC. In this appendix, we will briefly review the basic concepts of MBQC, but we direct the interested reader to more detailed reviews in [38, 39]. In this computational paradigm, established in [30, 80], a specific entangled state (e.g., a cluster state) is prepared in an array of qubits. Using the entanglement present in the system along with local measurements on a subset of the qubits, (arbitrary) unitary transformation may be implemented on the remaining qubits (if the cluster is large enough). Here, we will focus on MBQC based on 1D and 2D cluster states, i.e., graph states [37] based on regular, linear or rectangular lattices. Each vertex of the graph corresponds to a qubit initialised in the state $|+\rangle$, while edges connecting the vertices indicate that controlled phase gates CZ, given by



$$CZ_{ij} = |0\rangle\langle 0|_i \otimes \mathbb{1}_j + |1\rangle\langle 1|_i \otimes Z_j = CZ_{ji}, \quad (\text{A.108})$$

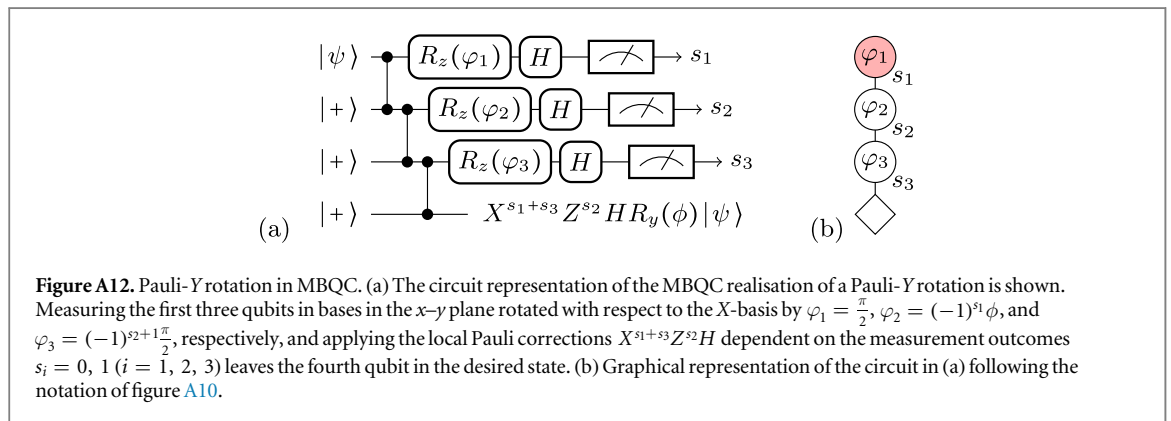
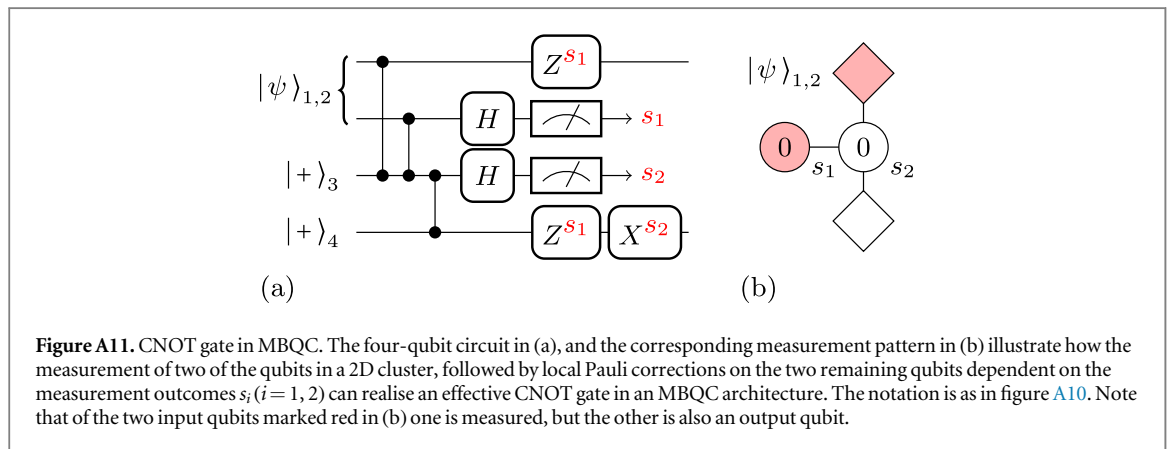
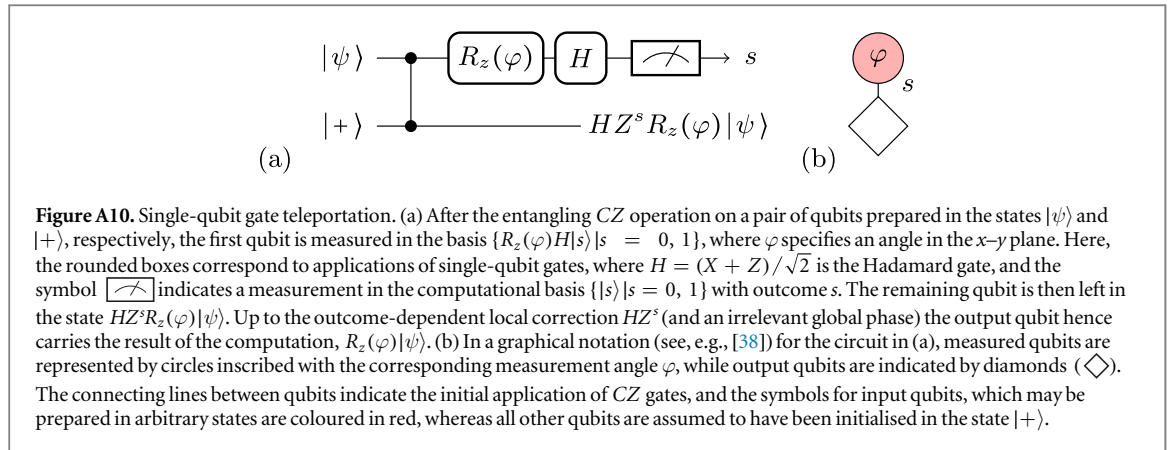
have been applied to these pairs of qubits. A simple example for a cluster state is shown in figure A9.

The essence of the working principle of a measurement-based computation is captured by single-qubit gate teleportation [81]. That is, by measuring one of the qubits of an entangled pair in a suitable local basis and applying local correction operators dependent on the outcome on the other qubit, a desired quantum gate can be effectively implemented on the remaining qubit, as illustrated in figure A10. Concatenating this procedure for a chain of qubits in a 1D cluster state, arbitrary single-qubit gates may be performed in such a way that only local corrections on the final qubit are required.

Although the measurement-based implementation of the CNOT gate [CX] in the notation of equation (A.108) is not possible in a 1D cluster, it can be achieved in two dimensions [31], as is demonstrated by a simple example in figure A11. Since the combination of arbitrary single-qubit gates with the CNOT gate is computationally universal, one may hence prepare an arbitrary quantum state (e.g., for performing parameter estimation) from a 2D cluster.

A.6.2. Probe state preparation in MBQC. In this last appendix, we present details on the conversion of the circuit for generating probe states (shown in figure 4 of the main text) to an MBQC measurement pattern. To do this, let us first see how a Y -rotation can be performed in MBQC, and consider the concatenation of three steps of single-qubit gate teleportation (see figure A10) as shown in figure A12. That is, we prepare a one-dimensional four-qubit cluster state, where the first qubit is initialised in an arbitrary state $|\psi\rangle$. The first three qubits are then measured with angles φ_1 , φ_2 , and φ_3 , respectively, leaving the fourth qubit in the state (up to a global phase)

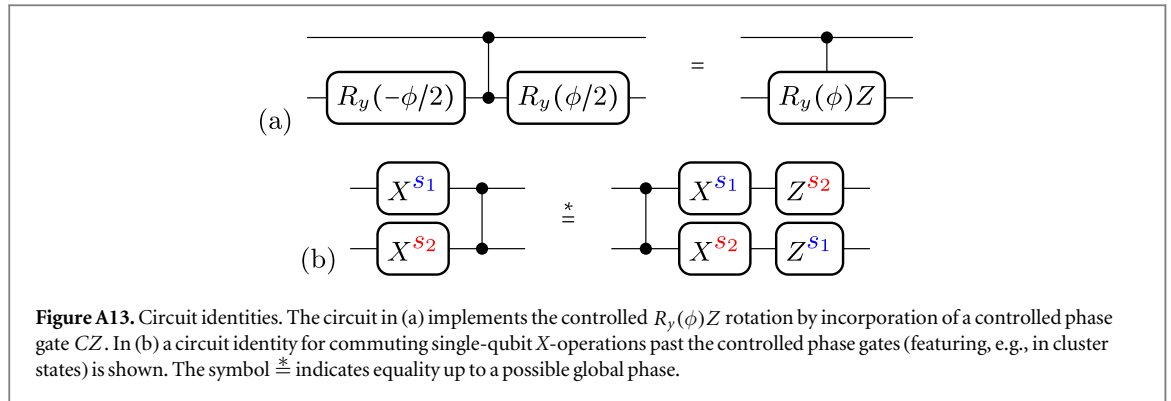
$$X^{s_1+s_3} Z^{s_2} H R_z((-1)^{s_2} \varphi_3) R_x((-1)^{s_1} \varphi_2) R_z(\varphi_1) |\psi\rangle. \quad (\text{A.109})$$



Noting that a Y-rotation about an arbitrary angle ϕ can be written as $R_y(\phi) = R_z(-\frac{\pi}{2})R_x(\phi)R_z(\frac{\pi}{2})$, selecting measurement angles $\varphi_1 = \frac{\pi}{2}$, $\varphi_3 = (-1)^{s_2+1}\frac{\pi}{2}$, and $\varphi_2 = (-1)^{s_1}\phi$ in figure A12 realises $R_y(\phi)$ up to appropriate local corrections on the last qubit.

With this strategy, we are able to implement $R_y(\phi_1)$. One may even commute the Hadamard correction with the Y-rotation to switch the initial state of the qubit from $|+\rangle$ to $|0\rangle$, as required in figure 4 of the main text. For the remaining controlled rotations, we make use of the simple identity $R_y(\phi)Z = ZR_y(-\phi)$, which allows us to utilise the CZ-gates naturally appearing in the cluster state to perform the operation $CR_y(\phi)$, as shown in the circuit in figure A13 (a). The spurious application of the operator Z before the rotation can be disregarded, since all qubits in the circuit in figure 4 are assumed to be in the state $|0\rangle$ in the beginning. This initialisation step can be included as for $R_y(\phi_1)$ before.

Since we already know from the circuit in figure A12 how to implement rotations $R_y(\phi)$ for arbitrary angles, all that is left to do to translate the preparation circuit in figure 4 to MBQC is to commute the local X-corrections past the CZ-gate appearing on the left-hand side of figure A13(a), as shown in figure A13(b), such that all local



corrections can be applied in the final step of the state preparation. We hence arrive at the MBQC measurement pattern generating the sine state $|\psi_{\text{sine}}\rangle$, which is shown in figure 5 of the main text.

ORCID

Nicolai Friis  <https://orcid.org/0000-0003-1950-8640>

Michalis Skotiniotis  <https://orcid.org/0000-0001-6935-7460>

Vedran Dunjko  <https://orcid.org/0000-0002-2632-7955>

References

- [1] Paris M G A 2009 *Int. J. Quantum. Inf.* **7** 125–37
- [2] Giovanetti V, Lloyd S and Maccone L 2011 *Nat. Photon.* **5** 222–9
- [3] Tóth G and Apellaniz I 2014 *J. Phys. A: Math. Theor.* **47** 424006
- [4] Demkowicz-Dobrzański R, Jarzyna M and Kołodyński J 2015 *Prog. Opt.* **60** 345–435
- [5] Audenaert K M R, Calsamiglia J, Masanes L, Muñoz-Tapia R, Acín A, Bagan E and Verstraete F 2007 *Phys. Rev. Lett.* **98** 160501
- [6] Audenaert K M R, Nussbaum M, Szkoła A and Verstraete F 2008 *Commun. Math. Phys.* **279** 251–83
- [7] Tsang M, Nair R and Lu X-M 2016 *Phys. Rev. X* **6** 031033
- [8] Abbott B P et al 2016 *Phys. Rev. Lett.* **116** 061102
- [9] Wasilewski W, Jensen K, Krauter H, Renema J J, Balabas M V and Polzik E S 2010 *Phys. Rev. Lett.* **104** 133601
- [10] Aiello C D, Hirose M and Cappelaro P 2013 *Nat. Commun.* **4** 1419
- [11] Mancini S and Tombesi P 2003 *Europhys. Lett.* **61** 8
- [12] Fermani R, Mancini S and Tombesi P 2003 *Phys. Rev. A* **70** 045801
- [13] Mitchell M W, Lundeen J S and Steinberg A M 2004 *Nature* **429** 161–4
- [14] Pan J-W, Chen Z-B, Lu C-Y, Weinfurter H, Zeilinger A and Zukowski M 2012 *Rev. Mod. Phys.* **84** 777–838
- [15] Wineland D J, Bollinger J J, Itano W M, Moore F L and Heinzen D J 1992 *Phys. Rev. A* **46** R6797(R)
- [16] Bollinger J J, Itano W M, Wineland D J and Heinzen J J 1996 *Phys. Rev. A* **54** R4649
- [17] Chwalla M, Kim K, Monz T, Schindler P, Riebe M, Roos C F and Blatt R 2007 *Appl. Phys. B* **89** 483–8
- [18] Giovanetti V, Lloyd S and Maccone L 2004 *Science* **306** 1330–6
- [19] Berry D W and Wiseman H M 2000 *Phys. Rev. Lett.* **85** 5098
- [20] Bagan E, Baig M and Muñoz-Tapia R 2004 *Phys. Rev. A* **70** 030301
- [21] Chiribella G, D’Ariano G M, Perinotti P and Sacchi M F 2004 *Phys. Rev. Lett.* **93** 180503
- [22] Briegel H J and Raussendorf R 2001 *Phys. Rev. Lett.* **86** 910–3
- [23] Wunderlich H, Wunderlich C, Singer K and Schmidt-Kaler F 2009 *Phys. Rev. A* **79** 052324
- [24] Skotiniotis M, Sekatski P and Dür W 2015 *New J. Phys.* **17** 073032
- [25] Jarzyna M and Demkowicz-Dobrzański R 2015 *New J. Phys.* **17** 013010
- [26] Macieszczak K, Fraas M and Demkowicz-Dobrzański R 2014 *New J. Phys.* **16** 113002
- [27] Demkowicz-Dobrzański R 2011 *Phys. Rev. A* **83** 061802(R)
- [28] Chiribella G, D’Ariano G M, Perinotti P and Sacchi M F 2004 *Phys. Rev. A* **70** 062105
- [29] Chiribella G, D’Ariano G M and Sacchi M F 2005 *Phys. Rev. A* **72** 042338
- [30] Raussendorf R and Briegel H J 2001 *Phys. Rev. Lett.* **86** 5188–91
- [31] Van den Nest M, Miyake A, Dür W and Briegel H J 2006 *Phys. Rev. Lett.* **97** 150504
- [32] Rosenkranz M and Jaksch D 2009 *Phys. Rev. A* **79** 022103
- [33] Berry D W, Higgins B L, Bartlett S D, Mitchell M W, Pryde G J and Wiseman H M 2009 *Phys. Rev. A* **80** 052114
- [34] Nielsen M A and Chuang I L 2000 *Quantum Computation and Quantum Information* (Cambridge: Cambridge University Press)
- [35] Paris M G A 2012 *Eur. Phys. J. Spec. Top.* **203** 61–86
- [36] Personick S D 1971 *IEEE Trans. Inf. Theory* **17** 240–6
- [37] Hein M, Dür W, Eisert J, Raussendorf R, Van den Nest M and Briegel H J 2006 Quantum computers, algorithms and chaos *Proc. Int. School Phys. ‘Enrico Fermi’* 162, p 115
- [38] Browne D E and Briegel H J 2006 One-way quantum computation—a tutorial introduction arXiv:quant-ph/0603226
- [39] Briegel H J, Browne D E, Dür W, Raussendorf R and Van den Nest M 2009 *Nat. Phys.* **5** 19–26
- [40] Holevo A S 1984 *Lect. Notes Math.* **1055** 153
- [41] Viola L, Knill E and Lloyd S 1998 *Phys. Rev. Lett.* **82** 2417–21

- [42] Cleve R and Watrous J 2000 *Proc. 41st IEEE Symp. on Foundations of Computer Science* pp 526–36
- [43] Takahashi Y, Kunihiro N and Ohta K 2007 *Quantum. Inf. Comput.* **7** 383–91
- [44] Vedral V, Barenco A and Ekert A 1996 *Phys. Rev. A* **54** 147
- [45] Cuccaro S A, Draper T G, Kutin S A and Moulton D P 2004 A new quantum ripple-carry addition circuit arXiv:quant-ph/0410184
- [46] Pius E 2015 Parallel quantum computing: from theory to practice *PhD Thesis* University of Edinburgh
- [47] Boixo S, Flammia S T, Caves C M and Geremia J M 2007 *Phys. Rev. Lett.* **98** 090401
- [48] Roy S M and Braunstein S L 2006 *Phys. Rev. Lett.* **100** 220501
- [49] Napolitano M, Koschorreck M, Dubost B, Behhood N, Sewell R J and Mitchell M W 2011 *Nature* **471** 486–9
- [50] Luis A 2010 *SPIE Rev.* **1** 018006
- [51] Rivas Á and Luis A 2010 *Phys. Rev. Lett.* **105** 010403
- [52] Boixo S, Datta A, Davis M J, Flammia S T, Shaji A and Caves C M 2008 *Phys. Rev. Lett.* **101** 040403
- [53] Sahota J and Quesada N 2015 *Phys. Rev. A* **91** 013808
- [54] Kok P, Dunningham J and Ralph J F 2017 *Phys. Rev. A* **95** 012326
- [55] Gaiba R and Paris M G A 2009 *Phys. Lett. A* **373** 934
- [56] Friis N, Skotiniotis M, Fuentes I and Dür W 2015 *Phys. Rev. A* **92** 022106
- [57] Sahota J, Quesada N and James D F V 2016 *Phys. Rev. A* **94** 033817
- [58] Demkowicz-Dobrzański R and Markiewicz M 2015 *Phys. Rev. A* **91** 062322
- [59] Escher B M, de Matos Filho R L and Davidovich L 2011 *Nat. Phys.* **7** 406–11
- [60] Demkowicz-Dobrzański R, Kolodyński J and Guţă M 2012 *Nat. Commun.* **3** 1063
- [61] Knysh S I, Chen E H and Durkin G A 2014 True limits to precision via unique quantum probe arXiv:1402.0495 [quant-ph]
- [62] Dür W, Skotiniotis M, Fröwis F and Kraus B 2014 *Phys. Rev. Lett.* **112** 080801
- [63] Sekatski P, Skotiniotis M and Dür W 2016 *New J. Phys.* **18** 073034
- [64] Sekatski P, Skotiniotis M, Kolodyński J and Dür W 2016 Quantum metrology with full and fast quantum control arXiv:1603.08944 [quant-ph]
- [65] Combes J, Ferrie C, Cesare C, Tiersch M, Milburn G J, Briegel H J and Caves C M 2014 *In situ* characterization of quantum devices with error correction arXiv:1405.5656 [quant-ph]
- [66] Orsucci D, Tiersch M and Briegel H J 2016 *Phys. Rev. A* **93** 042303
- [67] Pirandola S and Lupo C 2017 *Phys. Rev. Lett.* **118** 100502
- [68] Sanders B C, Milburn G J and Zhang Z 1997 *J. Mod. Opt.* **44** 1309–20
- [69] Fröwis F, Skotiniotis M, Kraus B and Dür W 2014 *New J. Phys.* **16** 083010
- [70] Skotiniotis M, Fröwis F, Dür W and Kraus B 2015 *Phys. Rev. A* **92** 022323
- [71] Knott P A 2016 *New J. Phys.* **18** 073033
- [72] Cramér H 1946 *Mathematical Methods of Statistics* (Princeton: Princeton University Press)
- [73] Frieden B R 2004 *Science from Fisher Information: A Unification* (Cambridge: Cambridge University Press)
- [74] Van Trees H L 1968 *Detection, Estimation, and Modulation Theory, Part I* (New York: Wiley)
- [75] Kay S M 1993 *Fundamentals of Statistical Signal Processing: Estimation Theory* (Englewood Cliffs, NJ: Prentice-Hall)
- [76] Braunstein S L and Caves C M 1994 *Phys. Rev. Lett.* **72** 3439–43
- [77] Liu J, Jing X X, Zhong W and Wang X G 2014 *Commun. Theor. Phys.* **61** 45–50
- [78] Gill R D and Levit B Y 1995 *Bernoulli* **1** 59–79
- [79] Berry D 2002 Adaptive phase measurements *PhD Thesis* University of Queensland arXiv:quant-ph/0202136
- [80] Raussendorf R, Browne D E and Briegel H J 2003 *Phys. Rev. A* **68** 022312
- [81] Gottesman D and Chuang I L 1999 *Nature* **402** 390–3