PAPER • OPEN ACCESS

Quantum phase estimation of multiple eigenvalues for small-scale (noisy) experiments

To cite this article: Thomas E O'Brien *et al* 2019 *New J. Phys.* **21** 023022

View the article online for updates and enhancements.

New Journal of Physics

The open access journal at the forefront of physics

PAPER

Quantum phase estimation of multiple eigenvalues for small-scale (noisy) experiments

Thomas E O'Brien[1,4], **Brian Tarasinski**[2] and **Barbara M Terhal**[2,3]

1 Instituut Lorentz, Universiteit Leiden, P.O. Box 9506, 2300 RA Leiden, The Netherlands
2 QuTech, Delft University of Technology, P.O. Box 5046, 2600 GA Delft, The Netherlands
3 JARA Institute for Quantum Information (PGI-11), Forschungszentrum Juelich, D-52425 Juelich, Germany
4 Author to whom any correspondence should be addressed.
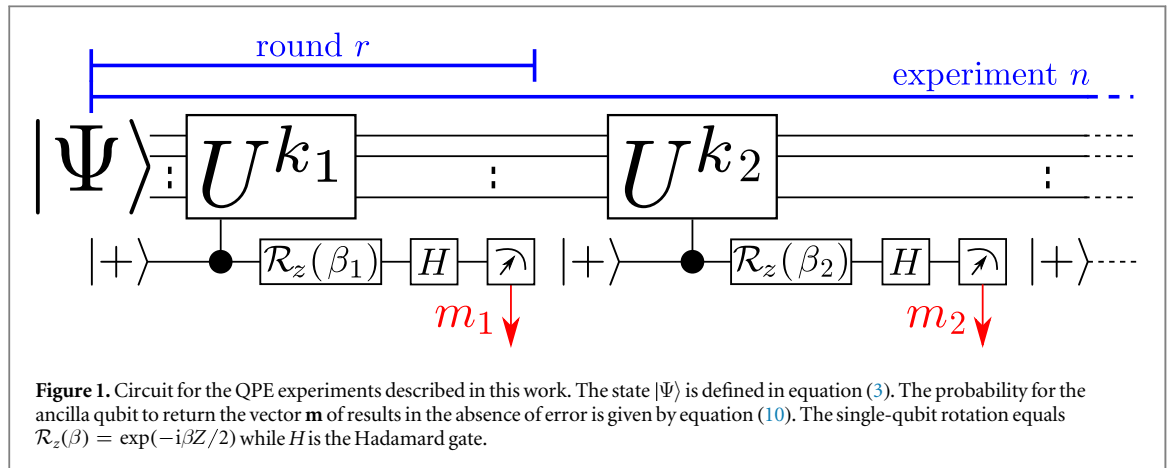
E-mail: obrien@lorentz.leidenuniv.nl

Abstract

Quantum phase estimation (QPE) is the workhorse behind any quantum algorithm and a promising method for determining ground state energies of strongly correlated quantum systems. Low-cost QPE techniques make use of circuits which only use a single ancilla qubit, requiring classical post-processing to extract eigenvalue details of the system. We investigate choices for phase estimation for a unitary matrix with low-depth noise-free or noisy circuits, varying both the phase estimation circuits themselves as well as the classical post-processing to determine the eigenvalue phases. We work in the scenario when the input state is not an eigenstate of the unitary matrix. We develop a new post-processing technique to extract eigenvalues from phase estimation data based on a classical time-series (or frequency) analysis and contrast this to an analysis via Bayesian methods. We calculate the variance in estimating single eigenvalues via the time-series analysis analytically, finding that it scales to first order in the number of experiments performed, and to first or second order (depending on the experiment design) in the circuit depth. Numerical simulations confirm this scaling for both estimators. We attempt to compensate for the noise with both classical post-processing techniques, finding good results in the presence of depolarizing noise, but smaller improvements in 9-qubit circuit-level simulations of superconducting qubits aimed at resolving the electronic ground state of a H_4-molecule.

1. Introduction

It is known that any problem efficiently solvable on a quantum computer can be formulated as eigenvalue sampling of a Hamiltonian or eigenvalue sampling of a sparse unitary matrix [1]. In this sense the algorithm of quantum phase estimation (QPE) is the only quantum algorithm which can give rise to solving problems with an exponential quantum speed-up. Despite it being such a central component of many quantum algorithms, very little work has been done so far to understand what QPE in the current noisy intermediate scale quantum (NISQ) era of quantum computing [2] where quantum devices are strongly coherence-limited. QPE comes in many variants, but a large subclass of these algorithms (e.g. the semi-classical version of textbook phase estimation [3, 4], Kitaev's phase estimation [5], Heisenberg-optimized versions [6]), are executed in an iterative sequential form using controlled-U^k gates with a single ancilla qubit [7, 8] (see figure 1), or by direct measurement of the system register itself [6]. Such circuits are of practical interest in the near term when every additional qubit requires a larger chip and brings in additional experimental complexity and incoherence.

Some of the current literature on QPE works under limiting assumptions. The first is that one does not start in an eigenstate of the Hamiltonian [9, 10]. A second limitation is that one does not take into account the (high) temporal cost of running U^k [8] for large k when optimizing phase estimation. The size and shallowness of the QPE circuit is important since, in the absence of error correction or error mitigation, one expects entropy build-up during computation. This means that circuits with large k may not be of any practical interest.

Figure 1. Circuit for the QPE experiments described in this work. The state $|\Psi\rangle$ is defined in equation (3). The probability for the ancilla qubit to return the vector **m** of results in the absence of error is given by equation (10). The single-qubit rotation equals $\mathcal{R}_z(\beta) = \exp(-\mathrm{i}\beta Z/2)$ while H is the Hadamard gate.

The scenario where the input state is not an eigenstate of the unitary matrix used in phase estimation is the most interesting one from the perspective of applications, and we will consider it in this work. Such an input state can be gradually projected onto an eigenstate by the phase estimation algorithm and the corresponding eigenvalue can be inferred. However, for coherence-limited low-depth circuits one may not be able to evolve sufficiently long to project well onto one of the eigenstates. This poses the question what one can still learn about eigenvalues using low-depth circuits. An important point is that it is experimentally feasible to repeat many relatively shallow experiments (or perform them in parallel on different machines). Hence we ask what the spectral-resolving power of such phase estimation circuits is, both in terms of the number of applications of the controlled-U circuit in a single experiment, and the number of times the experiment is repeated. Such repeated phase estimation experiments require classical post-processing of measurement outcomes, and we study two such algorithms for doing this. One is our adaptation of the Bayesian estimator of [10] to the multiple-eigenvalue scenario. A second is a new estimator based on a treatment of the observed measurements as a time-series, and construction of the resultant time-shift operator. This latter method is very natural for phase estimation, as one interprets the goal of phase estimation as the reconstruction of frequencies present in the output of a temporal sound signal. In fact, the time-series analysis that we develop is directly related to what are called Prony-like methods in the signal-processing literature, see e.g. [11]. The use of this classical method in quantum signal processing, including in quantum tomography [12], seems to hold great promise.

One can interpret our results as presenting a new hybrid classical-quantum algorithm for QPE. Namely, when the number of eigenstates in an input state is small, i.e. scaling polynomially with the number of qubits n_{sys}, the use of our classical post-processing method shows that there is no need to run a quantum algorithm which projects onto an eigenstate to learn the eigenvalues. We show that one can extract these eigenvalues efficiently by classically post-processing the data from experiments using a single-round QPE circuits (see section 2) and classically handling $\mathrm{poly}(n_{\mathrm{sys}}) \times \mathrm{poly}(n_{\mathrm{sys}})$ matrices. This constitutes a saving in the required depth of the quantum circuits.

The spectral-resolution power of QPE can be defined by its scaling with parameters of the experiment and the studied system. We are able to derive analytic scaling laws for the problem of estimating single eigenvalues with the time-series estimator. We find these to agree with the numerically-observed scaling of both studied estimators. For the more general situation, with multiple eigenvalues and experimental error, we study the error in estimating the lowest eigenvalue numerically. This is assisted by the low classical computation cost of both estimators. We observe scaling laws for this error in terms of the overlap between the ground and starting state (i.e. the input state of the circuit), the gap between the ground and excited states, and the coherence length of the system. In the presence of experimental noise we attempt to adjust our estimators to mitigate the induced estimation error. For depolarizing-type noise we find such compensation easy to come by, whilst for a realistic circuit-level simulation we find smaller improvements using similar techniques.

Even though our paper focuses on QPE where the phases corresponds to eigenvalues of a unitary matrix, our post-processing techniques may also be applicable to multi-parameter estimation problems in quantum optical settings. In these settings the focus is on determining an optical phase-shift [13–15] through an interferometric set-up. There is experimental work on (silicon) quantum photonic processors [16–18] on multiple-eigenvalue estimation for Hamiltonians which could also benefit from using the classical post-processing techniques that we develop in this paper.

2. Quantum phase estimation

QPE covers a family of quantum algorithms which measure a system register of n_{sys} qubits in the eigenbasis of a unitary operator U [5, 19]

$$U|\phi_j\rangle = e^{i\phi_j}|\phi_j\rangle, \tag{1}$$

to estimate one or many phases ϕ_j. QPE algorithms assume access to a noise free quantum circuit which implements U on our system register conditioned on the state of an ancilla qubit. Explicitly, we require the ability to implement

$$\mathcal{U}_c = |0\rangle\langle 0| \otimes \mathbb{I} + |1\rangle\langle 1| \otimes U, \tag{2}$$

where $|0\rangle$ and $|1\rangle$ are the computational basis states of the ancilla qubit, and $\mathbb{I}$ is the identity operator on the system register.

In many problems in condensed matter physics, materials science, or computational chemistry, the object of interest is the estimation of spectral properties or the lowest eigenvalue of a Hamiltonian $\mathcal{H}$. The eigenvalue estimation problem for $\mathcal{H}$ can be mapped to phase estimation for a unitary $U_\tau = \exp(-i\tau\mathcal{H})$ with a τ chosen such that the relevant part of the eigenvalue spectrum induces phases within $[-\pi, \pi]$. Much work has been devoted to determining the most efficient implementation of the (controlled)-$\exp(-i\tau\mathcal{H})$ operation, using exact or approximate methods [19–22]. Alternatively, one may simulate $\mathcal{H}$ via a quantum walk, mapping the problem to phase estimating the unitary $\exp(-i\arcsin(\mathcal{H})/\lambda)$ for some λ, which may be implemented exactly [23–26]. In this work we do not consider such variations, but rather focus on the error in estimating the eigenvalue phases of the unitary U that is actually implemented on the quantum computer. In particular, we focus on the problem of determining the value of a single phase ϕ_0 to high precision (this phase could correspond, for example, to the ground state energy of some Hamiltonian $\mathcal{H}$).

Phase estimation requires the ability to prepare an input, or *starting state*

$$|\Psi\rangle = \sum_j a_j|\phi_j\rangle, \, A_j \equiv |a_j|^2, \tag{3}$$

with good overlap with the ground state; $A_0 \gg 0$. Note here that the spectrum of U may have exact degeneracies (e.g. those enforced by symmetry) which phase estimation does not distinguish; we count degenerate eigenvalues as a single ϕ_j throughout this work. The ability to start QPE in a state which already has good overlap with the ground state is a non-trivial requirement for the applicability of the QPE algorithm. On the other hand, it is a well-known necessity given the QMA-completeness [27] of the lowest eigenvalue problem[5]. For many quantum chemistry and materials science problems it is known or expected that the Hartree–Fock state has good overlap with the ground state, although rigorous results beyond perturbation theory are far and few between (see e.g. [28]). Beyond this, either adiabatic evolution [20, 29] or variational quantum eigensolvers [30] can provide an approximate starting state to improve on via phase estimation.

Phase estimation is not limited to simply learning the value of ϕ_0; it may obtain information about all phases ϕ_j as long as $A_j > 0$. However, the resources required to estimate ϕ_j are bounded below by $1/A_j$. To see this, note that the controlled-unitary $\mathcal{U}_c$ does not mix eigenstates, and so there is no difference (in the absence of error) between starting with $|\Psi\rangle$ and the mixed state

$$\rho_\Psi = \sum_j A_j|\phi_j\rangle\langle\phi_j|. \tag{4}$$

The latter is then equivalent to preparing the pure state $|\phi_j\rangle$ with probability A_j, so if N preparations of $|\phi_j\rangle$ are required to estimate ϕ_j to an error ϵ, the same error margin requires at least N/A_j preparations of the state $|\Psi\rangle$. As the number of eigenstates N_{eig} with non-zero contribution to $|\Psi\rangle$ generally scales exponentially with the system size n_{sys}, estimating more than the first few ϕ_j (ordered by the magnitude A_j) will be unfeasible.

Low-cost (in terms of number of qubits) QPE may be performed by entangling the system register with a single ancilla qubit [5, 8, 10, 27]. In figure 1, we give the general form of the quantum circuit to be used throughout this paper. An experiment, labeled by a number $n = 1, \ldots, N$, can be split into one or multiple rounds $r = 1, \ldots, R_n$, following the preparation of the starting state $|\Psi\rangle$. In each round a single ancilla qubit prepared in the $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ state controls $\mathcal{U}_c^{k_r}$ where the integer k_r can vary per round. The ancilla qubit is then rotated by $\mathcal{R}_z(\beta_r) = \exp(-i\beta_r Z/2)$ (with the phase β_r possibly depending on other rounds in the same experiment) and read out in the X-basis, returning a measurement outcome $m_r \in \{0, 1\}$. We denote the

---

[5] QMA stands for Quantum Merlin Arthur, which is a complexity class which contains decision problems which are easy to verify on a quantum computer, though not necessarily easy to solve. This class is the natural quantum counterpart to the complexity class NP of problems that may be verified easily on a classical computer. A QMA-complete problem is one of the 'hardest possible' such problems (in analogy with NP-complete problems); the ability to solve these problems in polynomial time would allow polynomial-time solving of any other problem in QMA.

chosen strings of integers and phases of a single multi-round experiment by **k** and $\boldsymbol{\beta}$, respectively. We denote the number of controlled-U iterations per experiment as $K = \sum_{r=1}^{R_n} k_r$. We denote the total number of controlled-U iterations over all experiments as

$$K_{\text{tot}} = \sum_{n=1}^{N} \sum_{r=1}^{R_n} k_r. \tag{5}$$

As the system register is held in memory during the entire time of the experiment, the choice of K is dictated by the coherence time of the underlying quantum hardware. Hence, we introduce a dimensionless coherence length

$$K_{\text{err}} = \frac{T_{\text{err}}}{n_{\text{sys}} T_U}. \tag{6}$$

Here T_U is the time required to implement a single application of controlled-U in equation (7), and T_{err} is the time-to-error of a single qubit, so that $T_{\text{err}}/n_{\text{sys}}$ is the time-to-failure of n_{sys} qubits. The idea is that K_{err} bounds the maximal number of applications of U in an experiment, namely $K \leqslant K_{\text{err}}$.

A new experiment starts with the same starting state $|\Psi\rangle$. Values of k_r and β_r may be chosen independently for separate experiments n, i.e. we drop the label n for convenience. We further drop the subscript r from single-round experiments (with $R = 1$).

In the absence of error, one may calculate the action of the QPE circuit on the starting state (defined in equation (3)). Working in the eigenbasis of U on the system register, and the computational basis on the ancilla qubit, we calculate the state following the controlled-rotation $\mathcal{U}_c^{k_1}$, and the rotation $\mathcal{R}_z(\beta_1)$ on the ancilla qubit to be

$$\frac{1}{\sqrt{2}} \sum_j a_j (|0\rangle + e^{i(k_1\phi_j + \beta_1)}|1\rangle)|\phi_j\rangle. \tag{7}$$

The probability to measure the ancilla qubit in the X-basis as $m_1 \in \{0, 1\}$ is then

$$\sum_j A_j \cos^2\left(\frac{k_1\phi_j}{2} + \frac{\beta_1 - m_1\pi}{2}\right), \tag{8}$$

and the unnormalized post-selected state of the system register is

$$\sum_j a_j e^{\frac{i}{2}(k_1\phi_j + \beta_1)} \cos\left(\frac{k_1\phi_j}{2} + \frac{\beta_1 - m_1\pi}{2}\right)|\phi_j\rangle. \tag{9}$$

The above procedure may then be repeated for r rounds to obtain the probability of a string **m** of measurement outcomes of one experiment as

$$P_{\mathbf{k},\boldsymbol{\beta}}(\mathbf{m}|\boldsymbol{\phi}, \mathbf{A}) = \sum_j A_j \prod_{r=1}^{R} \cos^2\left(\frac{k_r\phi_j}{2} + \frac{\beta_r - m_r\pi}{2}\right). \tag{10}$$

Here, $\boldsymbol{\phi}$ is the vector of phases ϕ_j and $\mathbf{A}$ the vector of probabilities for different eigenstates. We note that $P_{\mathbf{k},\boldsymbol{\beta}}(\mathbf{m}|\boldsymbol{\phi}, \mathbf{A})$ is independent of the order in which the rounds occur in the experiment. Furthermore, when $N_{\text{eig}} = 1$, $P_{\mathbf{k},\boldsymbol{\beta}}(\mathbf{m}|\boldsymbol{\phi}) = P_{\mathbf{k},\boldsymbol{\beta}}(\mathbf{m}|\boldsymbol{\phi}, \mathbf{A})$ is equal to the product of the single-round probabilities $P_{k_r,\beta_r}(m_r|\boldsymbol{\phi})$, as there is no difference between a multi-round experiment and the same rounds repeated across individual experiments.

One can make a direct connection with parameter estimation work by considering the single-round experiment scenario in figure 1. The Hadamard gate putting the ancilla qubit in $|+\rangle$ and measuring the qubit in the X-basis are, in the optical setting, realized by beam-splitters, so that only the path denoted by the state $|1\rangle$ will pick up an unknown phase-shift. When the induced phase-shift is not unique but depends, say, on the state of another quantum system, we may like to estimate all such possible phases corresponding to our scenario of wishing to estimate multiple eigenvalues. Another physical example is a dispersively coupled qubit-cavity mode system where the cavity mode occupation number will determine the phase accumulation of the coupled qubit [31].

3. Classical data analysis

Two challenges are present in determining ϕ_0 from QPE experiments. First, we only ever have inexact sampling knowledge of $P_{\mathbf{k},\boldsymbol{\beta}}(\mathbf{m}|\boldsymbol{\phi}, \mathbf{A})$. That is, repeated experiments at fixed **k**, $\boldsymbol{\beta}$ do not directly determine $P_{\mathbf{k},\boldsymbol{\beta}}(\mathbf{m}|\boldsymbol{\phi}, \mathbf{A})$, but rather sample from the multinomial distribution $P_{\mathbf{k},\boldsymbol{\beta}}(\mathbf{m}|\boldsymbol{\phi}, \mathbf{A})$. From the measurement outcomes we can try to estimate $P_{\mathbf{k},\boldsymbol{\beta}}(\mathbf{m}|\boldsymbol{\phi}, \mathbf{A})$ (and from this ϕ_0) as a hidden variable. Secondly, when $N_{\text{eig}} > 1$ determining ϕ_0 from $P_{\mathbf{k},\boldsymbol{\beta}}(\mathbf{m}|\boldsymbol{\phi}, \mathbf{A})$ poses a non-trivial problem.

Let us first consider the case $N_{\text{eig}} = 1$. Let us assume that we do single-round experiments with a fixed k for each experiment. Naturally, taking $k = 1$ would give rise to the lowest-depth experiments. If we start these experiments with $k = 1$ in the eigenstate $|\phi_0\rangle$, then one can easily prove that taking $\beta = 0$ or $\beta = \frac{\pi}{2}$ for half of the experiments, suffices to estimate ϕ_0 with variance scaling as $1/N = 1/K_{\text{tot}}$. This result can be derived using standard Chernoff

bounds, see e.g. [32, 33], and represent standard sampling or shot noise behavior. When $N_{\text{eig}} = 1$, NK-round experiments each with $k = 1$ are indistinguishable from $N \times K$ single-round experiments with $k = 1$. This implies that the same scaling holds for such multi-round experiments, i.e. the variance scales as $1/(NK) = 1/K_{\text{tot}}$.

Once the phase ϕ_0 is known to sufficient accuracy, performing QPE experiments with $k > 1$ is instrumental in resolving ϕ_0 in more detail, since the probability of a single-round outcome depends on $k\phi_0$ [6]. Once one knows with sufficient certainty that $\phi_0 \in [(2m - 1)\pi/k, (2m + 1)\pi/k)$ (for integer m), one can achieve variance scaling as $1/k^2 N$ (conforming to so-called local estimation Cramer–Rao bounds suggested in [10, 34]). A method achieving Heisenberg scaling, where the variance scales as $1/K_{\text{tot}}^2$ (see equation (5)), was analyzed in [6, 32]. This QPE method can also be compared with the information-theoretic optimal maximum-likelihood phase estimation method of [8] where $N \sim \log K$ experiments are performed, each choosing a random $k \in \{1, \ldots, K\}$ to resolve ϕ_0 with error scaling as $1/K$. The upshot of these previous results is that, while the variance scaling in terms of the total number of unitaries goes like $1/K_{\text{tot}}$ when using $k = 1$, clever usage of $k > 1$ data can lead to $1/K_{\text{tot}}^2$ scaling. However, as K is limited by K_{err} in near-term experiments, this optimal Heisenberg scaling may not be accessible.

When $N_{\text{eig}} > 1$, the above challenge is complicated by the need to resolve the phase ϕ_0 from the other ϕ_j. This is analogous to the problem of resolving a single note from a chord. Repeated single-round experiments at fixed k and varying β can only give information about the value of the function:

$$g(k) = \sum_j A_j e^{ik\phi_j}, \tag{11}$$

at this fixed k, since

$$P_{k,\beta}(m|\phi) = \frac{1}{2} + \frac{1}{2}\cos(\beta + m\pi)\text{Re}[g(k)]$$
$$- \frac{1}{2}\sin(\beta + m\pi)\text{Im}[g(k)]. \tag{12}$$

This implies that information from single-round experiments at fixed k is insufficient to resolve ϕ_0 when $N_{\text{eig}} > 1$, as $g(k)$ is then not an invertible function of ϕ_0 (try to recover a frequency from a sound signal at a single point in time!). In general, for multi-round experiments using a maximum of K total applications of $\mathcal{U}_c$, we may only ever recover $g(k)$ for $k \leqslant K$. This can be seen from expanding $P_{\mathbf{k},\boldsymbol{\beta}}(\mathbf{m}|\phi, \mathbf{A})$ as a sum of $\sum_j A_j \cos^m(\phi_j)\sin^n(\phi_j)$ terms with $m + n \leqslant K$, which are in turn linear combinations of $g(k)$ for $k \leqslant K$. As we will show explicitly in the next section 3.1 this allows us to recover up to K ϕ_j. However, when $N_{\text{eig}} > K$, these arguments imply that we cannot recover any phases exactly. In this case, the accuracy to which we can estimate our target ϕ_0 is determined by the magnitude of the amplitude A_0 in the inital state $|\Psi\rangle$ as well as the gap towards the other eigenvalues. For example, in the limit $A_0 \to 1$, an unbiased estimation of ϕ_0 using data from $k = 1$ would be

$$\text{Arg}[g(1)] = \text{Im}\left[\ln\left(\sum_j A_j e^{i\phi_j}\right)\right], \tag{13}$$

and the error in such estimation is

$$|\text{Arg}[g(1)] - \phi_0| = \left| \frac{1}{A_0} \sum_{j=1}^{N_{\text{eig}}-1} A_j \sin(\phi_j - \phi_0) + O(A_0^{-2}) \right|$$
$$\leqslant \frac{1 - A_0}{A_0},$$

with our bound being independent of N_{eig}. We are unable to extend this analysis beyond the $k = 1$ scenario, and instead we study the scaling in this estimation numerically in section 4. In the remainder of this section, we present two estimators for multi-round QPE. The first is an estimator based on a time-series analysis of the function $g(k)$ using Prony-like [11] methods that has a low computation overhead. The second is a Bayesian estimator similar to that of [10], but adapted for multiple eigenphases ϕ_j.

3.1. Time-series analysis
Let us assume that the function $g(k)$ in equation (11) is a well-estimated function at all points $0 \leqslant k \leqslant K$, since the number of experiments N is sufficiently large. We may extend this function to all points $-K \leqslant k \leqslant K$ using the identity $g(-k) = g^*(k)$ to obtain a longer signal[6]. We wish to determine the dominant frequencies ϕ_j in the signal $g(k)$ as a function of 'time' k. This can be done by constructing and diagonalizing a time-shift matrix $\mathfrak{T}$ whose eigenvalues are the relevant frequencies in the signal, as follows.

---

[6] Extending $g(k)$ from $0 \leqslant k \leqslant K$ to $-K \leqslant k \leqslant K$ is not required to perform a time-series analysis, however numerically we observe that this obtains up to order of magnitude improvement in estimating ϕ_0.

We first demonstrate the existence of the time-shift matrix $\mathfrak{T}$ in the presence of $N_{\text{eig}} < K$ separate frequencies. Since we may not know N_{eig}, let us first estimate it as l. We then define the vectors $\mathbf{g}(k) = (g(k), g(k + 1), \dots g(k + l))^T$, $k = -K, \dots, K$. These vectors can be decomposed in terms of single-frequency vectors $\mathbf{b}_j = (1, e^{i\phi_j}, \dots, e^{il\phi_j})^T$

$$\mathbf{g}(k) = \sum_j A_j e^{ik\phi_j} \mathbf{b}_j. \tag{14}$$

We can make a $l \times N_{\text{eig}}$ matrix B with the components $\mathbf{b}_j$ as columns

$$B_{k,j} = e^{ik\phi_j}. \tag{15}$$

When $N_{\text{eig}} \leqslant l$, the columns of B are typically linearly independent[7], hence the non-square matrix B is invertible and has a (left)-pseudoinverse B^{-1} such that $B^{-1}B = \mathbf{1}$. Note however, when $N_{\text{eig}} > l$ the columns of B are linearly-dependent, so B cannot be inverted. If B is invertible, we can construct the shift matrix $\mathfrak{T} = BDB^{-1}$ with $D_{i,j} = \delta_{i,j}e^{i\phi_j}$. By construction, $\mathfrak{T}\mathbf{b}_j = e^{i\phi_j}\mathbf{b}_j$ (as $\mathfrak{T}B = BD$), and thus

$$\begin{aligned} \mathfrak{T}\mathbf{g}(k) &= \sum_j A_j e^{ik\phi_j} \mathfrak{T}\mathbf{b}_j \\ &= \sum_j A_j e^{i(k+1)\phi_j} = \mathbf{g}(k + 1). \end{aligned} \tag{16}$$

This implies that $\mathfrak{T}$ acts as the time-shift operator mapping $\mathbf{g}(k)$ to $\mathbf{g}(k + 1)$. As the eigenvalues of $\mathfrak{T}$ are precisely the required phases $e^{i\phi_j}$ in case $N_{\text{eig}} \leqslant l$, constructing and diagonalizing $\mathfrak{T}$ will obtain our desired phases including ϕ_0. When $N_{\text{eig}} > l$, the eigen-equation for $\mathfrak{T}$ cannot have the solution $\mathbf{b}_j$ since these are not linearly independent.

The above proof of existence does not give a method of constructing the time-shift operator $\mathfrak{T}$, as we do not have access to the matrices B or D. To construct $\mathfrak{T}$ from the data that we do have access to, we construct the $l \times (2K + 1 - l)$ Hankel matrices $G^{(0)}, G^{(1)}$ by

$$G^{(a)}_{i,j} = g(i + j + a - K), \tag{17}$$

indexing $0 \leqslant i \leqslant l - 1$, $0 \leqslant j \leqslant 2K - l$. The kth column of $G^{(a)}$ is the vector $\mathbf{g}(k + a - K)$, and so $\mathfrak{T}G^{(0)} = G^{(1)}$. We can thus attempt to find $\mathfrak{T}$ as a solution of the (least-squares) problem of minimizing $\|\mathfrak{T}G^{(0)} - G^{(1)}\|$. The rank of the obtained $\tilde{\mathfrak{T}}$ is bounded by the rank of $G^{(0)}$. We have that $\text{rank}(G^{(0)})$ is at most N_{eig} since it is a sum over rank-1 matrices. At the same time $\text{rank}(G^{(0)}) \leqslant \min(l, 2K + 1 - l)$. This implies that we require both $l \geqslant N_{\text{eig}}$ and $2K + 1 - l \geqslant N_{\text{eig}}$ to obtain a shift matrix $\mathfrak{T}$ with N_{eig} eigenvalues. This is only possible when $K \geqslant N_{\text{eig}}$, giving an upper bound for the number of frequencies obtainable. When $G^{(0)}$ is not full rank (because $N_{\text{eig}} < l$), this problem may have multiple zeros $\tilde{\mathfrak{T}}$. However, when $N_{\text{eig}} < l$ each of these must satisfy $\tilde{\mathfrak{T}}\mathbf{g}(k) = \mathbf{g}(k + 1)$ for $-K < k < K - l$.

Then, as long as $\text{rank}(G^{(0)}) \geqslant N_{\text{eig}}$, equation (14) is invertible by an operator C

$$\sum_k C_{i,k} A_j e^{ik\phi_j} = \delta_{i,j} \rightarrow \mathbf{b}_j = \sum_k C_{j,k} \mathbf{g}(k). \tag{18}$$

It follows that

$$\sum_k C_{j,k} \mathbf{g}(k + 1) = \sum_{k,l} C_{j,k} A_l e^{ik\phi_l}(e^{i\phi_l}\mathbf{b}_l) = e^{i\phi_j}\mathbf{b}_j, \tag{19}$$

and then

$$\tilde{\mathfrak{T}}\mathbf{b}_j = \sum_k C_{k,j}\tilde{\mathfrak{T}}\mathbf{g}(k) = \sum_k C_{k,j}\mathbf{g}(k + 1) = e^{i\phi_j}\mathbf{b}_j, \tag{20}$$
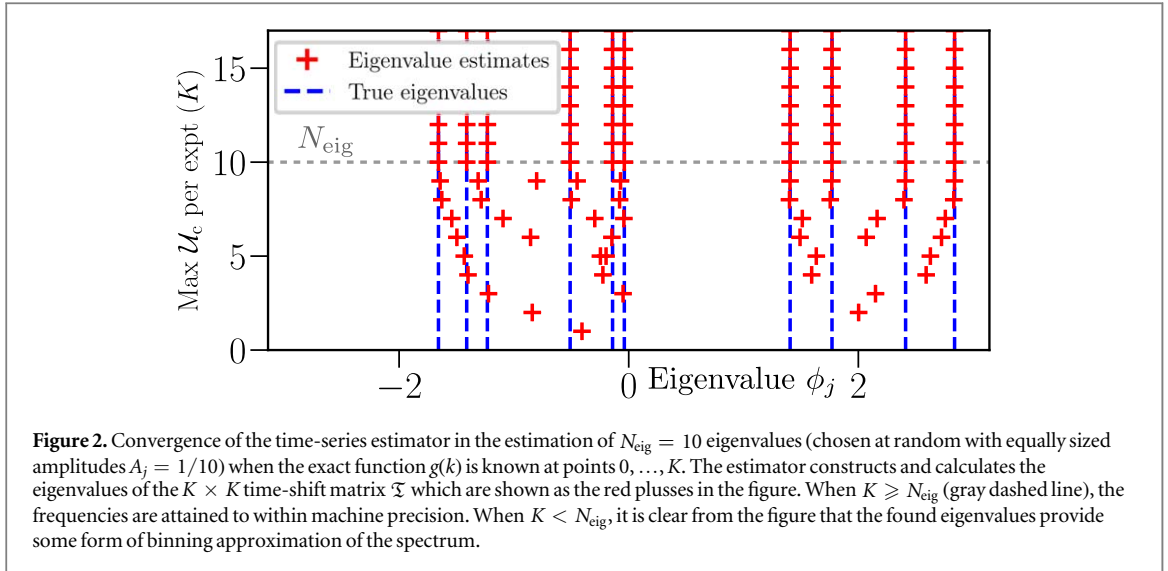
so every $\tilde{\mathfrak{T}}$ obtained in this way must have eigenvalues $e^{i\phi_j}$.

The above analysis is completely independent of the coefficients A_j. However, once the eigenvalues ϕ_j are known, the matrix B (equation (15)) may be constructed, and the A_j may be recovered by a subsequent least-squares minimization of

$$\|B\mathbf{A} - \mathbf{g}(0)\|. \tag{21}$$

This allows us to identify spurious eigenvalues if $l > N_{\text{eig}}$ (as these will have a corresponding zero amplitude). Numerically, we find no disadvantage to then choosing the largest l permitted by our data, namely $l = K$.

---

[7] Counterexamples may exist, but are hard to construct and have not occurred in any numerics.

Figure 2. Convergence of the time-series estimator in the estimation of $N_{eig} = 10$ eigenvalues (chosen at random with equally sized amplitudes $A_j = 1/10$) when the exact function $g(k)$ is known at points $0, \dots, K$. The estimator constructs and calculates the eigenvalues of the $K \times K$ time-shift matrix $\tilde{\mathfrak{T}}$ which are shown as the red plusses in the figure. When $K \geqslant N_{eig}$ (gray dashed line), the frequencies are attained to within machine precision. When $K < N_{eig}$, it is clear from the figure that the found eigenvalues provide some form of binning approximation of the spectrum.

Assuming a sufficient number of repetitions N these arguments imply that this strategy requires that $K \geqslant N_{eig}$ to determine all eigenvalues accurately. However, when $K < N_{eig}$ there still exists a least-squares solution $\tilde{\mathfrak{T}}$ that minimizes $||\tilde{\mathfrak{T}}G^{(0)} - G^{(1)}||$. When $A_0 \gg 1/K$, we expect that $\tilde{\mathfrak{T}}$ should have eigenvalues $e^{i\tilde{\phi}_0} \approx e^{i\phi_0}$ that we can take as the estimator for ϕ_0; the same is true for any other ϕ_j with sufficiently large A_j. In figure 2 we show an example of convergence of this estimation for multiple eigenvalues ϕ_j as $K \to N_{eig}$ in the case where $g(k)$ is known precisely (i.e. in the absence of sampling noise). The error $|\tilde{\phi}_0 - \phi_0|$ when $K < N_{eig}$ depends on the eigenvalue gap above ϕ_0, as well as the relative weights A_j, as we will see in section 4.3.

In appendix B we derive what variance can be obtained with this time-series method in the case $l = N_{eig} = 1$, using single-round circuits with $k = 1$ up to K. Our analysis leads to the following scaling in N and K:

$$\text{Var}(\phi) \propto \frac{1}{K^2 N}. \tag{22}$$

We will compare these results to numerical simulations in section 4.1.

3.1.1. Estimating $g(k)$

The function $g(k)$ cannot be estimated directly from experiments, but may instead be created as a linear combination of $P_{\mathbf{k},\beta}(\mathbf{m}|\phi, \mathbf{A})$ for different values of $\mathbf{k}$ and β. For single-round experiments, this combination is simple to construct:

$$\begin{aligned} g(k) = &P_{k,0}(0|\phi, \mathbf{A}) - P_{k,0}(1|\phi, \mathbf{A}) \\ &- iP_{k,\frac{\pi}{2}}(0|\phi, \mathbf{A}) + iP_{k,\frac{\pi}{2}}(1|\phi, \mathbf{A}). \end{aligned} \tag{23}$$

For multi-round experiments, the combination is more complicated. In general, $P_{\mathbf{k},\beta}(\mathbf{m}|\phi, \mathbf{A})$ is a linear combination of real and imaginary parts of $g(l)$ with $l < K = \sum_r k_r$. This combination may be constructed by writing $\cos^2(k\phi_j/2 + \beta/2)$ and $\sin^2(k\phi_j/2 + \beta/2)$ in terms of exponentials, and expanding. However, inverting this linear equation is a difficult task and subject to numerical imprecision. For some fixed choices of experiments, it is possible to provide an explicit expansion. Here we focus on K-round $k = 1$ experiments with $K/2\ \beta = 0$ and $K/2\ \beta = \frac{\pi}{2}$ final rotations during each experiment (choosing K even). The formula for $P_{\mathbf{k},\beta}(\mathbf{m}|\phi, \mathbf{A})$ is independent of the order in which these rounds occur. Let us write $\mathbb{P}(\mathfrak{m}, \mathfrak{n}|\phi, \mathbf{A})$ as the probability of seeing both $\mathfrak{m} \in \{0, \dots, K/2\}$ outcomes with $m_r = 1$ in the $K/2$ rounds with $\beta_r = 0$ and $\mathfrak{n} \in \{0, \dots, K/2\}$ outcomes with $n_r = 1$ in the $K/2$ rounds with $\beta_r = \pi/2$. In other words, $\mathfrak{m}, \mathfrak{n}$ are the Hamming weights of the measurement vectors split into the two types of rounds described above. Then, one can prove that, for $0 \leqslant k \leqslant K/2$:

$$g(k) = \sum_{m=0}^{K/2} \sum_{n=0}^{K/2} \chi_k(\mathfrak{m}, \mathfrak{n}) \mathbb{P}(\mathfrak{m}, \mathfrak{n}|\phi, \mathbf{A}), \tag{24}$$

where

$$\chi_k(\mathfrak{m}, \mathfrak{n}) = \sum_{l=0}^{k} (-i)^{k-l} \binom{k}{l}$$

$$\times \left[\sum_{p_1=0}^{\lfloor l/2 \rfloor} \frac{\binom{\mathfrak{m}}{2p_1}\binom{K/2-\mathfrak{m}}{l-2p_1}}{\binom{K/2}{l}} - 1 \right]$$

$$\times \left[\sum_{p_2=0}^{\lfloor (k-l)/2 \rfloor} \frac{\binom{\mathfrak{n}}{2p_2}\binom{K/2-\mathfrak{n}}{k-l-2p_2}}{\binom{K/2}{k-l}} - 1 \right]. \tag{25}$$

The proof of this equality can be found in appendix A.

Calculating $g(k)$ from multi-round ($k = 1$) experiments contains an additional cost: combinatorial factors in equation (24) relate the variance in $g(k)$ to the variance in $\mathbb{P}(\mathfrak{m}, \mathfrak{n}|\phi, \mathbf{A})$ but the combinatorial pre-factor $\binom{k}{l}$ can increase exponentially in k. This can be accounted for by replacing the least squares fit used above with a weighted least squares fit, so that one effectively relies less on the correctness of $g(k)$ for large k. To do this, we construct the matrix $\mathfrak{T}$ row-wise from the rows $\mathbf{g}_i^{(1)}$ of $G^{(1)}$. That is, for the ith row $\mathbf{t}_i$ we minimize

$$||\mathbf{t}_i G^{(0)} - \mathbf{g}_i^{(1)}||. \tag{26}$$

This equation may be weighted by multiplying $G^{(0)}$ and $g_i^{(1)}$ by the weight matrix

$$w_{j,k}^{(i)} = \delta_{j,k} \frac{1}{\sigma_{G_{i,j}^{(1)}}}, \tag{27}$$

where $\sigma_{G_{i,j}^{(1)}}$ is the standard deviation in our estimate of $G_{i,j}^{(1)}$. Note that the method of weighted least-squares is only designed to account for error in the independent variable of a least squares fit, in our case this is $G^{(1)}$. This enhanced effect of the sampling error makes the time-series analysis unstable for large K. We can analyze how this weighting alters the previous variance analysis when $N_{\mathrm{eig}} = 1$. If we take this into account (see derivation in appendix B), we find that

$$\mathrm{Var}(\phi) \propto \frac{1}{KN}, \tag{28}$$

for a time-series analysis applied to multi-round $k = 1$ experiments.

3.1.2. Classical computation cost

In practice, the time-series analysis can be split into three calculations; (1) estimation of $P_{\mathbf{k},\beta}(\mathbf{m}|\phi, \mathbf{A})$ or $\mathbb{P}(\mathfrak{m}, \mathfrak{n}|\phi, \mathbf{A})$, (2) calculation of $g(k)$ from these probabilities via equation (23) or equation (24), and (3) estimation of the phases ϕ from $g(k)$. Clearly (2) and (3) only need to be done once for the entire set of experiments.

The estimation of the phases ϕ requires solving two least squares equations, with cost $O(l^2 K)$ (recalling that l is the number of frequencies to estimate, and K is the maximum known value of $g(k)$), and diagonalizing the time-shift matrix $\mathfrak{T}$ with cost $O(l^3)$. For single-round phase estimation this is the dominant calculation, as calculating $g(k)$ from equation (23) requires simply K additions. As a result this estimator proves to be incredibly fast, able to estimate one frequency from a set of $N = 10^6$ experiments of up to $K = 10\,000$ in <100 ms, and $l = 1000$ frequencies from $N = 10^6$ experiments with $K = 1000$ in <1 min. However, for multi-round phase estimation the calculation of $g(k)$ in equation (24) scales as $O(K^4)$. This then dominates the calculation, requiring 30 s to calculate 50 points of $g(k)$. (All calculations performed on a 2.4 GHz Intel i3 processor.) We note that all the above times are small fractions of the time required to generate the experimental data when $N \gg K$, making this a very practical estimator for near-term experiments.

3.2. Efficient Bayesian analysis

When the starting state is the eigenstate $|\phi_0\rangle$, the problem of determining ϕ_0 based on the obtained multi-experiment data has a natural solution via Bayesian methods [10, 35]. Here we extend such Bayesian methodology to a general starting state. For computational efficiency we store a probability distribution over phases $P(\phi)$ using a Fourier representation of this periodic function $P(\phi)$ (see appendix C). This technique can also readily be applied to the case of Bayesian phase estimation applied to a single eigenstate.

A clearly information-theoretic optimal Bayesian strategy is to choose the ϕ and $\mathbf{A}$ based on the data obtained in some N experiments [8]. After these N experiments, leading to qubit measurement outcomes

$\{\mathbf{m}_i\}_{i=1}^N$, one can simply choose $\mathbf{A}$, ϕ which maximizes the posterior distribution:

$$P_{\text{post}}(\phi, \mathbf{A}) = \frac{P_{\{\mathbf{k}_i\},\{\beta_i\}}(\{\mathbf{m}_i\}|\phi, \mathbf{A})}{P(\{\mathbf{m}_i\})}P_{\text{prior}}(\phi, \mathbf{A}), \tag{29}$$

In other words, one chooses

$$\begin{aligned}(\phi_{\text{opt}}, \mathbf{A}_{\text{opt}}) &= \arg\max_{\phi,\mathbf{A}} \log P_{\text{post}}(\phi, \mathbf{A}) \\ &= \arg\max_{\phi,\mathbf{A}} [\log P_{\{\mathbf{k}_i\},\{\beta_i\}}(\{\mathbf{m}_i\}|\phi, \mathbf{A}) + \log P_{\text{prior}}(\phi, \mathbf{A})].\end{aligned}$$

A possible way of implementing this strategy is to (1) assume the prior distribution to be independent of $\mathbf{A}$ and ϕ and (2) estimate the maximum by assuming that the derivative with respect to $\mathbf{A}$ and ϕ vanishes at this maximum.

Instead of this method we update our probability distribution over ϕ and $\mathbf{A}$ after each experiment. After experiment n the posterior distribution $P_n(\phi, \mathbf{A})$ via Bayes' rule reads

$$P_n(\phi, \mathbf{A}) = \frac{P_{\mathbf{k},\beta}(\mathbf{m}|\phi, \mathbf{A})}{P(\mathbf{m})}P_{n-1}(\phi, \mathbf{A}). \tag{30}$$

To calculate the updates we will assume that the distribution over the phases ϕ_j and probabilities A_j are independent, that is

$$P_n(\phi, \mathbf{A}) = P_n^{\text{red}}(\mathbf{A}) \prod_{j=0}^{N_{\text{eig}}-1} P_n^j(\phi_j). \tag{31}$$

As prior distribution we take $P_0(\phi, \mathbf{A}) = P_{\text{prior}}(\mathbf{A})P_{\text{prior}}(\phi)$ with a flat prior $P_{\text{prior}}(\phi) = \left(\frac{1}{2\pi}\right)^{N_{\text{eig}}}$, given the absence of a more informed choice. We take $P_{\text{prior}}(\mathbf{A}) = e^{-(\mathbf{A}-\mathbf{A}_0)^2/2\Sigma^2}$, with $\mathbf{A}_0$ and Σ^2 approximate mean and covariance matrices. We need to do this to break the symmetry of the problem, so that $\tilde{\phi}_0$ is estimating ϕ_0 and not any of the other ϕs. We numerically find that the estimator convergence is relatively independent of our choice of $\mathbf{A}_0$ and Σ^2.

The approximation in equation (31) allows for relatively fast calculations of the Bayesian update of $P_n^j(\phi_j)$, and an approximation to the maximum-likelihood estimation of $P_n^{\text{red}}(\mathbf{A})$. Details of this computational implementation are given in (C1).

3.2.1. Classical computation cost

In contrast to the time-series estimator, the Bayesian estimator incurs a computational cost in processing the data from each individual experiment. On the other hand, obtaining the estimate $\tilde{\phi}_0$ for ϕ_0 is simple, once one has the probability distribution $P^{j=0}(\phi)$:

$$\tilde{\phi}_0 = \arg\left(\int d\phi P^{j=0}(\phi)e^{i\phi}\right).$$

A key parameter here is the number of frequencies #freq stored in the Fourier representation of $P(\phi)$; each update requires multiplying a vector of length #freq by a sparse matrix. Our approximation scheme for calculating the update to $\mathbf{A}$ makes this multiplication the dominant time cost of the estimation. As we argue in (C1) one requires #freq $\geqslant K_{\text{tot}}$ to store a fully accurate representation of the probability vector. For the single-round scenario with $k_r = 1$, hence $K_{\text{tot}} = N$, we find a large truncation error when #freq $\ll N$, and so the computation cost scales as N^2. In practice we find that processing the data from $N < 10^4$ experiments takes seconds on a classical computer, but processing more than 10^5 experiments becomes rapidly unfeasible.

3.3. Experiment design

Based on the considerations above we seek to compare some choices for the meta-parameters in each experiment, namely the number of rounds, and the input parameters k_r and β_r for each round.

Previous work [10, 36], which took as a starting state the eigenstate $|\phi_0\rangle$, formulated a choice of k and β, using single-round experiments and Bayesian processing, namely

$$k = \min\left(\left\lceil\frac{1.25}{\sigma_{P_n^{j=0}(\phi_0)}}\right\rceil, K_{\text{err}}\right), \qquad \beta \sim P_n^{j=0}(\phi_0 = \beta), \tag{32}$$

Roughly, this heuristic adapts to the expected noise in the circuit by not using any k such that the implementation of U^k takes longer than $T_{\text{err}}/n_{\text{sys}}$. It also adapts k to the standard-deviation of the current posterior probability

distribution over ϕ_0: a small standard-deviation after the *n*th experiment implies that *k* should be chosen large to resolve the remaining bits in the binary expansion of ϕ_0[8].

In this work we use a starting state which is not an eigenstate, and as such we must adjust the choice in equation (32). As noted in section 3, to separate different frequency contributions to $g(k)$ we need good accuracy beyond that at a single value of *k*. The optimal choice of the number of frequencies to estimate depends on the distribution of the A_j, which may not be well known in advance. Following the inspiration of [10], we choose for the Bayesian estimator

$$k \in \{1, \ldots, K\}$$
$$K = \min\left(\left\lceil \frac{1.25}{\sigma_{P_n^{j=0}(\phi_0)}} \right\rceil, K_{\mathrm{err}}\right). \tag{33}$$

We thus similarly bound *K* depending how well one has already converged to a value for ϕ_0 which constitutes some saving of resources. At large *N* we numerically find little difference between choosing *k* at random from $\{1, \ldots, K\}$ and cycling through $k = 1, \ldots, K$ in order. For this Bayesian estimator we draw β at random from a uniform distribution $[0, 2\pi)$. We find that the choice of β has no effect on the final estimation (as long as it is not chosen to be a single number) For the time-series estimator applied to single-round experiments, we choose to cycle over $k = 1, \ldots, K$ so that it obtains a complete estimate of $g(k)$ as soon as possible, taking an equal number of experiments with final rotation $\beta = 0$ and $\beta = \pi/2$ at each *k*. Here again $K \leqslant K_{\mathrm{err}}$, so that we choose the same number of experiments for each $k \leqslant K$. For the time-series estimator applied to multi-round experiments, we choose an equal number of rounds with $\beta = 0$ and $\beta = \pi/2$, taking the total number of rounds equal to $R = K$.

4. Results without experimental noise

We first focus on the performance of our estimators in the absence of experimental noise, to compare their relative performance and check the analytic predictions in section 3.1. Although with a noiseless experiment our limit for *K* is technically infinite, we limit it to a make connection with the noisy results of the following section. Throughout this section we generate results directly by calculating the function $P_{\mathbf{k},\beta}(\mathbf{m}|\phi, \mathbf{A})$ and sampling from it. Note that $P_{\mathbf{k},\beta}(\mathbf{m}|\phi, \mathbf{A})$ only depends on N_{eig} and not on the number of qubits in the system.

4.1. Single eigenvalues
To confirm that our estimators achieve the scaling bounds discussed previously, we first test them on the single eigenvalue scenario $N_{\mathrm{eig}} = 1$. In figure 3, we plot the scaling of the average absolute error in an estimation $\tilde{\phi}$ of a single eigenvalue $\phi \in [-\pi, \pi)$, defined so as to respect the 2π-periodicity of the phase:

$$\epsilon := \langle \min(|\phi - \tilde{\phi}|, 2\pi - |\phi - \tilde{\phi}|) \rangle = \langle |\mathrm{Arg}(e^{i(\phi - \tilde{\phi})})| \rangle, \tag{34}$$
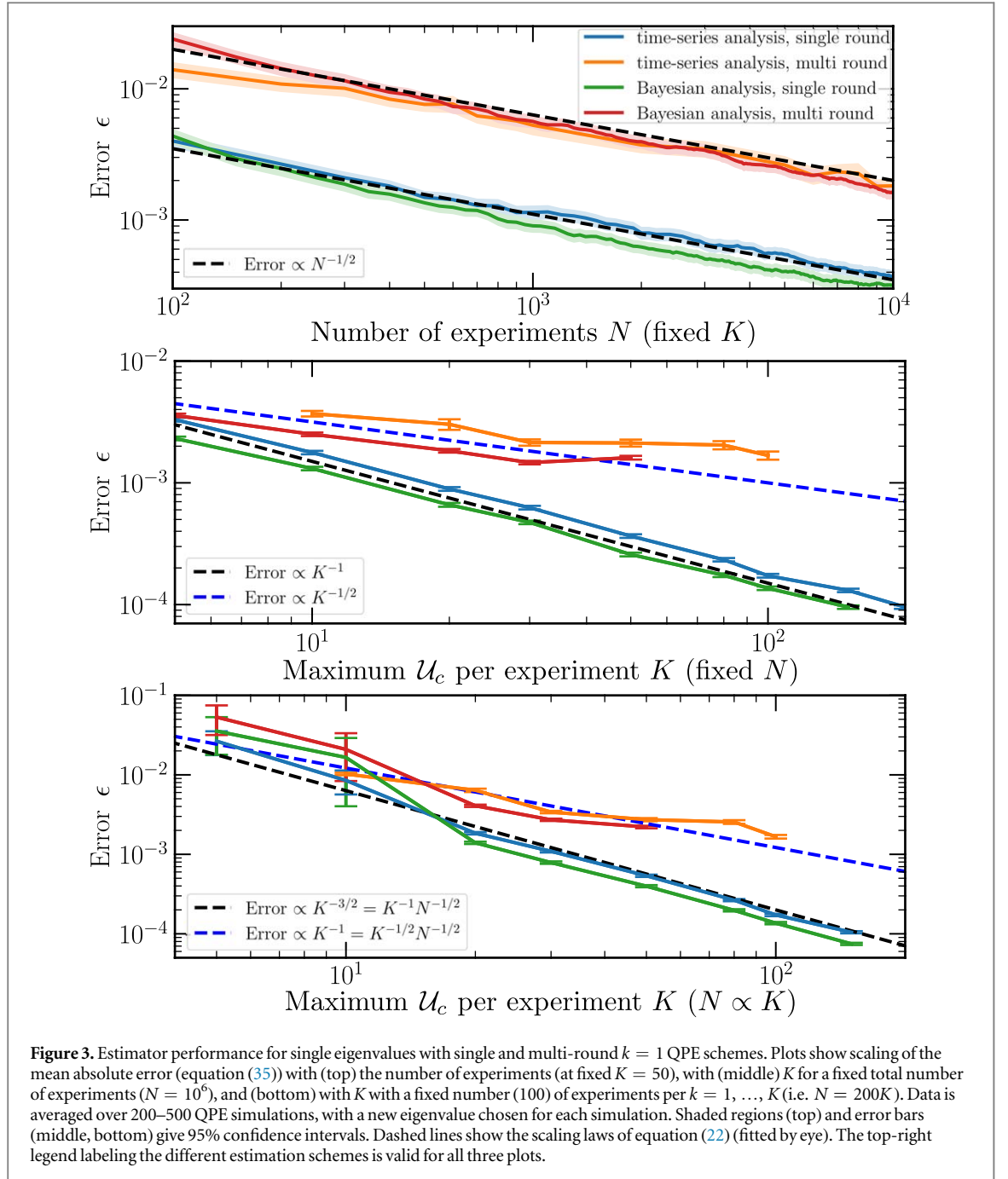
as a function of varying *N* and *K*. Here $\langle \rangle$ represents an average over repeated QPE simulations, and the Arg function is defined using the range $[-\pi, \pi)$ (otherwise the equality does not hold).

We see that both estimators achieve the previously-derived bounds in 3.1 (overlaid as dashed lines), and both estimators achieve almost identical convergence rates. The results for the Bayesian estimation match the scaling observed in [10]. Due to the worse scaling in *K*, the multi-round $k = 1$ estimation significantly underperforms single-round phase estimation. This is a key observation of this paper, showing that if the goal is to estimate a phase rather than to project onto an eigenstate, it is preferable to do single-round experiments.

4.2. Example behavior with multiple eigenvalues
The performance of QPE is dependent on both the estimation technique and the system being estimated. Before studying the system dependence, we first demonstrate that our estimators continue to perform at all in the presence of multiple eigenvalues. In figure 4, we demonstrate the convergence of both the Bayesian and time-series estimators in the estimation of a single eigenvalue $\phi_0 = -0.5$ of a fixed unitary *U*, given a starting state $|\Psi_0\rangle$ which is a linear combination of 10 eigenstates $|\phi_j\rangle$. We fix $|\langle \phi_0|\Psi_0\rangle|^2 = 0.5$, and draw other eigenvalues and amplitudes at random from $[0, \pi]$ (making the minimum gap $\phi_j - \phi_0$ equal to 0.5). We perform 2000 QPE simulations with $K = 50$, and calculate the mean absolute error ϵ (equation (35), solid), Holevo variance $|\langle e^{i\tilde{\phi}}\rangle|^{-2} - 1$ (dashed), and root mean squared error ϵ_{rms} (dotted), given by
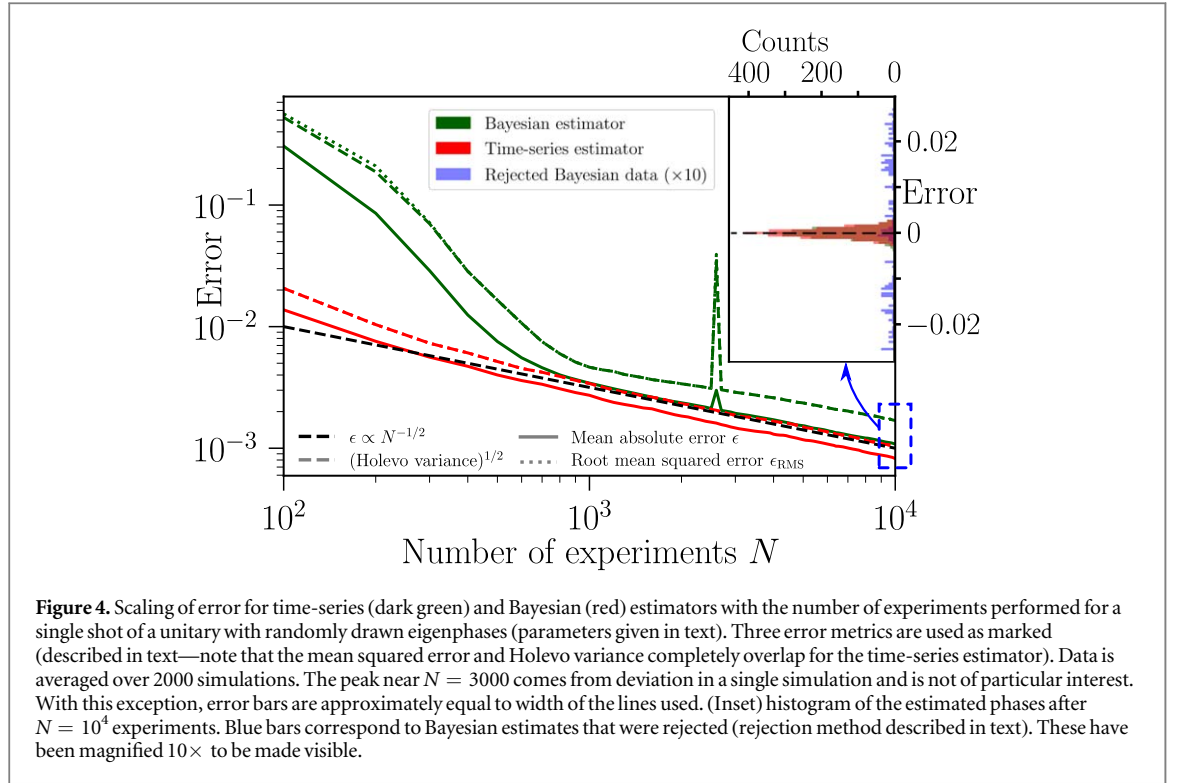
---

[8] Note that this strategy is the opposite of textbook phase estimation in which one necessarily learns the least-significant bit of ϕ_0 *first* by choosing the largest *k*. One chooses the next smallest *k* and β so that the next measurement outcome gives the next more-significant bit etc.

Figure 3. Estimator performance for single eigenvalues with single and multi-round $k = 1$ QPE schemes. Plots show scaling of the mean absolute error (equation (35)) with (top) the number of experiments (at fixed $K = 50$), with (middle) K for a fixed total number of experiments ($N = 10^6$), and (bottom) with K with a fixed number (100) of experiments per $k = 1, \ldots, K$ (i.e. $N = 200K$). Data is averaged over 200–500 QPE simulations, with a new eigenvalue chosen for each simulation. Shaded regions (top) and error bars (middle, bottom) give 95% confidence intervals. Dashed lines show the scaling laws of equation (22) (fitted by eye). The top-right legend labeling the different estimation schemes is valid for all three plots.

$$\epsilon_{\mathrm{rms}}^2 := \langle \min(|\phi - \tilde{\phi}|, 2\pi - |\phi - \tilde{\phi}|)^2 \rangle = \langle |\mathrm{Arg}(e^{i(\phi - \tilde{\phi})})|^2 \rangle. \tag{35}$$

We observe that both estimators retain their expected $\epsilon \propto N^{-1/2}$, with one important exception. The Bayesian estimator occasionally (10% of simulations) estimates multiple eigenvalues near ϕ_0. When this occurs, the estimations tend to repulse each other, making neither a good estimation of the target. This is easily diagnosable without knowledge of the true value of ϕ_0 by inspecting the gap between estimated eigenvalues. While using this data to improve estimation is a clear target for future research, for now we have opted to reject simulations where such clustering occurs (in particular, we have rejected data points where $\min(\bar{\phi}_0 - \bar{\phi}_j) < 0.05$). That this is required is entirely system-dependent: we find the physical Hamiltonians studied later in this text to not experience this effect. We attribute this difference to the distribution of the amplitudes A_j—physical Hamiltonians tend to have a few large A_j, whilst in this simulation the A_j were distributed uniformly.

In the inset to figure 4, we plot a histogram of the estimated eigenphases after $N = 10^4$ experiments. For the Bayesian estimator, we show both the selected (green) and rejected (blue) eigenphases. We see that regardless of whether rejection is used, the distribution appears symmetric about the target phase ϕ_0. This suggests that in the absence of experimental noise, both estimators are unbiased. Proving this definitively for any class of systems is difficult, but we expect both estimators to be unbiased provided $A_0 \gg 1/K$. When $A_0 \leqslant 1/K$, one can easily

Figure 4. Scaling of error for time-series (dark green) and Bayesian (red) estimators with the number of experiments performed for a single shot of a unitary with randomly drawn eigenphases (parameters given in text). Three error metrics are used as marked (described in text—note that the mean squared error and Holevo variance completely overlap for the time-series estimator). Data is averaged over 2000 simulations. The peak near $N = 3000$ comes from deviation in a single simulation and is not of particular interest. With this exception, error bars are approximately equal to width of the lines used. (Inset) histogram of the estimated phases after $N = 10^4$ experiments. Blue bars correspond to Bayesian estimates that were rejected (rejection method described in text). These have been magnified $10\times$ to be made visible.

construct systems for which no phase estimation can provide an unbiased estimation of ϕ_0 (following the arguments of section 3). We further see that the scaling of the rms error ϵ_{rms} and the Holevo variance match the behavior of the mean absolute error ϵ, implying that our results are not biased by the choice of estimator used.

4.3. Estimator scaling with two eigenvalues

The ability of QPE to resolve separate eigenvalues at small K can be tested in a simple scenario of two eigenvalues, ϕ_0 and ϕ_1. The input to the QPE procedure is then entirely characterized by the overlap A_0 with the target state $|\phi_0\rangle$, and the gap $\delta = |\phi_0 - \phi_1|$.

In figure 5, we study the performance of our time-series estimator in estimating ϕ_0 after $N = 10^6$ experiments with $K = 50$, measured again by the mean error ϵ (equation (35)). We show a two-dimensional plot (averaged over 500 simulations at each point A_0, δ) and log–log plots of one-dimensional vertical (lower left) and horizontal (lower right) cuts through this surface. Due to computational costs, we are unable to perform this analysis with the Bayesian estimator, or for the multi-round scenario. We expect the Bayesian estimator to have similar performance to the time-series estimator (given their close comparison in sections 4.1 and 4.2). We also expect the error in multi-round QPE to follow similar scaling laws in A_0 and δ as single-round QPE (i.e. multi-round QPE should be suboptimal only in its scaling in K).

The ability of our estimator to estimate ϕ_0 in the presence of two eigenvalues can be split into three regions (marked as (a), (b), (c) on the surface plot). In region (a), we have performed insufficient sampling to resolve the eigenvalues ϕ_0 and ϕ_1, and QPE instead estimates the weighted average phase $A_0\phi_0 + A_1\phi_1$. The error in the estimation of ϕ_0 then scales by how far it is from the average, and how well the average is resolved

$$\epsilon \propto (1 - A_0)\delta K^{-1} N^{-1/2}. \tag{36}$$

In region (b), we begin to separate ϕ_0, from the unwanted frequency ϕ_1, and our convergence halts

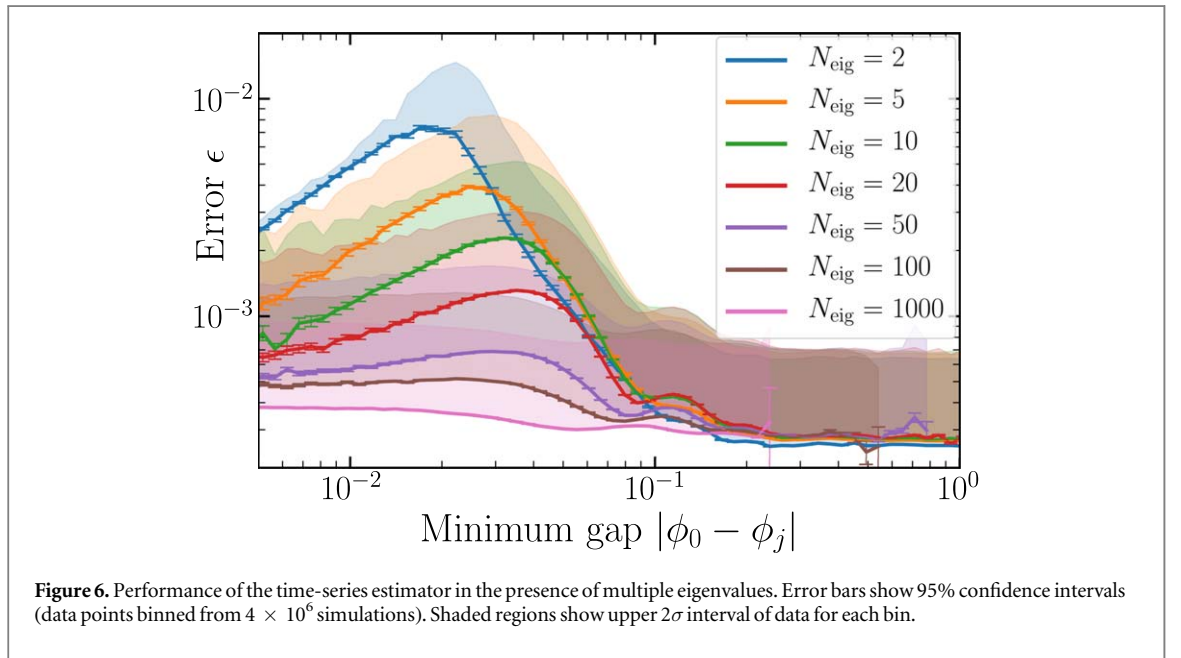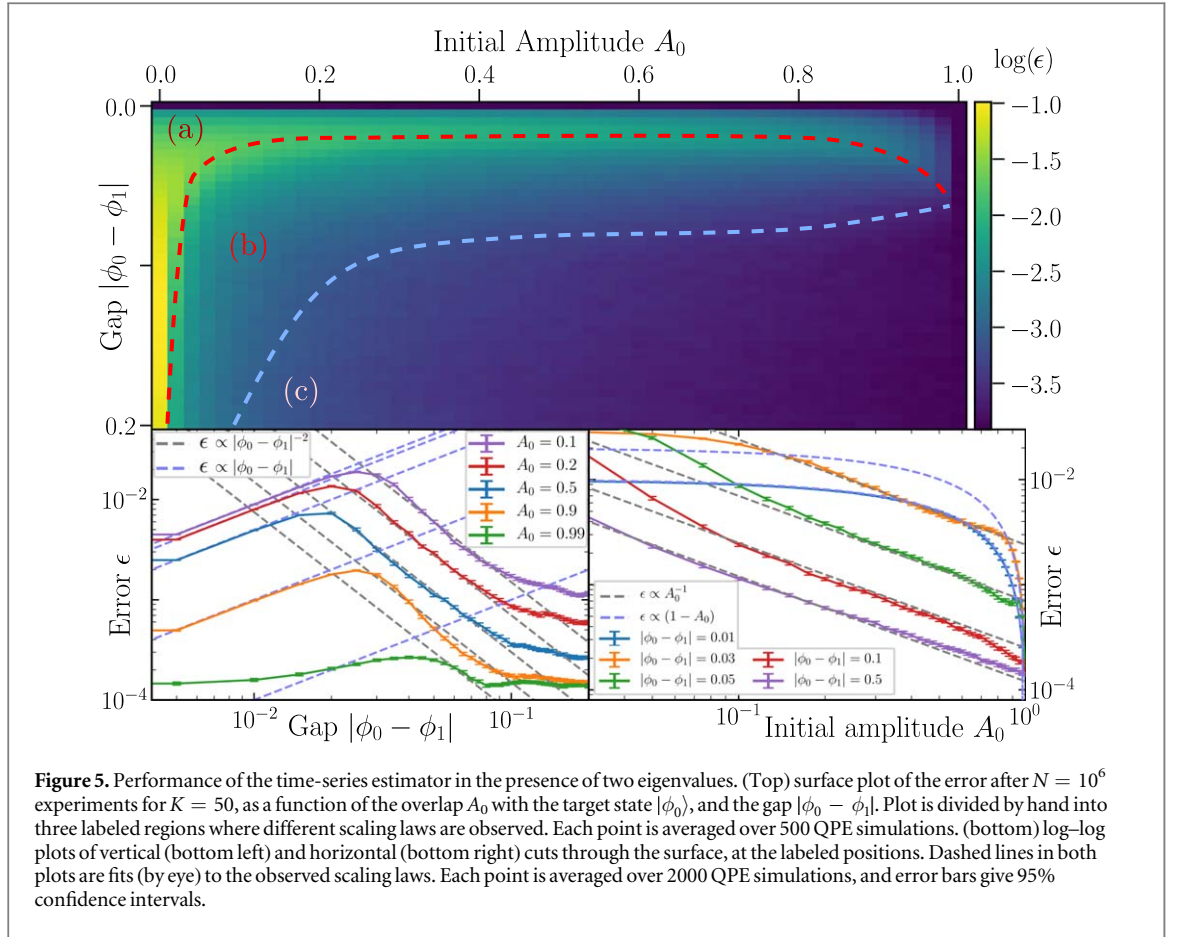$$\epsilon \propto A_0^{-1}\delta^{-2}. \tag{37}$$

In region (c), the gap is sufficiently well resolved and our estimation returns to scaling well with N and K

$$\epsilon \propto A_0^{-1}K^{-1}N^{-1/2}. \tag{38}$$

The scaling laws in all three regions can be observed in the various cuts in the lower plots of figure 5. We note that the transition between the three regions is not sharp (boundaries estimated by hand), and is K and N-dependent.
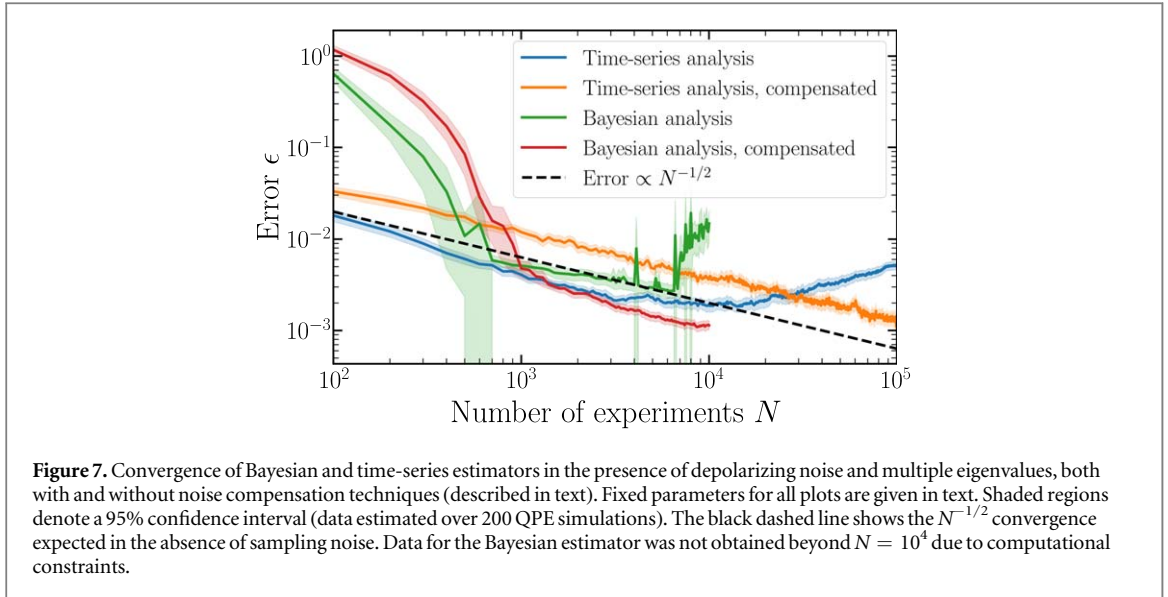
4.4. Many eigenvalues

To show that our observed scaling is applicable beyond the toy 2-eigenvalue system, we now shift to studying systems of random eigenvalues with $N_{\mathrm{eig}} > 1$. In keeping with our insight from the previous section, in figure 6

Figure 5. Performance of the time-series estimator in the presence of two eigenvalues. (Top) surface plot of the error after $N = 10^6$ experiments for $K = 50$, as a function of the overlap A_0 with the target state $|\phi_0\rangle$, and the gap $|\phi_0 - \phi_1|$. Plot is divided by hand into three labeled regions where different scaling laws are observed. Each point is averaged over 500 QPE simulations. (bottom) log–log plots of vertical (bottom left) and horizontal (bottom right) cuts through the surface, at the labeled positions. Dashed lines in both plots are fits (by eye) to the observed scaling laws. Each point is averaged over 2000 QPE simulations, and error bars give 95% confidence intervals.

Figure 6. Performance of the time-series estimator in the presence of multiple eigenvalues. Error bars show 95% confidence intervals (data points binned from 4×10^6 simulations). Shaded regions show upper 2σ interval of data for each bin.

we fix $\phi_0 = 0$, and study the error ϵ as a function of the gap

$$\delta = \min_{j>1}(|\phi_j - \phi_0|). \tag{39}$$

We fix $A_0 = 0.5$, and draw the other parameters for the system from a uniform distribution: $\phi_j \sim [\delta, \pi]$, $A_j \sim [0, 0.5]$ (fixing $\sum_{j=1}^{N_{eig}} A_j = 1 - A_0$). We plot both the average error ϵ (line) and the upper 47.5% confidence interval

Figure 7. Convergence of Bayesian and time-series estimators in the presence of depolarizing noise and multiple eigenvalues, both with and without noise compensation techniques (described in text). Fixed parameters for all plots are given in text. Shaded regions denote a 95% confidence interval (data estimated over 200 QPE simulations). The black dashed line shows the $N^{-1/2}$ convergence expected in the absence of sampling noise. Data for the Bayesian estimator was not obtained beyond $N = 10^4$ due to computational constraints.

$[\epsilon, \epsilon + 2\sigma_\epsilon]$ (shaded region) for various choices of N_{eig}. We observe that increasing the number of spurious eigenvalues does not critically affect the error in estimation; indeed the error generally decreases as a function of the number of eigenvalues. This makes sense; at large N_{eig} the majority of eigenvalues sit in region (c) of figure 5, and we do not expect these to combine to distort the estimation. Then, the nearest eigenvalue $\min_{j \neq 0} \phi_j$ has on average an overlap $A_j \propto 1/N_{\text{eig}}$, and its average contribution to the error in estimating ϕ_0 (inasmuch as this can be split into individual contributions) scales accordingly. We further note that the worst-case error remains that of two eigenvalues at the crossover between regions (a) and (b). In appendix D we study the effect of confining the spurious eigenvalues to a region $[\delta, \phi_{\text{max}}]$. We observe that when most eigenvalues are confined to regions (a) and (b), the scaling laws observed in the previous section break down, however the worst-case behavior remains that of a single spurious eigenvalue. This implies that sufficiently long K is not a requirement for QPE, even in the presence of large systems or small gaps δ; it can be substituted by sufficient repetition of experiments. However, we do require that the ground state is guaranteed to have sufficient overlap with the starting state—$A_0 > 1/K$ (as argued in section 3). As QPE performance scales better with K than it does with N, a quantum computer with coherence time $2T$ is still preferable to two quantum computers with coherence time T (assuming no coherent link between the two).

5. The effect of experimental noise

Experimental noise currently poses the largest impediment to useful computation on current quantum devices. As we suggested before, experimental noise limits K so that for $K \gtrsim K_{\text{err}}$ the circuit is unlikely to produce reliable results. However, noise on quantum devices comes in various flavors, which can have different corrupting effects on the computation. Some of these corrupting effects (in particular, systematic errors) may be compensated for with good knowledge of the noise model. For example, if we knew that our system applied $U = e^{-i\mathcal{H}(t+\epsilon)}$ instead of $U = e^{-i\mathcal{H}t}$, one could divide $\tilde{\phi}_0$ by $(t + \epsilon)/t$ to precisely cancel out this effect. In this study we have limited ourselves to studying and attempting to correct two types of noise: depolarizing noise, and circuit-level simulations of superconducting qubits. Given the different effects observed, extending our results to other noise channels is a clear direction for future research. In this section we do not study multi-round QPE, so each experiment consists of a single round. A clear advantage of the single-round method is that the only *relevant* effect of any noise in a single-round experiment is to change the outcome of the ancilla qubit, independent of the number of system qubits n_{sys}.

5.1. Depolarizing noise
A very simple noise model is that of depolarizing noise, where the outcome of each experiment is either correct with some probability p or gives a completely random bit with probability $1 - p$. We expect this probability p to depend on the circuit time and thus the choice of $k \geqslant 0$, i.e.

$$p = p(k) = e^{-k/K_{\text{err}}}. \tag{40}$$

We can simulate this noise by directly applying it to the calculated probabilities $P_{k,\beta}(m|\phi)$ for a single round

$$P_{k,\beta}(m|\phi) \rightarrow P_{k,\beta}(m|\phi)p(k) + \frac{1 - p(k)}{2}. \tag{41}$$

In figure 7, we plot the convergence of the time-series (blue) and Bayesian (green) estimators as used in the previous section as a function of the number of experiments, with fixed $K = 50 = K_{err}/2$ fixed, $A_0 = 0.5$, $N_{eig} = 10$ and $\delta = 0.5$. We see that both estimators obey $N^{-1/2}$ scaling for some portion of the experiment, however this convergence is unstable, and stops beyond some critical point.

Both the Bayesian and time-series estimator can be adapted rather easily to compensate for this depolarizing channel. To adapt the time-series analysis, we note that the effect of depolarizing noise is to send $g(k) \rightarrow g(k)p(k)$ when $k > 0$, via equation (23) and equation (41). Our time-series analysis was previously performed over the range $k = -K, \ldots, K$ (getting $g(-k) = g^*(k)$ for free), and over this range

$$g(k) \rightarrow g(k)p(|k|). \tag{42}$$

$g(k)$ is no longer a sum of exponential functions over our interval $[-K, K]$, as it is not differentiable at $k = 0$, which is the reason for the failure of our time-series analysis. However, over the interval $[0, K]$ this is not an issue, and the time-series analysis may still be performed. If we construct a shift operator T using $g(k)$ from $k = 0, \ldots, K$, this operator will have eigenvalues $e^{i\phi_j - 1/K_{err}}$. This then implies that the translation operator T can be calculated using $g(k)$ with $k > 0$, and the complex argument of the eigenvalues of T give the correct phases ϕ_j. We see that this is indeed the case in figure 7 (orange line). Halving the range of $g(k)$ that we use to estimate ϕ_0 decreases the estimator performance by a constant factor, but this can be compensated for by increasing N.

Adapting the Bayesian estimator requires simply that we use the correct conditional probability, equation (41). This in turn requires that we either have prior knowledge of the error rate K_{err}, or estimate it alongside the phases ϕ_j. For simplicity, we opt to choose the former. In an experiment K_{err} can be estimated via standard QCVV techniques, and we do not observe significant changes in estimator performance when it is detuned. Our Fourier representation of the probability distribution of ϕ_0 can be easily adjusted to this change. The results obtained using this compensation are shown in figure 7: we observe that the data follows a $N^{-1/2}$ scaling again.

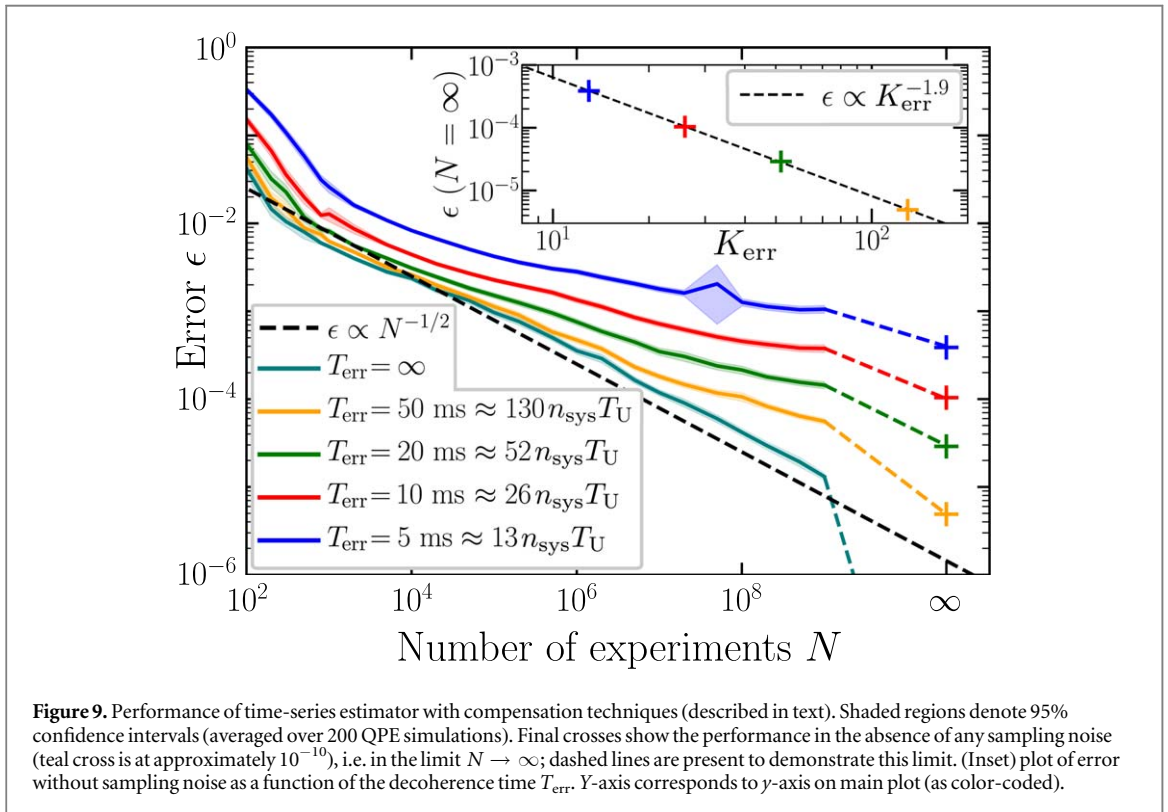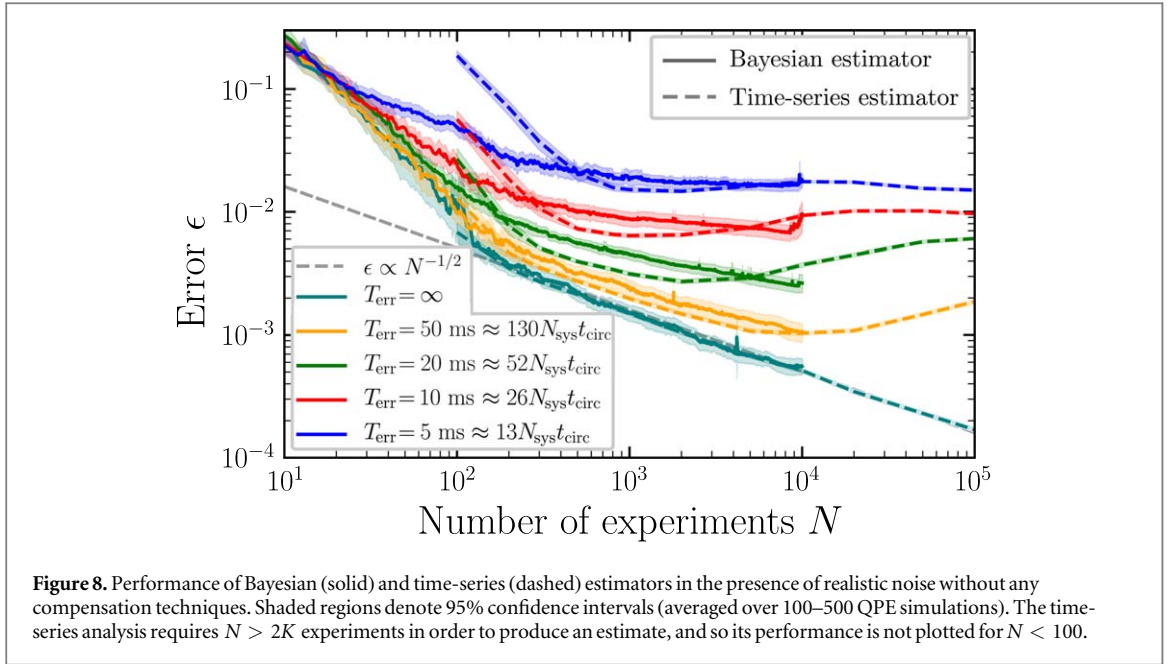## 5.2. Realistic circuit-level noise

Errors in real quantum computers occur at a circuit-level, where individual gates or qubits get corrupted via various error channels. To make connection to current experiments, we investigate our estimation performance on an error model of superconducting qubits. Full simulation details can be found in appendix E. Our error model is primarily dominated by T_1 and T_2 decoherence, incoherent two-qubit flux noise, and dephasing during single-qubit gates. We treat the decoherence time $T_{err} = T_1 = T_2$ as a free scale parameter to adjust throughout our simulations, whilst keeping all other error parameters tied to this single scale parameter for simplicity. In order to apply circuit-level noise we must run quantum circuit simulations, for which we use the quantumsim density matrix simulator first introduced in [37]. We then choose to simulate estimating the ground state energy of four hydrogen atoms in varying rectangular geometries, with Hamiltonian $\mathcal{H}$ taken in the STO-3G basis calculated via psi4 [38], requiring $n_{sys} = 8$ qubits. We make this estimation via a lowest-order Suzuki-Trotter approximation [39] to the time-evolution operator $e^{-i\mathcal{H}t}$. To prevent energy eigenvalues wrapping around the circle we fix $t = 1/\sqrt{\mathrm{Trace}[\mathcal{H}^\dagger\mathcal{H}]/(2^{n_{sys}})}$ [9]. The resultant 9-qubit circuit is made using the OpenFermion package [9].

In lieu of any circuit optimizations (e.g. [23, 40]), the resulting circuit has a temporal length per unitary of $T_U = 42 \ \mu s$ (with single- (two-) qubit gate times 20 ns (40 ns)). This makes the circuit unrealistic to operate at current decoherence times for superconducting circuits, and we focus on decoherence times $1-2$ orders of magnitude above what is currently feasible, i.e. $T_{err} = 5-50$ ms. However one may anticipate that the ratio T_U/T_{err} can be enlarged by circuit optimization or qubit improvement. Naturally, choosing a smaller system, less than 8 qubits, or using error mitigation techniques could also be useful.

We observe realistic noise to have a somewhat different effect on both estimators than a depolarizing channel. Compared to the depolarizing noise, the noise may (1) be biased towards 0 or 1 and/or (2) its dependence on k may not have the form of equation (40).

In figure 8, we plot the performance of both estimators at four different noise levels (and a noiseless simulation to compare), in the absence of any attempts to compensate for the noise. Unlike for the depolarizing channel, where a $N^{-1/2}$ convergence was observed for some time before the estimator became unstable, here we

---

[9] This normalization is not good for large systems since it makes t exponentially small in system size. A scalable choice for normalization is to first determine upper and lower bounds on the eigenvalues of $\mathcal{H}$ present in the starting state, assume that they occur in a some numerical window W. Given W (which is at most poly(n_{sys})), one sets $U = \exp(-i\pi\mathcal{H}/W)$. The implementation of this U in Trotterized form with sufficient accuracy determines T_U.

Figure 8. Performance of Bayesian (solid) and time-series (dashed) estimators in the presence of realistic noise without any compensation techniques. Shaded regions denote 95% confidence intervals (averaged over 100–500 QPE simulations). The time-series analysis requires $N > 2K$ experiments in order to produce an estimate, and so its performance is not plotted for $N < 100$.

Figure 9. Performance of time-series estimator with compensation techniques (described in text). Shaded regions denote 95% confidence intervals (averaged over 200 QPE simulations). Final crosses show the performance in the absence of any sampling noise (teal cross is at approximately 10^{-10}), i.e. in the limit $N \to \infty$; dashed lines are present to demonstrate this limit. (Inset) plot of error without sampling noise as a function of the decoherence time T_{err}. Y-axis corresponds to y-axis on main plot (as color-coded).

see both instabilities and a loss of the $N^{-1/2}$ decay to begin with. Despite this, we note that reasonable convergence (to within 1%−2%) is achieved, even at relatively low coherence times such as $K_{err} = 10$. Regardless, the lack of eventual convergence to zero error is worrying, and we now shift to investigating how well it can be improved for either estimator.

Adjusting the time-series estimator to use only $g(k)$ for positive k gives approximately 1−2 orders of magnitude improvement. In figure 9, we plot the estimator convergence with this method. We observe that the estimator is no longer unstable, but the $N^{-1/2}$ convergence is never properly regained. We may study this convergence in greater deal for this estimator, as we may extract $g(k)$ directly from our density-matrix simulations, and thus investigate the estimator performance in the absence of sampling noise (crosses on screen). We note that similar extrapolations in the absence of noise, or in the presence of depolarizing noise
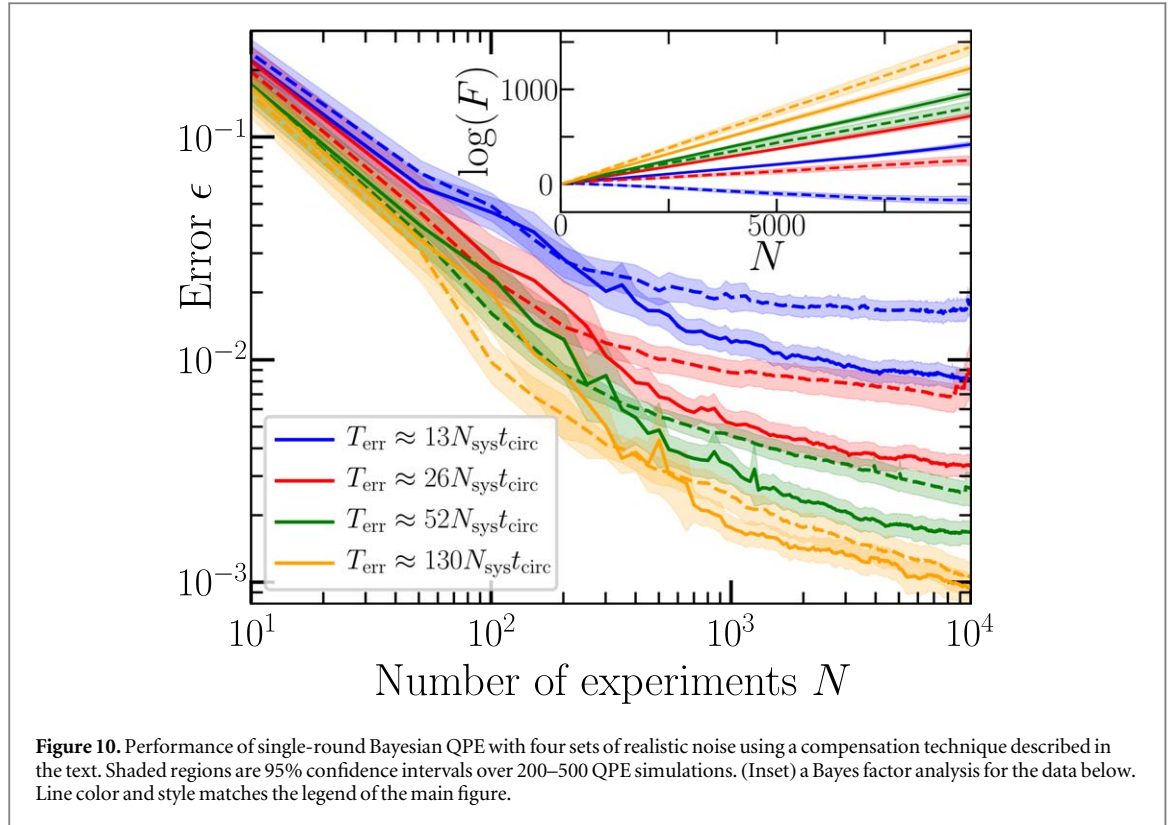
Figure 10. Performance of single-round Bayesian QPE with four sets of realistic noise using a compensation technique described in the text. Shaded regions are 95% confidence intervals over 200–500 QPE simulations. (Inset) a Bayes factor analysis for the data below. Line color and style matches the legend of the main figure.

Table 1. Table comparing metrics of interest between the two studied estimators. All metrics are implementation-specific, and may be improvable.

	Time-series estimator	Bayesian estimator
Speed (scaling)	$O(K)$	$O(N^2)$
Speed (timing)	Processes large datasets in milliseconds	Takes hours to process 10^5 experiments
Accuracy	$\epsilon \propto N^{-1/2}K^{-1}A_0^{-1}\delta^{-2}$ demonstrated.	$\epsilon \propto N^{-1/2}K^{-1}$ demonstrated $\epsilon \propto A_0^{-1}\delta^{-2}$ expected.
Number of eigenvalues estimated	$100-200$ with relative ease	Limited to $2-5$
Improve accuracy via classical approximation	Not obvious	Can get speedup via choice of prior (not attempted in this work)
Account for error	Limited ability	Limited ability

(when compensated) give an error rate of around 10^{-10}, which we associate to fixed-point error in the solution to the least squares problem (this is also observed in the curve without noise in figure 9). Plotting this error as a function of K_{err} shows a power-law decay - $\epsilon \propto K_{\text{err}}^{-\alpha} \propto T_{\text{err}}^{-\alpha}$ with $\alpha = 1.9 \approx 2$. We do not have a good understanding of the source of the obtained power law.

The same compensation techniques that restored the performance of the Bayesian estimator in the presence of depolarizing noise do not work nearly as well for realistic noise. Most likely this is due to the fact that the actual noise is not captured by a k-dependent depolarizing probability. In figure 10 we plot the results of using a Bayesian estimator when attempting to compensate for circuit-level noise by approximating it as a depolarizing channel with a decay rate (equation (40)) of $K_{\text{err}} = T_{\text{err}}/T_U\,n_{\text{sys}}$. This can be compared with the results of figure 8 where this compensation is not attempted. We observe a factor 2 improvement at low T_{err}, however the $N^{-1/2}$ scaling is not regained, and indeed the estimator performance appears to saturate at roughly this point. Furthermore, at $T_{\text{err}} = 50$ ms, the compensation techniques do not improve the estimator, and indeed appear to make it more unstable.

To investigate this further, in figure 10 (inset) we plot a Bayes factor analysis of the Bayesian estimators with and without compensation techniques. The Bayes factor analysis is obtained by calculating the Bayes factors

$$F = \prod_{\text{expt } n} \frac{P(m_n|M)}{P(m_n|M_0)}, \tag{43}$$

where M is the chosen Bayesian model (including the prior knowledge), and M_0 is a reference model, and $P(m|M)$ is the probability of observing measurement m given model M. As a reference model we take that of random noise— $P(m|M_0) = 0.5$. We observe that at large T_{err} the Bayes factor with compensation falls below that without, implying that the compensation techniques make the model worse. We also observe that at very small T_{err}, the estimator makes worse predictions than random noise ($\log(F) < 0$). Despite our best efforts we have been unable to further improve the Bayesian estimator in noisy single-round QPE experiments.

6. Discussion

In this work, we have presented and studied the performance of two estimators for QPE at low K for different experiment protocols, different systems (in particular those with one versus many eigenvalues), and under simplistic and realistic noise conditions. These findings are summarized in table 1. From our numerical studies, we observe scaling laws for our time-series estimator; we find it first-order sensitive to the overlap A_0 between starting state and ground state, second-order sensitive to the gap between the ground state and the nearest eigenstates, and second-order sensitive to the coherence time of the system. The Bayesian estimator appears to perform comparably to the time-series estimator in all circumstances, and thus should obey similar scaling laws.

We further observe that realistic noise has a worse effect on QPE than a depolarizing channel, for which the effects can largely be mitigated. We have numerically explored (but not reported) multi-round QPE in the presence of noise. Since each experiment has multiple outputs, it is harder to adapt the classical data analysis to the presence of noise and our results for realistic noise have not been convincing so far. Since the performance of multi-round noiseless QPE is already inferior to single-round noiseless QPE, we do not advocate it as a near-term solution, although, for noiseless long circuits it does have the ability to project onto a single eigenstate, which single-round QPE certainly does not.

Despite our slightly pessimistic view of the effect of errors on the performance of QPE, we should note that the obtained error of 10^{-3} at $T_{err} \approx 13 n_{sys} T_U$ or $K_{err} = 13$ would be sufficient to achieve chemical accuracy in a small system. However, as the energy of a system scales with the number of particles, if we require a Hamiltonian's spectrum to fit in $[-\pi, \pi)$, we will need a higher resolution for QPE, making error rates of 10^{-3} potentially too large. This could potentially be improved by improving the compensation techniques described in the text, applying error mitigation techniques to effectively increase T_{err}, or by using more well-informed prior distributions in the Bayesian estimator to improve accuracy. All of the above are obvious directions for future work in optimizing QPE for the NISQ era. Another possible direction is to investigate QPE performance in other error models than the two studied here. Following [6], we expect SPAM errors to be as innocuous as depolarizing noise. However, coherent errors can be particularly worrying as they imitate alterations to the unitary U. The time-series estimator is a clear candidate for such a study, due to its ease in processing a large number of experiments and its ability to be studied in the absence of sampling noise. We also expect that it is possible to combine the time-series estimator with the Heisenberg-limited scaling methods of [6, 32] so as to extend these optimal methods to the multiple-eigenvalue scenario with $N_{eig} > 1$ eigenvalues, and that these methods could be extended to analog or ancilla-free QPE settings such as described in [6].

In this work we do not compare the performance of QPE with purely classical methods. Let us assume that we have a classical efficient representation of the starting state Ψ and one can efficiently calculate $\text{Tr}\,\mathcal{H}^k |\Psi\rangle\langle\Psi|$ for $k = 1, \ldots, K$ with $K = O(1)$ (for fermionic Gaussian starting states and fermionic Hamiltonians this is possible as a single fermionic term in $\mathcal{H}^k$ can be estimated as the Pfaffian of some matrix). Then, if there are at most $K = O(1)$ eigenstates in this initial state, the time-series method would allow us to extract these eigenvalues efficiently. Thus in this setting and under these assumptions QPE would not offer an exponential computational advantage.

Appendix A. Derivation of the identity in equation (25)

One first writes for $0 \leqslant k \leqslant K/2$:

$$\sum_j A_j \exp(ik\phi_j) = \sum_{\mathbf{m},\mathbf{n}} \Pi_{i=1}^k [(-1)^{m_i} - i(-1)^{n_i}]$$
$$\times \mathbb{P}(m_1, \ldots, m_{K/2}, n_1, \ldots n_{K/2} | \phi, \mathbf{A}), \tag{A1}$$

where $\mathbb{P}(m_1, \ldots, m_{K/2}, n_1, \ldots n_{K/2} | \phi, \mathbf{A})$ is the probability for a specific series of outcomes $m_1, \ldots, m_{K/2}$ for $\beta = 0$ and $n_1, \ldots, n_{K/2}$ for $\beta = \pi/2$. To see that the above is true, note that it is quickly true for $N_{\text{eig}} = 1$ by using equation (23) for $g(1)$. By linearity on the left and right-hand side it then holds generally.

Since the order of the outcomes of the rounds does not matter, i.e. $\mathbb{P}(m_1, \ldots, m_{K/2}, n_1, \ldots n_{K/2} | \phi, \mathbf{A})$ only depends on the Hamming weights $\mathfrak{m} = |\mathbf{m}|$ and $\mathfrak{n} = |\mathbf{n}|$, we can symmetrize the coefficient over permutations of the rounds and replace $\mathbb{P}(m_1, \ldots, m_{K/2}, n_1, \ldots n_{K/2} | \phi, \mathbf{A})$ by $\mathbb{P}(\mathfrak{m}, \mathfrak{n} | \phi, \mathbf{A}) / \left(\binom{K/2}{\mathfrak{m}} \binom{K/2}{\mathfrak{n}} \right)$. This gives the following expression for $\chi_k(\mathfrak{m}, \mathfrak{n})$:

$$\chi_k(\mathfrak{m}, \mathfrak{n}) = \frac{1}{((K/2)!)^2} \sum_{\pi_1 \in S_{K/2}, \pi_2 \in S_{K/2}} \prod_{i=1}^k ((-1)^{m_{\pi_1(i)}} - i(-1)^{n_{\pi_2(i)}}),$$

where m_i is the ith bit of a bitstring with Hamming weight $\mathfrak{m}$ (and similarly n_i), and $S_{K/2}$ is the symmetric group of permutations. We can expand this last expression as

$$\chi_k(\mathfrak{m}, \mathfrak{n}) = \sum_{k=0}^l \binom{k}{l} (-i)^{k-l} \rho(l, \mathfrak{m}) \rho(k-l, \mathfrak{n})$$
$$\rho(l, \mathfrak{m}) = \frac{1}{(K/2)!} \sum_\pi (-1)^{m_{\pi(1)}} \ldots (-1)^{m_{\pi(l)}}$$
$$= -1 + \frac{2}{(K/2)!} \sum_{\pi : m_{\pi(1)} \ldots m_{\pi(l)} \text{ is even}} 1.$$

The sum $\sum_{\pi : m_{\pi(1)} \ldots m_{\pi(l)} \text{ is even}}$ can be written as a sum over permutations such that $m_{\pi(1)} \ldots m_{\pi(l)}$ has Hamming weight $2p$ with $p = 0, 1, \ldots \lfloor l/2 \rfloor$. Then one counts the number of permutations of a $K/2$-bitstring of Hamming weight $\mathfrak{m}$ such that some segment of length l has Hamming weight $2p$ which equals $\binom{\mathfrak{m}}{2p} \binom{K/2 - \mathfrak{m}}{l - 2p} l! \ (K/2 - l)!$. All together this leads to $\chi_k(\mathfrak{m}, \mathfrak{n})$ in equation (25). It is not clear whether one can simplify this equality or verify it directly using other combinatorial identities or (Chebyshev) polynomials.

Appendix B. Variance calculations for time-series estimator

For the case of estimating a single eigenvalue using single-round QPE with the time-series estimator, one can directly calculate the error in the estimation. In this situation, our matrices G_0 and G_1 are column vectors

$$G_0^T = (g(-K), g(-K+1), \ldots, g(K-1)), \tag{B1}$$

$$G_1^T = (g(-K+1), g(-K+2), \ldots, g(K)). \tag{B2}$$

The least-squares solution for $\mathfrak{T}$ is then

$$\mathfrak{T} = (G_0^\dagger G_0)^{-1} G_0^\dagger G_1 = \frac{\sum_{k=-K}^{K-1} g^*(k) g(k+1)}{\sum_{k=-K}^{K-1} g^*(k) g(k)}. \tag{B3}$$

For a single frequency, $g(k) = e^{ik\phi}$, and immediately $\mathfrak{T} = e^{i\phi}$. However, we estimate the real and imaginary components of $g(k)$ separately. Let us write in terms of our independent components

$$\mathfrak{T} = \mathfrak{T}_r + i\mathfrak{T}_i, \qquad\qquad g(k) = g_k^0 + ig_k^1, \tag{B4}$$

remembering that $g_k^0 = g_{-k}^0$ and $g_k^1 = -g_{-k}^1$ (i.e. the variables are correlated). Our target angle $\phi = \tan^{-1} \mathfrak{T}_i / \mathfrak{T}_r$, and so we can calculate

$$\begin{aligned}
\mathrm{Var}(\phi) &= \sum_{a,k} \left[\frac{\partial \phi}{\partial g_k^a} \right]^2 \mathrm{Var}[g_k^a] \\
&= \left[\frac{1}{\mathfrak{T}_r^2 + \mathfrak{T}_i^2} \right]^2 \sum_{a,k} \left[\mathfrak{T}_r \frac{\partial \mathfrak{T}_i}{\partial g_k^a} - \mathfrak{T}_i \frac{\partial \mathfrak{T}_r}{\partial g_k^a} \right]^2 \mathrm{Var}[g_k^a].
\end{aligned} \tag{B5}$$

Let us expand out our real and imaginary components of $\mathfrak{T}$:

$$\mathfrak{T}_r = \frac{\sum_{k=-K}^{K-1} (g_k^0 g_{k+1}^0 + g_k^1 g_{k+1}^1)}{\sum_{k=-K}^{K-1} (g_k^0)^2 + (g_k^1)^2}, \tag{B6}$$

$$\mathfrak{T}_i = \frac{\sum_{k=-K}^{K-1} (g_k^0 g_{k+1}^1 - g_k^0 g_{k+1}^1)}{\sum_{k=-K}^{K-1} (g_k^0)^2 + (g_k^1)^2}. \tag{B7}$$

Then, we can calculate their derivatives as (recalling again that $g_k^0 = g_{-k}^0$ and $g_k^1 = g_{-k}^1$)

$$\frac{\partial \mathfrak{T}_r}{\partial g_k^a} = \frac{2}{1 + \delta_{k,0}} \left[\frac{(1 - \delta_{k,K}) g_{k+1}^a + g_{k-1}^a - 2\mathfrak{T}_r g_k^a}{\sum_{k=-K}^{k+1} ((g_k^0)^2 + (g_k^1)^2)} \right], \tag{B8}$$

$$\frac{\partial \mathfrak{T}_i}{\partial g_k^a} = \frac{2(-1)^a}{1 + \delta_{k,0}} \left[\frac{(1 - \delta_{k,K}) g_{k+1}^{1-a} - g_{k-1}^{1-a} - 2\mathfrak{T}_i g_k^a}{\sum_{k=-K}^{k+1} ((g_k^0)^2 + (g_k^1)^2)} \right]. \tag{B9}$$

Substituting in for g_k^a, we find that everything precisely cancels when $k \neq K$!

$$\frac{\partial \mathfrak{T}_r}{\partial g_k^0} = -\frac{\partial \mathfrak{T}_i}{\partial g_k^1} = -2\delta_{k,K} \frac{\cos((K+1)\phi)}{\sum_{k=-K}^{k+1} ((g_k^0)^2 + (g_k^1)^2)}, \tag{B10}$$

$$\frac{\partial \mathfrak{T}_i}{\partial g_k^0} = \frac{\partial \mathfrak{T}_r}{\partial g_k^1} = -2\delta_{k,K} \frac{\sin((K+1)\phi)}{\sum_{k=-K}^{k+1} ((g_k^0)^2 + (g_k^1)^2)}. \tag{B11}$$

Our variance is then

$$\begin{aligned}
\mathrm{Var}(\phi) &= \left[\frac{2}{\left(\mathfrak{T}_r^2 + \mathfrak{T}_i^2 \right) \sum_{k=-K}^{k+1} ((g_k^0)^2 + (g_k^1)^2)} \right]^2 \\
&\quad \times \{ \mathrm{Var}[g_K^0] (-\cos(\phi)\sin((K+1)\phi) + \sin(\phi)\cos((K+1)\phi))^2 \\
&\quad + \mathrm{Var}[g_K^1] (\cos(\phi)\cos((K+1)\phi) + \sin(\phi)\sin((K+1)\phi))^2 \} \\
&= \left[\frac{1}{K} \right]^2 \{ \mathrm{Var}[g_K^0] \sin^2(K\phi) + \mathrm{Var}[g_K^1] \cos^2(K\phi) \}. \tag{B12}
\end{aligned}$$

If g_K^a is estimated with N shots, we expect $\mathrm{Var}[g_K^0] = \frac{1}{N}$, and

$$\mathrm{Var}(\phi) \propto \frac{1}{K^2 N}. \tag{B13}$$

As described in section 3.1.1, for multi-round experiments we weight the least-squares inversion as per equation (27). This weighting adjusts the g_k^a values in equations (B8), (B9) so that $\frac{\partial \phi}{\partial g_k^A}$ is no longer zero when $k < K$. The sum over k in equation (B5) then lends an extra factor of K to the variance, reducing it to

$$\mathrm{Var}(\phi) \propto \frac{1}{KN}. \tag{B14}$$

Appendix C. Fourier representation for Bayesian updating

For simplicity, we first consider when the starting state is a simple eigenstate $|\phi_j\rangle$. After each multi-round experiment we would like to update the probability distribution $P(\phi_j = \phi)$, i.e. $P_n(\phi) = \frac{P_{k,\beta}(\mathbf{m} \mid \phi)}{P(\mathbf{m})} P_{n-1}(\phi)$. We will represent the 2π-periodic probability distribution $P_n(\phi)$ by a Fourier series with a small number of Fourier coefficients #freq which are updated after each experiment, that is, we write

$$P(\phi) = p_0 + \sum_{j=1}^{\#\mathrm{freq}-1} (p_{2j-1} \sin(j\phi) + p_{2j} \cos(j\phi)) \equiv \boldsymbol{p}. \tag{C1}$$

We thus collect the coefficients as a #freq-component vector $\boldsymbol{p}$. The Fourier representation has the advantage that integration is trivial i.e. $\int_{-\pi}^{\pi} P(\phi)\mathrm{d}\phi = 2\pi p_0$ so that the probability distribution is easily normalized. In addition, the current estimate $\tilde{\phi}$ is easy to obtain:

$$\tilde{\phi} = \arg(\langle e^{i\phi}\rangle_P) = \arg(p_2 + ip_1). \tag{C2}$$

Another observation is that the Holevo phase variance is easily obtained from this Fourier representation as

$$\mathrm{Var}(P(\phi)) = \frac{1}{|\langle e^{i\phi}\rangle_P|^2} - 1 = \frac{1}{\pi^2(p_2^2 + p_1^2)} - 1. \tag{C3}$$

Note that this is the Holevo phase variance of the posterior distribution of a single simulation instance. By comparison, in figure 4 we have calculated the same quantity over repeat simulations. However, in general we find the two to be equivalent.

The other advantage of the Fourier representation is that a single-round in an experiment is the application of a sparse matrix on $\mathbf{p}$. One has $P(\phi) \to P_{k_r,\beta_r}(m_r|\phi)P(\phi) = \cos^2(k_r\phi/2 + \gamma/2)P(\phi)$, where $\gamma = \beta_r + m_r\pi$ which is equivalent to

$$\mathbf{p} \to \frac{1}{2}\mathbf{p} + \frac{1}{4}\cos(\gamma)M^0(k_r)\mathbf{p} + \frac{1}{4}\sin(\gamma)M^1(k_r)\mathbf{p}. \tag{C4}$$

The coefficients of the update matrices $M^{0,1}(k_r)$ can be simply calculated using the double angle formulae and employing

$$\cos^2(k\phi/2 + \gamma/2)\cos(j\phi)$$
$$= \frac{1}{2}\cos(j\phi) + \frac{1}{4}\cos(\gamma)(\cos((j+k)\phi) + \cos((j-k)\phi))$$
$$+ \frac{1}{4}\sin(\gamma)(\sin((j-k)\phi) - \sin((j+k)\phi)), \tag{C5}$$

and

$$\cos^2(k\phi/2 + \gamma/2)\sin(j\phi)$$
$$= \frac{1}{2}\sin(j\phi) + \frac{1}{4}\cos(\gamma)(\sin((j+k)\phi) + \sin((j-k)\phi))$$
$$+ \frac{1}{4}\sin(\gamma)(\cos((j+k)\phi) - \cos((j-k)\phi)). \tag{C6}$$

The matrices $M^a(n)$ are then calculated from the above equations. When $j > k$, we have

$$[M^0(k)]_{2j+2k,2j} = 1, \quad [M^0(k)]_{2j-2k,2j} = 1,$$
$$[M^0(k)]_{2j+2k-1,2j-1} = 1, \quad [M^0(k)]_{2j-2k-1,2j-1} = 1,$$
$$[M^1(k)]_{2j+2k-1,2j} = -1, \quad [M^1(k)]_{2j-2k-1,2j} = 1,$$
$$[M^1(k)]_{2j+2k,2j-1} = 1, \quad [M^1(k)]_{2j-2k,2j-1} = -1.$$

When $j \leqslant k$, we have to account for the sign change in $\sin((j - k)\phi)$:

$$[M^0(k)]_{j+2k,j} = 1, \quad [M^0(k)]_{2k-2j,2j} = 1,$$
$$[M^0(k)]_{2k-2j-1,2j-1} = -1$$
$$[M^0(k)]_{2k,0} = -2, \quad [M^0(k)]_{4k-1,2k-1} = 1$$
$$[M^1(k)]_{2j+2k-1,2j} = -1, \quad [M^1(k)]_{2k-2j-1,2j} = -1,$$
$$[M^1(k)]_{2j+2k,2j-1} = 1, \quad [M^1(k)]_{2k-2j,2j-1} = -1,$$
$$[M^1(k)]_{2k-1,0} = 2, \quad [M^1(k)]_{4k-1,2k} = 1.$$

For a multi-round experiment with R rounds, one thus applies such sparse matrices to the vector **p** R times. Note that each round with given k_r requires at most k_r more Fourier components, hence an experiment with at most K controlled-U applications adds at most K Fourier components. Thus, when the total number of unitary rotations summed over all experiments $K_{\text{tot}} = \sum_n \sum_r k_r > \#\text{freq}$, our representation of the distribution is no longer accurate. When $K_{\text{tot}} \leqslant \#\text{freq}$ on the other hand, it will be accurate.

C.1. Bayesian updating for multi-eigenvalue starting state

In this section we detail the method by which we store the distributions $P_n^j(\phi_j)$ and $P_n^{\text{red}}(\mathbf{A})$ of equation (31) and perform the Bayesian update of equation (30). We do so by representing the marginal probabilities $P_n^j(\phi_j)$ by a Fourier series with a small number of Fourier coefficients which are updated after each experiment as shown in the previous section. We assume that there are most N_{eig} coefficients $A_j > 0$ and thus $N_{\text{eig}} \phi_j$.

From our independence assumption, individual updates of $P^j(\phi_j)$ may be calculated by integrating out the other unknown variables in equation (30):

$$P_n^j(\phi_j) = \int \left(\prod_{l \neq j} d\phi_l P_{n-1}^l(\phi_l) \right) \int d\mathbf{A} \, P_{n-1}^{\text{red}}(\mathbf{A}) P_{\mathbf{k},\beta}(\mathbf{m}|\phi, \mathbf{A}) P_{n-1}^j(\phi_j). \tag{C7}$$

Expanding the conditional probability of equation (10) and rewriting leads to the form

$$P_n^j(\phi_j) = \frac{1}{P_{\mathbf{k},\beta}(\mathbf{m})} \left(C + B_j \prod_r P_{k_r,\beta_r}(m_r|\phi_j) \right) P_{n-1}^j(\phi_j), \tag{C8}$$

with

$$C = \sum_{k \neq j} B_k \int d\phi_k P_{n-1}^k(\phi_k) \prod_r P_{k_r,\beta_r}(m_r|\phi_k),$$

and $B_j = \int d\mathbf{A} \, P_{n-1}^{\text{red}}(\mathbf{A}) A_j$. Here we have used that $\int d\phi_l P_{n-1}^l(\phi_l) = 1$. One can concisely write B_j as the components of a vector **B**. Computing equation (30) then involves creating an 'update' distribution for each ϕ_j, calculating the integral of each distribution, and then forming the new distribution from a weighted sum from the 'update' distributions.

Calculating the distribution $P_n^{\text{red}}(\mathbf{A})$ is complicated slightly by the restriction that $\sum_j A_j = 1$, $A_j \geqslant 0$, meaning that we cannot assume the distribution of individual A_j terms is uncorrelated. The marginal probability distribution equals

$$P_n^{\text{red}}(\mathbf{A}) = \frac{P_{n-1}^{\text{red}}(\mathbf{A})}{P_{\mathbf{k},\beta}(\mathbf{m})} \sum_j A_j \int d\phi_j P_{n-1}^j(\phi_j) \prod_r P_{k_r,\beta_r}(m_r|\phi_j). \tag{C9}$$

or

$$P_n^{\text{red}}(\mathbf{A}) = \frac{P_{n-1}^{\text{red}}(\mathbf{A})}{P_{\mathbf{k},\beta}(\mathbf{m})} \mathbf{A} \cdot \mathbf{q}_{n-1}, \tag{C10}$$

where the *j*th component $(q_{n-1})_j$ is the integral

$$(q_{n-1})_j = \int \mathrm{d}\phi_j P_{n-1}^j(\phi_j) \prod_r P_{k_r,\beta_r}(m_r|\phi_j).$$ (C11)

As **A** only enters our estimation through the vector $\mathbf{B} = (B_0, \ldots, B_{N_{\mathrm{eig}}})$, we only need approximate this value. Assuming we know the marginal probabilities $P_n(\phi_j)$ for all experiments $n = 1, \ldots, N$, we can estimate **B** after all experiments by the maximum likelihood value $\mathbf{A}^{(\mathrm{max})}$

$$\mathbf{A}_N^{(\mathrm{max})} = \arg\max_{\mathbf{A}} f(\mathbf{A})$$

$$f(\mathbf{A}) = \log\left(P_{\mathrm{prior}}(\mathbf{A}) \prod_{n=1}^N \mathbf{A} \cdot \mathbf{q}_n \right)$$

$$= \log(P_{\mathrm{prior}}(\mathbf{A})) + \sum_{n=1}^N \log(\mathbf{A} \cdot \mathbf{q}_n).$$

Evaluating this equation for up $N = 1000$ experiments, taking #freq $= 10\,000$ frequency components of $N_{\mathrm{eig}} = 2$ eigenvalues takes less than a second on a laptop using a method such as sequential least-squares programming [41]. However, beyond this it becomes fairly computationally intensive. Thus, after $N > 100$ experiments have been performed, we switch to a local optimization method. We determine the optimal $\mathbf{B}_n$ after *n* experiments from its prior value $\mathbf{B}_{n-1}$ via a single step of an approximate Newton's method, that is, we take

$$\mathbf{B}_n = \mathbf{B}_{n-1} - \Pi[\mathbf{H}^{-1}(f(\mathbf{B}_{n-1}))\, (\vec{\nabla}f)(\mathbf{B}_{n-1})],$$

where $\vec{\nabla}f(\mathbf{A})$ is the first derivative of *f* at **A** and **H** is the Hessian matrix of *f*, i.e. $H_{ij} = \partial_{A_i}\partial_{A_j} f(\mathbf{A})$. Here $\Pi[\mathbf{A}]$ is the projector onto the plane $\sum_{j=0}^{N_{\mathrm{eig}}} A_j = 1$ so that the update preserves the normalization. We have

$$\partial_{A_i} f(\mathbf{A}) = \frac{\partial_{A_i} P_{\mathrm{prior}}(\mathbf{A})}{P_{\mathrm{prior}}(\mathbf{A})} + \sum_{n=1}^N \frac{(q_n)_i}{\mathbf{A} \cdot \mathbf{q}_n}$$

We approximate the second term for each step as coming from only from the added term, i.e.

$$\vec{\nabla}f(\mathbf{B}_{n-1}) \approx \frac{\mathbf{q}_n}{\mathbf{B}_{n-1} \cdot \mathbf{q}_n}.$$ (C12)

The Hessian equals

$$H_{ij}(f(\mathbf{A})) = -\sum_{n=1}^N \frac{(q_n)_i (q_n)_j}{(\mathbf{A} \cdot \mathbf{q}_n)^2},$$ (C13)

but we approximate this at the *n*th step

$$H_{ij}^{(n)}(f(\mathbf{B}_{n-1})) \approx H_{ij}^{(n-1)} - \frac{(q_n)_i (q_n)_j}{(\mathbf{B}_n \cdot \mathbf{q}_n)^2}.$$ (C14)
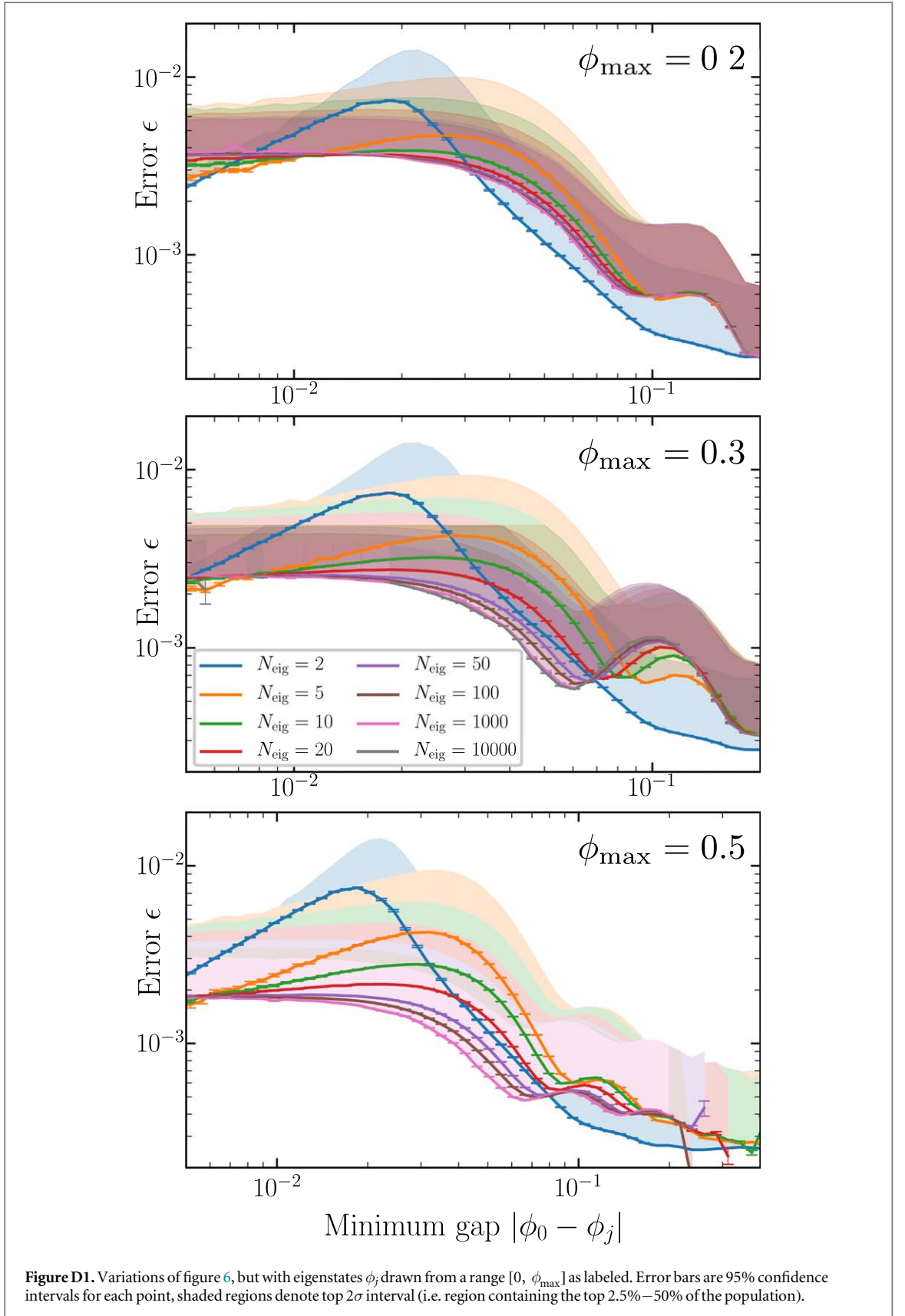
This approximation allows *H* to be updated without summing over each experiment.

With the above implemented, we observe that our estimator can process data from $N = 10\,000$ experiments to estimate $N_{\mathrm{eig}} = 2$ eigenvalues with $N = 20\,000$ Fourier components within approximately two minutes on a laptop. Unfortunately, this method scales as N^2, as the number of frequencies required for accurate estimation grows as the total number of unitaries applied.

As the mean, variance and integration calculations only require the first few frequencies of the distribution, it may be possible to reduce this cost by finding approximation techniques for higher frequency components.

Appendix D. Convergence of the (noiseless) time-series analysis in case of multiple eigenvalues

In this section we present an expansion of figure 6, namely figure D1, by drawing the spurious eigenvalues ϕ_j from a range closer to the target eigenvalue ϕ_0. This negates the drop in estimation error observed in figure 6 that was caused by the majority of eigenvalues lying in region (c) of figure 5. We observe that for certain gaps δ, multiple eigenvalues confined to a thin region $[\delta, \phi_{\mathrm{max}}]$ can have a worse effect on our ability to estimate ϕ_0 than that of a single eigenvalue at δ. However, this loss in accuracy does not get critically worse with the addition of more eigenvalues. Neither is it worse than the worst-possible estimation with two eigenvalues.

Figure D1. Variations of figure 6, but with eigenstates ϕ_j drawn from a range $[0, \phi_{max}]$ as labeled. Error bars are 95% confidence intervals for each point, shaded regions denote top 2σ interval (i.e. region containing the top $2.5\% - 50\%$ of the population).

Appendix E. Details of realistic simulation

In this appendix we give details of the method for the realistic noisy circuit simulation of section 5.2. Our density-matrix simulator is fairly limited in terms of qubit number, and so we opt to simulate H_4 in the STO-3G basis. This molecule has 8 spin orbitals and thus requires 9 qubits for the QPE simulation (with the additional

Table E1. Parameters of the H_4 geometries used in the text. Terms are described in appendix E. $\|\mathcal{H}\|_F = \sqrt{\text{Trace}[\mathcal{H}^\dagger \mathcal{H}]/2^{n_{sys}}}$.

d_x [Å]	d_y [Å]	E_0	A_0	$\delta/\|\mathcal{H}\|_F$	$\epsilon_{\text{Trotter}}/\|\mathcal{H}\|_F$
0.4	0.5	-0.26	0.98	0.09	3.7×10^{-4}
0.6	0.7	-1.46	0.94	0.17	3.1×10^{-3}
0.8	0.9	-1.84	0.88	0.24	0.016
1.0	1.1	-1.96	0.80	0.23	0.017
1.2	1.3	-1.98	0.71	0.18	0.013
1.6	1.7	-1.94	0.55	0.09	6.0×10^{-3}
0.2	1.8	0.32	0.996	0.67	2.0×10^{-4}
0.4	1.6	-1.80	0.993	1.14	2.6×10^{-3}
0.6	1.4	-2.15	0.98	1.27	0.014
0.8	1.2	-2.09	0.96	0.73	0.021

Table E2. Standard parameters of error models used in density matrix simulation. Table adapted from [37] with all parameters taken from therein (with the exception of the $1/f$ flux noise, which is made incoherent as described in text).

Parameter	Symbol	Standard value	Scaling
Qubit relaxation time	T_1	30 μs	λ
Qubit dephasing time	T_2	30 μs	λ
Single-qubit gate time	T_{sq}	20 ns	1
Two-qubit gate time	T_{2q}	40 ns	1
In-axis rotation error	p_{axis}	10^{-4}	λ^{-1}
In-plane rotation error	p_{plane}	5×10^{-4}	λ^{-1}
Incoherent flux noise	A	$(1\mu\Phi_0)^2$	λ^{-1}
Measurement time	T_{meas}	300 ns	1
Depletion time	T_{dep}	300 ns	1
Readout infidelity	ϵ_{RO}	5×10^{-3}	λ^{-1}
Measurement induced decay	$p_{\text{d,i}}$, $p_{\text{d,f}}$	0.005, 0.001 5	λ^{-1}

qubit being the ancilla). We choose 10 rectangular molecular geometries for the H_4 system, parametrized by a horizontal distance d_x and a vertical distance d_y (i.e. the four H atoms are in the positions $(\pm d_x/2, \pm d_y/2, 0)$). We calculate the Hartree–Fock and full-CI solutions to the ground state using the psi4 package [38] with the openfermion interface [9]. This allows to calculate the true ground state energy E_0 for each geometry, and the overlap A_0 between the ground state and the Hartree–Fock state, which we choose as our starting state $|\Psi\rangle$. Due to symmetry and particle number conservation, $|\Psi\rangle$ has non-zero overlap with only 8 eigenstates of the full-CI solution, separated from the ground state by a minimum gap δ. (When $d_x = d_y$, the true ground state of H_4 is actually orthogonal to the Hartree–Fock state, and so we do not include any such geometries in our calculation.) The full error in our calculation of the energy (at a fixed geometry) is then a combination of three separate contributions: basis set error (i.e. from the choice of orbitals), Trotter error, and the estimator error studied in this work (which includes error from experimental noise). The Trotter error $\epsilon_{\text{Trotter}}$ is reasonably large due to our use of only the first-order Suzuki-Trotter approximation $U = \prod_i e^{-iH_i t} \approx e^{-i\mathcal{H}t}$. Higher-order Suzuki-Trotter expansions require longer quantum circuits, which in turn increase the estimator error from experimental noise. Balancing these two competing sources of error is key to obtaining accurate calculations and a clear target for future study. In table E1, we list some parameters of interest for each studied geometry. We normalize the gap and the Trotter error by the Frobenius norm $\|\mathcal{H}\|_F = \sqrt{\text{Trace}[\mathcal{H}^\dagger \mathcal{H}]/2^{n_{sys}}}$, as we chose an evolution time $t = 1/\|\mathcal{H}\|_F$, making this the relevant scale for comparison with scaling laws and errors calculated in the text.

E.1. Error model and error parameters
Throughout this work we simulate circuits using an error model of superconducting qubits first introduced in [37]. This captures a range of different error channels with parameters either observed in experimental data or estimated via theory calculations. All error channels used are listed in table E2, and we will now describe them in further detail.

Transmon qubits are dominated primarily by decoherence, which is captured via T_1 and T_2 channels [4]. Typical T_1 and T_2 times in state-of-the-art devices are approximately $10-100$ μs. As other error parameters are derived from experimental results on a device with $T_1 = T_2 \approx 30$ μs, we take these as a base set of parameters [42, 43]. Single-qubit gates in transmon qubits incur slight additional dephasing due to inaccuracies or fluctuations in microwave pulses. We assume such dephasing is Markovian, in which case it corresponds to a shrinking of the Bloch sphere along the axis of rotation by a value $1 - p_{\text{axis}}$, and into the perpendicular plane by a value $1 - p_{\text{plane}}$. We take typical values for these parameters as $p_{\text{axis}} = 10^{-4}$, $p_{\text{plane}} = 5 \times 10^{-4}$ [37].

Two-qubit gates in transmon qubits incur dephasing due to $1/f$ flux noise. Assuming that the phase in an ideal C-phase gate $G = \text{diag}(1, 1, 1, e^{i\phi}))$ is controlled by adjusting the time of application, this suggests a model for the applied gate which is

$$G(\delta_{\text{flux}}) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & e^{i\delta_{\text{flux}}\phi} & 0 \\ 0 & 0 & 0 & e^{i(1+\delta_{\text{flux}}/2)\phi} \end{pmatrix}, \tag{E1}$$

where δ_{flux} is drawn from a normal distribution around 0 with standard deviation σ_{flux}. One can estimate $\sigma_{\text{flux}} \approx 0.01$ rad for a typical gate length of 40 ns [37]. The noise is in general non-Markovian, as δ_{flux} fluctuates on longer timescale than a single gate. However, to make the simulation tractable, we approximate it as Markovian. The Pauli transfer matrix of this averaged channel [44] reads

$$\Lambda[G] = \int d\delta_{\text{flux}} P(\delta_{\text{flux}})\Lambda[G(\delta_{\text{flux}})], \tag{E2}$$

where the Pauli transfer matrix of a channel G is given by $\Lambda[G]_{i,j} = \text{Tr}[\sigma_i G \sigma_j]$.

During qubit readout, we assume that the qubit is completely dephased and projected into the computational basis. We then allow for a $T_{\text{meas}} = 300$ ns period of excitation and de-excitation (including that from T_1-decay), during which the qubit state is copied onto a classical bit. This copying is also assumed to be imperfect, with a probability ϵ_{RO} of returning the wrong result. The qubit then has an additional $T_{\text{dep}} = 300$ ns waiting period before it may participate in gates again (to allow resonator depletion [42]), over which additional excitation and de-excitation may occur. Though simple, this description is an accurate model of experimental results. Typically experiments do not observe measurement-induced excitation to the $|1\rangle$ state, but do observe measurement-induced decay [37]. Typical values of such decay are 0.005 prior to the copy procedure, and 0.015 after.

Though reasonably accurate, this error model does fail to capture some details of real experimental systems. In particular, we do not include leakage to the $|2\rangle$ state, which is a dominant source of two-qubit gate error. Furthermore, we have not included cross-talk between qubits.

To study the effect of changing noise levels while staying as true as possible to our physically-motivated model, we scale our noise parameters by a dimensionless parameter λ such that the contribution from each error channel to the simulation remains constant. In table E2 we show the power of λ that each error term is multiplied by during this scaling. We report $T_{\text{err}} := T_1 = T_2$ in the main text instead of λ to make connection to parameters regularly reported in experimental works.

References

[1] Wocjan P and Zhang S 2006 Several natural BQP-complete problems arXiv:quant-ph/0606179
[2] Preskill J 2018 Quantum computing in the NISQ era and beyond *Quantum* **2** 79
[3] Griffiths R B and Niu C-S 1996 Semiclassical Fourier transform for quantum computation *Phys. Rev. Lett.* **76** 3228–31
[4] Nielsen M A and Chuang I L 2000 *Quantum Computation and Quantum Information* (Cambridge: Cambridge University Press)
[5] Kitaev A Y 1995 Quantum measurements and the Abelian stabilizer problem arXiv:quant-ph/9511026
[6] Kimmel S, Low G H and Yoder T J 2015 Robust calibration of a universal single-qubit gate set via robust phase estimation *Phys. Rev. A* **92** 062315
[7] Knill E, Ortiz G and Somma R D 2007 Optimal quantum measurements of expectation values of observables *Phys. Rev. A* **75** 012328
[8] Svore K M, Hastings M B and Freedman M 2013 Faster phase estimation *Quant. Inf. Comput.* **14** 306–28 arXiv:1304.0741
[9] McClean J R *et al* 2017 OpenFermion: the electronic structure package for quantum computers arXiv:1710.07629
[10] Wiebe N and Granade C 2016 Efficient Bayesian phase estimation *Phys. Rev. Lett.* **117** 010503
[11] Potts D and Tasche M 2013 Parameter estimation for nonincreasing exponential sums by Prony-like methods *Linear Algebr. Appl.* **439** 1024–39
[12] Onorati E, Werner A H and Eisert J 2018 Randomized benchmarking for individual quantum gates arXiv:1811.11775
[13] Xiang G Y *et al* 2011 Entanglement-enhanced measurement of a completely unknown optical phase *Nat. Photon.* **5** 43–7
[14] Daryanoosh S *et al* 2018 Experimental optical phase measurement approaching the exact Heisenberg limit *Nat. Commun.* **9** 4606
[15] Berni A A *et al* 2015 Ab-initio quantum-enhanced optical phase estimation using real-time feedback control *Nat. Photon.* **9** 577–81
[16] Zhou X-Q, Kalasuwan P, Ralph T C and O'Brien J L 2013 Calculating unknown eigenvalues with a quantum algorithm *Nat. Photon.* **7** 02
[17] Paesani S *et al* 2017 Experimental Bayesian quantum phase estimation on a silicon photonic chip *Phys. Rev. Lett.* **118** 100503
[18] Santagati R *et al* 2018 Witnessing eigenstates for quantum simulation of Hamiltonian spectra *Sci. Adv.* **4** eaap9646

[19] Abrams D S and Lloyd S 1999 Quantum algorithm providing exponential speed increase for finding eigenvalues and eigenvectors *Phys. Rev. Lett.* **83** 5162–5

[20] Whitfield J D, Biamonte J and Aspuru-Guzik A 2011 Simulation of electronic structure Hamiltonians using quantum computers *Mol. Phys.* **109** 735–50

[21] Berry D W *et al* 2015 Simulating Hamiltonian dynamics with a truncated Taylor series *Phys. Rev. Lett.* **114** 090502

[22] Low G H and Chuang I L 2017 Optimal Hamiltonian simulation by quantum signal processing *Phys. Rev. Lett.* **118** 010501

[23] Berry D W *et al* 2018 Improved techniques for preparing eigenstates of fermionic Hamiltonians *npj Quant. Inf.* **4** 22

[24] Poulin D *et al* 2018 Quantum algorithm for spectral measurement with a lower gate count *Phys. Rev. Lett.* **121** 010501

[25] Berry D W and Childs A M 2012 Black-box Hamiltonian simulation and unitary implementation *Quant. Inf. Comput.* **12** 29–62

[26] Childs A M 2010 On the relationship between continuous- and discrete-time quantum walk *Commun. Math. Phys.* **294** 581–603

[27] Kitaev A Y, Shen A H and Vyalyi M N 2002 Classical and quantum computation *Graduate Studies in Mathematics* (Providence, RI: American Mathematical Society)

[28] Bravyi S, Gosset D, Koenig R and Temme K Approximation algorithms for quantum many-body problems arXiv:1808.01734

[29] Reiher M, Wiebe N, Svore K M, Wecker D and Troyer M 2017 Elucidating reaction mechanisms on quantum computers *Proc. Nat. Acad. Sci.* **114** 7555–60

[30] Peruzzo A *et al* 2014 A variational eigenvalue solver on a photonic quantum processor *Nat. Commun.* **5** 4213

[31] Gambetta J, Blais A, Schuster D I, Wallraff A, Frunzio L, Majer J, Devoret M H, Girvin S M and Schoelkopf R J 2006 Qubit-photon interactions in a cavity: measurement-induced dephasing and number splitting *Phys. Rev.* A **74** 042318

[32] Higgins B L *et al* 2009 Demonstrating Heisenberg-limited unambiguous phase estimation without adaptive measurements *New J. Phys.* **11** 073023

[33] Terhal B M and Weigand D 2016 Encoding a qubit into a cavity mode in circuit QED using phase estimation *Phys. Rev.* A **93** 012315

[34] Wiebe N, Granade C and Cory D G 2015 Quantum bootstrapping via compressed quantum Hamiltonian learning *New J. Phys.* **17** 022005

[35] Babbush R and Ding N 2017 Quantum phase estimation of multiple eigenvalues https://patents.google.com/patent/WO2017116446A1/

[36] Wiebe N, Granade C, Ferrie C and Cory D 2014 Hamiltonian learning and certification using quantum resources *Phys. Rev. Lett.* **112** 190501

[37] O'Brien T E, Tarasinski B and DiCarlo L 2017 Density-matrix simulation of small surface codes under current and projected experimental noise *npj Quant. Inf.* **3** 39

[38] Parrish R M *et al* 2017 Psi4 1.1: an open-source electronic structure program emphasizing automation, advanced libraries, and interoperability *J. Chem. Theory Comput.* **13** 3185–97

[39] Suzuki M 1976 Generalized Trotter formula and systematic approximants of exponential operators and inner derivations with applications to many-body problems *Commun. Math. Phys.* **51** 183–90

[40] Hastings M B, Wecker D, Bauer B and Troyer M 2015 Improving quantum algorithms for quantum chemistry *Quant. Inf. Comput.* **15** 0001–21 arXiv:1403.1539

[41] Kraft D 1994 Algorithm 733: TOMP-fortran modules for optimal control calculations *ACM Trans. Math. Softw.* **20** 262–81

[42] Bultink C C *et al* 2016 Active resonator reset in the nonlinear dispersive regime of circuit QED *Phys. Rev. Appl.* **6** 034008

[43] Rol M A *et al* 2017 Restless tuneup of high-fidelity qubit gates *Phys. Rev. Appl.* **7** 041001

[44] Chow J M *et al* 2012 Complete universal quantum gate set approaching fault-tolerant thresholds with superconducting qubits *Phys. Rev. Lett.* **109** 060501