Optimal quantum measurements of expectation values of observables

Emanuel Knill*
Mathematical and Computational Sciences Division, National Institute of Standards and Technology, Boulder, Colorado 80305, USA

Gerardo Ortiz†
Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA
and Department of Physics, Indiana University, Bloomington, Indiana 47405, USA

Rolando D. Somma‡
Physics Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA
(Received 3 August 2006; published 24 January 2007)

Experimental characterizations of a quantum system involve the measurement of expectation values of observables for a preparable state $|\psi\rangle$ of the quantum system. Such expectation values can be measured by repeatedly preparing $|\psi\rangle$ and coupling the system to an apparatus. For this method, the precision of the measured value scales as $\frac{1}{\sqrt{N}}$ for N repetitions of the experiment. For the problem of estimating the parameter ϕ in an evolution $e^{-i\phi H}$, it is possible to achieve precision $\frac{1}{N}$ [the quantum metrology limit; see Giovannetti *et al.*, Phys. Rev. Lett. **96**, 010401 (2006)] provided that sufficient information about H and its spectrum is available. We consider the more general problem of estimating expectations of operators A with minimal prior knowledge of A. We give explicit algorithms that approach precision $\frac{1}{N}$ given a bound on the eigenvalues of A or on their tail distribution. These algorithms are particularly useful for simulating quantum systems on quantum computers because they enable efficient measurement of observables and correlation functions. Our algorithms are based on a method for efficiently measuring the complex overlap of $|\psi\rangle$ and $U|\psi\rangle$, where U is an implementable unitary operator. We explicitly consider the issue of confidence levels in measuring observables and overlaps and show that, as expected, confidence levels can be improved exponentially with linear overhead. We further show that the algorithms given here can typically be parallelized with minimal increase in resource usage.

DOI: 10.1103/PhysRevA.75.012328 PACS number(s): 03.67.Mn, 03.65.Ud

I. INTRODUCTION

When we extract information from a physical system, fundamental physical limits on the achievable precision are set by uncertainty relations such as Heisenberg's uncertainty principle. The goal of quantum metrology is to measure properties of states of quantum systems as precisely as possible given available resources. Typically, these properties are determined by experiments that involve repeated preparation of a quantum system in a state ρ followed by a measurement. The property is derived from the measurement outcomes. Because the repetitions are statistically independent, the precision with which the property is obtained scales as $\frac{1}{\sqrt{N}}$, where N is the number of preparations performed. This is known as the standard quantum limit or the shot-noise limit, and it is associated with a purely classical statistical analysis of errors. It has been shown that in many cases of interest, the precision can be improved to $\frac{1}{N}$ by using the same resources, but with initial states entangled over multiple instances of the quantum system, or by preserving quantum coherence from one experiment to the next. It is known that it is usually not possible to attain a precision that scales better than $\frac{1}{N}$. (See [1] for a review of quantum-enhanced measurements.) A setting where this limit can be achieved is the parameter estimation problem, where the property is given by the parameter ϕ in an evolution $e^{-i\phi H}$ for a known Hamiltonian H [2], which captures some common measurement problems. The standard method for determining ϕ requires the ability to apply $e^{-i\phi H}$ and to prepare and measure an eigenstate of H with known eigenvalue. If it is not possible to prepare such an eigenstate or if we wish to determine expectations with respect to arbitrary states, this method fails. Here we are interested in the more general and physically important expectation estimation problem, where the property to be determined is an expectation $\langle A\rangle=\mathrm{tr}(A\rho)$ of an observable (Hermitian operator) or unitary A, for a possibly mixed state ρ. Both A and ρ are assumed to be experimentally sufficiently controllable, but other than a bound on the eigenvalues of A or their tail distribution, no other properties of A or ρ need to be known. In particular, we need not be able to prepare eigenstates of A or know the spectrum of A. The parameter estimation problem is a special instance of the expectation estimation problem. Parameter estimation reduces to the problem of determining $\mathrm{tr}(e^{-i\phi H}|\psi\rangle\langle\psi|)$ for $|\psi\rangle$ an eigenstate of H with nonzero eigenvalue. We show that for solving the expectation estimation problem, precision scalings of $\frac{1}{N^{1-\alpha}}$ for arbitrarily small $\alpha>0$ can be achieved with sequential algorithms and the algorithms can be parallelized with minimal additional resources.

Our motivation for this work is the setting of quantum physics simulations on quantum computers. This is one of

*Electronic address: knill@boulder.nist.gov
†Electronic address: g_ortiz@lanl.gov, ortizg@indiana.edu
‡Electronic address: somma@lanl.gov

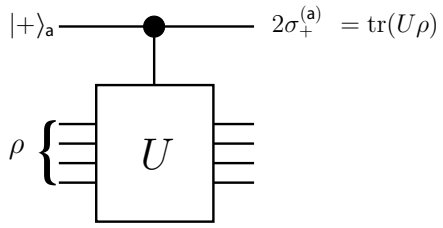

FIG. 1. Quantum network for the one-ancilla algorithm to measure $\langle U\rangle=\mathrm{tr}(U\rho)$ with $|+\rangle_a=(|0\rangle_a+|1\rangle_a)/\sqrt{2}$ in the logical basis. The desired expectation is given by $\mathrm{tr}(U\rho)=\langle 2\sigma_+^{(a)}\rangle=\langle\sigma_x^{(a)}\rangle+i\langle\sigma_y^{(a)}\rangle$, where $\langle\sigma_x^{(a)}\rangle$ and $\langle\sigma_y^{(a)}\rangle$ are the expectations of the Pauli matrices $\sigma_x^{(a)}$ and $\sigma_y^{(a)}$ for the final state, which are estimated by repeating the experiment and measuring either $\sigma_x^{(a)}$ or $\sigma_y^{(a)}$ on the control (ancilla) qubit labelled a. Because these measurements have ±1 as possible outcomes, their statistics are determined by the binomial distribution.

the most promising applications of quantum computing [3] and enables a potentially exponential speedup for the correlation function evaluation problem [4–6]. The measurement of these correlation functions reduces to the measurement of the expectation of an operator for one or more states. Because the measurement takes place within a scalable quantum computer, the operators and states are manipulatable via arbitrarily low-error quantum gates. The quantum computational methods that have been described for the determination of these expectations have order $\frac{1}{\sqrt{N}}$ precision. An example is the one-ancilla algorithm for measuring $\langle U\rangle=\mathrm{tr}(U\rho)$ for unitary U described in [5,7,8], which applies U conditional on an ancilla a prepared in a superposition state (Fig. 1). Improving the precision without special knowledge of the operator or state requires more sophisticated algorithms.

Here we give quantum algorithms based on phase and amplitude estimation [9,10] to improve the resource requirements needed to achieve a given precision. We begin by giving an "overlap estimation" algorithm (OEA) for determining the amplitude and phase of $\mathrm{tr}(U\rho)$ for U unitary. We assume that quantum procedures for preparing ρ from a standard initial state and for applying U are known and that it is possible to reverse these procedures. We determine the number of times, N, that these procedures are used to achieve a goal precision p and show that N is of order $1/p$. To determine $\mathrm{tr}(A\rho)$ for observables A not expressible as a small sum of unitary operators, we assume that it is possible to evolve under A. This means that we can apply e^{-iAt} for positive times t. The OEA can be used to obtain $\mathrm{tr}(A\rho)t\approx i[\mathrm{tr}(e^{-iAt}\rho)-1]$ for small t. The problem of how to measure $\mathrm{tr}(A\rho)$ with precision p requires determining $\mathrm{tr}(e^{-iAt}\rho)$ with precision better than pt and choosing t small enough that the error in the approximation does not dominate. We solve this problem by means of an "expectation estimation" algorithm (EEA) with minimal additional knowledge on the eigenvalue distribution of A. For this situation, the relevant resources are not only the number N of uses of e^{-iAt} and of the state preparation algorithm, but also the total time T of evolution under A. We show that to achieve a goal precision p, N and T are of order $1/(p^{1+\alpha})$ and $1/p$, respectively, with $\alpha>0$ arbitrarily small. The term α in the resource bound is due partly to the tail distribution of the eigenvalues of A with respect to ρ. When it is known that ρ is an eigenstate of A, so that the distribution is a delta function, then we have $\alpha=0$. This applies to the parameter estimation problem. In the case where A is unbounded, α is still arbitrarily small if the tail distribution is exponentially decaying. But if only small moments of A can be bounded, in which case the best bound on the tail distribution decays polynomially, then α becomes finite.

It is important to properly define the meaning of the term "precision." Here, when we say that we are measuring $\mathrm{tr}(A\rho)$ with precision p, we mean that the probability that the measured value a_{meas} is within p of $\mathrm{tr}(A\rho)$ is bounded below by a constant $c>0$. In other words, the "confidence level" that $a_{\mathrm{meas}}-p\leqslant\mathrm{tr}(A\rho)\leqslant a_{\mathrm{meas}}+p$ is at least c. Thus $a_{\mathrm{meas}}\pm p$ defines "confidence bounds" of the measurement for confidence level c. One interpretation of confidence levels is that if the measurement is independently repeated, the fraction of times the measured value is within the confidence bound is at least the confidence level. For measurement values a_{meas} that have an (approximately) Gaussian distribution, it is conventional to use $c=0.68$ to identify the precision p with the standard deviation. In this case, the confidence level that the measurement outcome is within xp can be bounded by $\mathrm{erf}(x/\sqrt{2})$, where $\mathrm{erf}(y)$ is the error function, $\mathrm{erf}(x/\sqrt{2})\geqslant 1-e^{-x^2/2}$. This bound is often too optimistic, which is one reason to specify confidence levels explicitly. This becomes particularly important in our use of the "phase estimation" algorithm (PEA), whose standard version [9] has confidence levels that converge slowly toward 1 as x goes to infinity. Because of these issues, our algorithms are stated so that they solve the problem of determining $\mathrm{tr}(A\rho)$ with precision p and confidence level c, where p and c are specified at the beginning. This requires that the resource usage be parametrized by both p and c, and we show that the resource usage grows by a factor of order $|\log(1-c)|$ to achieve high confidence levels c.

An important problem in measuring properties of quantum systems is how well the measurement can be parallelized with few additional resources. The goal of parallelizing is to minimize the time for the measurement by using more parallel resources. Ideally, the time for the measurement is independent of the problem. Typically we are satisfied if the time grows at most logarithmically. It is well known that for the parameter estimation problem, one can readily parallelize the measurement by exploiting entanglement in state preparation [11]. That this is still possible for the OEA and EEA given here is not obvious. In fact, we show that there are cases where parallelization either involves a loss of precision or requires additional resources. However, the entanglement method for parallelizing measurements works for expectation estimation and for overlap estimation when $|\mathrm{tr}(U\rho)|$ is not close to 1.

Our algorithms are well suited for quantum computers where the contribution of noisy gates to simulation error is smaller than the desired precision. In principle, this condition can be met at the cost of a polylogarithmic overhead using fault-tolerantly implemented qubits and gates [12]. Thus, after taking into account this overhead, our algorithms are still

more efficient than independently repeated direct measurements when high precision is desired. If low precision suffices or when using quantum devices with limited quantum control or without fault tolerance, which algorithm is best requires further analysis.

II. OVERLAP ESTIMATION

Let U be a unitary operator and ρ a state of quantum system S. We assume that we can prepare ρ and apply U to any quantum system S' that is equivalent to S. Both the preparation procedure and U must be reversible. In addition, we require that the quantum systems be sufficiently controllable and that U can be applied conditionally (see below). We use labels to clarify which quantum system is involved. Thus, $\rho^{(\mathsf{S}')}$ is the state ρ of system S' and $U^{(\mathsf{S}')}$ is U acting on system S'. This notation allows us to describe the preparation of ρ and the application of U in parallel on multiple quantum systems.

When we say that we can prepare ρ, we mean that we can do this fully coherently. That is, we have access to a unitary operator $V^{(\mathsf{SE})}$ that can be applied to a standard initial state $|0\rangle_{\mathsf{SE}}$ of S and an ancillary system E (environment) such that $\rho^{(\mathsf{S})}=\mathrm{tr}_{\mathsf{E}}[V^{(\mathsf{SE})}|0\rangle_{\mathsf{SE}}{}^{\mathsf{SE}}\langle 0|(V^{(\mathsf{SE})})^{\dagger}]$. The state $V^{(\mathsf{SE})}|0\rangle_{\mathsf{SE}}$ is a so-called purification of $\rho^{(\mathsf{S})}$. For our purposes and without loss of generality, we can assume that ρ is pure by merging systems E and S and letting unitaries act on the merged system. With this simplification we can write $\rho=|\psi\rangle\langle\psi| = V|0\rangle\langle 0|V^{\dagger}$ and use $\mathsf{S},\mathsf{S}',\ldots$ to refer to equivalent merged systems. The goal of the OEA is now to estimate the overlap $\langle\psi|U|\psi\rangle$ of $|\psi\rangle$ with $U|\psi\rangle$.

The OEA and EEA require that S be sufficiently controllable. In particular, we require that it be possible to couple S to ancilla qubits and to implement conditional selective sign changes of $|0\rangle_{\mathsf{S}}$. Let $P_0^{(\mathsf{S})}=\mathbf{I}^{(\mathsf{S})}-2|0\rangle_{\mathsf{S}}^{\mathsf{S}}\langle 0|$ be the selective sign change of $|0\rangle_{\mathsf{S}}$, with $\mathbf{I}^{(\mathsf{S})}$ the identity (or no-action) operator. If an ancilla (control) qubit is labeled a, an instance of the conditional selective sign change is defined by

$$ {}^{c}P_0^{(\mathsf{aS})} = |0\rangle_{\mathsf{a}}^{\mathsf{a}}\langle 0|\mathbf{I}^{(\mathsf{S})} + |1\rangle_{\mathsf{a}}^{\mathsf{a}}\langle 1|P_0{}^{(\mathsf{S})}. \tag{1}$$

If S consists of qubits and $|0\rangle_{\mathsf{S}}$ is the usual starting state with all qubits in logical state $|0\rangle$, then this is essentially a many-controlled sign flip and has efficient implementations [13].

As mentioned above, for the OEA we require that U can be applied conditionally. This means that the unitary operator

$$ {}^{c}U^{(\mathsf{aS})} = |0\rangle_{\mathsf{a}}^{\mathsf{a}}\langle 0|\mathbf{I}^{(\mathsf{S})} + |1\rangle_{\mathsf{a}}^{\mathsf{a}}\langle 1|U^{(\mathsf{S})} \tag{2}$$

is available for use. When U is associated with an evolution simulated on a quantum computer, this is no problem since all quantum gates are readily "conditionalized" [13]. Nevertheless, we note that ${}^{c}U$ is not required if only the amplitude $|\langle\psi|U|\psi\rangle|$ of $\langle\psi|U|\psi\rangle$ is needed.

The "amplitude estimation" algorithm (AEA) [10] can almost immediately be applied to obtain $|\langle\psi|U|\psi\rangle|$. To accomplish our goals we need to adapt it for arbitrarily prepared states and use a version that avoids the complexities of the full quantum Fourier transform [14]. Before we describe and analyze the version of the AEA needed here, we show how the OEA uses it to estimate the phase and amplitude of $\langle\psi|U|\psi\rangle$. Let $\mathrm{AE}(U,|\psi\rangle,p)$ be the estimate of $|\langle\psi|U|\psi\rangle|$ obtained by the AEA for goal precision p. (We specify the meaning of the precision parameter below.)

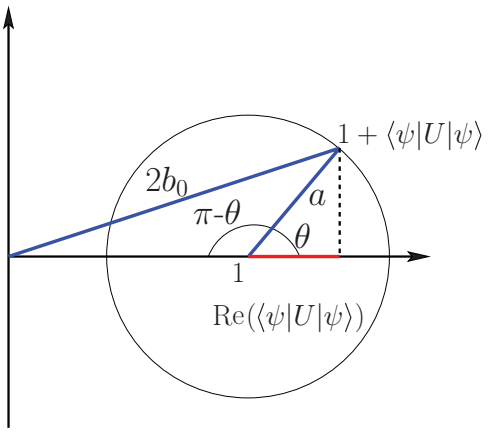

FIG. 2. (Color online) Geometrical construction for computing $\mathrm{Re}(\langle\psi|U|\psi\rangle)$ from $a=|\langle\psi|U|\psi\rangle|$ and $2b_0=|(1+\langle\psi|U|\psi\rangle)|$. According to the law of cosines, $(2b_0)^2=a^2+1+2a\cos(\theta)$, and we have $\mathrm{Re}(\langle\psi|U|\psi\rangle)=a\cos(\theta)=[(2b_0)^2-a^2-1]/2$.

Overlap estimation algorithm. Given are U, $|\psi\rangle$ (in terms of a preparation unitary $V:|0\rangle\mapsto|\psi\rangle$), and the goal precision p. An estimate of $\langle\psi|U|\psi\rangle$ is to be returned.

(1) Obtain $a=\mathrm{AE}(U,|\psi\rangle,p/4)$, so that a is an estimate of $|\langle\psi|U|\psi\rangle|$ with precision $p/4$.

(2) Obtain $b_0=\mathrm{AE}({}^{c}U^{(\mathsf{aS})},|+\psi\rangle_{\mathsf{aS}}=|+\rangle_{\mathsf{a}}|\psi\rangle,p/16)$.

Note that ${}^{\mathsf{aS}}\langle+\psi|{}^{c}U^{(\mathsf{aS})}|+\psi\rangle_{\mathsf{aS}}=(1+\langle\psi|U|\psi\rangle)/2$.

(3) Obtain $b_{\pi/2}=\mathrm{AE}(e^{i\sigma_z^{(\mathsf{a})}\pi/4}\,{}^{c}U^{(\mathsf{aS})},|+\psi\rangle_{\mathsf{aS}},p/16)$.

Note that ${}^{\mathsf{aS}}\langle+\psi|e^{i\sigma_z^{(\mathsf{a})}\pi/4}\,{}^{c}U^{(\mathsf{aS})}|+\psi\rangle_{\mathsf{aS}}=e^{i\pi/4}(1-i\langle\psi|U|\psi\rangle)/2$.

(4) Estimate the phase θ of $\langle\psi|U|\psi\rangle$ by computing the argument of the complex number y defined by

$$\mathrm{Re}(y) = (4b_0^2 - a^2 - 1)/2,$$

$$\mathrm{Im}(y) = (4b_{\pi/2}^2 - a^2 - 1)/2. \tag{3}$$

If a, b_0, and $b_{\pi/2}$ were the exact values of the amplitudes estimated by the three instances of the AEA, then we would have $y=\langle\psi|U|\psi\rangle$. For example, the formula for $\mathrm{Re}(y)$ may be obtained by geometrical reasoning, as shown in Fig. 2.

(5) Estimate $\langle\psi|U|\psi\rangle$ as $e^{i\theta}a$. The reason for not using y directly is that if the overlap has amplitude near 1, then the error in the amplitude of y can be substantially larger than the error in a. (This is because of the way we estimate y using a PEA; see below.)

We define $\mathrm{OE}(U,|\psi\rangle,p)$ to be the value returned by the OEA. A flowchart for the algorithm is depicted in Fig. 3.

When $a=|\langle\psi|U|\psi\rangle|$ is close to 1, the absolute precision with which a is obtained is as much as quadratically better for the same resources. To avoid this nonuniformity of the precision to resource relationship, we define the precision δ of an overlap by means of a parametrization of $\langle\psi|U|\psi\rangle$ using the points (x_1,x_2,x_3) on the upper hemisphere of the surface

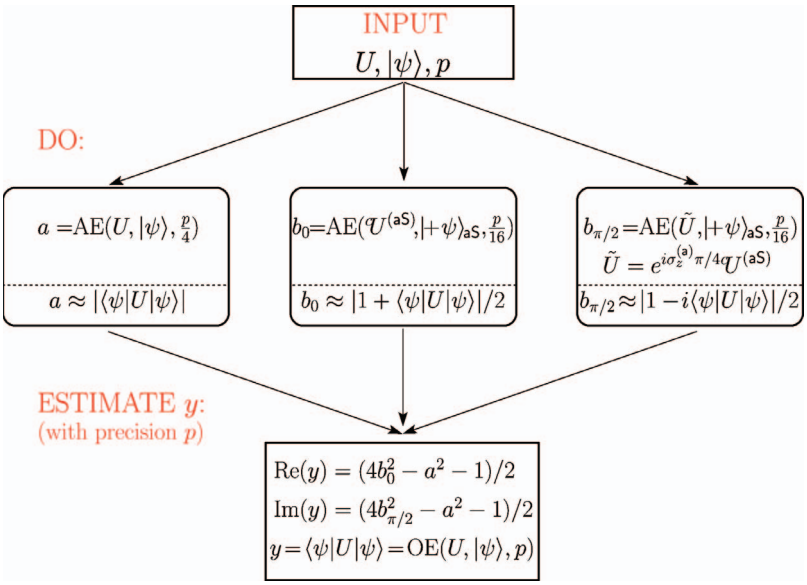

FIG. 3. (Color online) OEA flowchart. An estimate of the overlap $\langle\psi|U|\psi\rangle$ is obtained. The algorithm requires three state preparations and calls the AEA three times. The amplitude of the returned value shown in the flowchart may need to be adjusted according to the value of a to optimize the precision. For details see the text.

of a unit sphere in three dimensions. For this purpose, define $h(x_1,x_2,x_3)=x_1+ix_2$ for $x_1^2+x_2^2+x_3^2=1$ and $x_3 \geqslant 0$. Define the distance between (x_1,x_2,x_3) and (x_1',x_2',x_3') to be the angular distance along a great circle. The precision of the value o returned by the OEA is determined by the distance δ between the liftings $h^{-1}(o)$ and $h^{-1}(\langle\psi|U|\psi\rangle)$ (see Fig. 4). We define the precision of the value returned by the AEA similarly, by restricting the parametrization to the positive reals. The precision parameters with which the AEA is called in the OEA are chosen so that the returned overlap has precision $\delta \leqslant p$ with respect to our parametrization (see Endnote [15]).

The AEA is based on a trick for converting amplitude into phase information, so that an efficient PEA can be applied. Let $|\psi_0\rangle=|\psi\rangle$ and $|\psi_1\rangle=U|\psi\rangle$. Let $S_0=\mathbf{I}-2|\psi_0\rangle\langle\psi_0|=VP_0V^\dagger$ be the selective sign change of $|\psi_0\rangle$ and $S_1=\mathbf{I}-2|\psi_1\rangle\langle\psi_1| = UVP_0V^\dagger U^\dagger$ the selective sign change of $|\psi_1\rangle$. The composition $S=S_0S_1$ is a unitary operator that rotates $|\psi_0\rangle$ toward $|\psi_1\rangle$ in the two-dimensional subspace $\mathcal{Q}$ spanned by $|\psi_0\rangle$ and $|\psi_1\rangle$. The rotation is by a Bloch-sphere angle of $2\phi = 4\arccos(|\langle\psi_0|\psi_1\rangle|)$. Thus, the eigenvalues of S in $\mathcal{Q}$ are $e^{\pm i\phi}$. The Bloch-sphere picture of the states and the rotation are shown in Fig. 5. When $|\langle\psi_0|\psi_1\rangle|=|\langle\psi|U|\psi\rangle|=1$, S is the identity operator. The PEA for S with initial state $|\psi_0\rangle$ determines the phase ϕ of one of these eigenvalues, where each of the signs has equal probability of being returned. The overlap $|\langle\psi|U|\psi\rangle|$ is obtained from ϕ by the formula $|\langle\psi|U|\psi\rangle| = \cos(\phi/2)$. The PEA requires use of the conditional S operator ${}^cS$. As defined, this needs to be decomposed into a product of ${}^cP_0$, ${}^cU$, and ${}^cV$. A significant simplification is to not condition U and V and to write ${}^cS = V\,{}^cP_0V^\dagger UV\,{}^cP_0V^\dagger U^\dagger$. This works because if the controlling qubit is in state $|0\rangle$, all the U's and V's are canceled by matching $U^\dagger$'s and $V^\dagger$'s [7].

Let $\mathrm{PE}(W,|\psi'\rangle,p)$ be a phase returned by the PEA for unitary operator W and initial state $|\psi'\rangle$ with precision goal p. The AEA may be summarized as follows.

Amplitude estimation algorithm. Given are U, $|\psi\rangle$ (in terms of a preparation unitary $V:|0\rangle\mapsto|\psi\rangle$), and the goal pre-

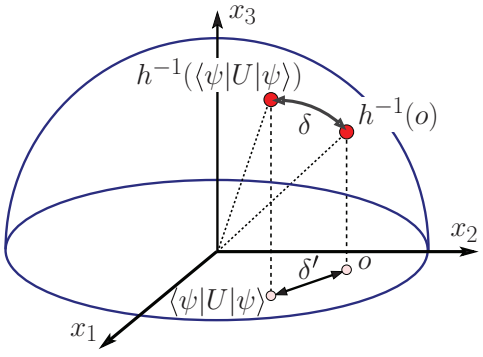

FIG. 4. (Color online) Visualization of the parameterization of the overlap in terms of points on the upper hemisphere of a unit sphere. The function h is defined by $h(x_1,x_2,x_3)=x_1+ix_2$. Note that for overlaps $|\langle\psi|U|\psi\rangle|$ approaching 1 and small δ, δ' approaches $\delta^2/2 \ll \delta$.

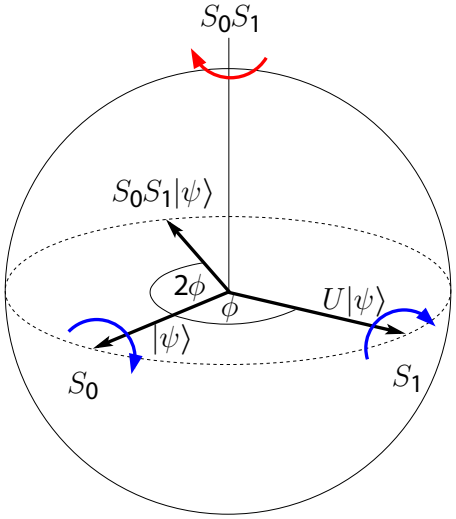

FIG. 5. (Color online) Bloch-sphere representation of the rotations induced on the subspace spanned by $|\psi\rangle$ and $U|\psi\rangle$ by the operators S_0 and S_1.

cision p. An estimate of $|\langle\psi|U|\psi\rangle|$ is to be returned.

(1) Let $\phi=\mathrm{PE}(S,|\psi\rangle,2p)$ with $S=S_0S_1=VP_0V^\dagger UVP_0V^\dagger U^\dagger$.

(2) Estimate $|\langle\psi|U|\psi\rangle|$ as $|\cos(\phi/2)|$.

The precision parameter for the PEA has the conventional interpretation (modulo 2π). Because $\arccos(|\langle\psi|U|\psi\rangle|)$ is the angle along the semicircle in the parametrization of the overlap defined above, the precision $2p$ of the value returned by the PEA translates directly to the desired precision in the value to be returned by the AEA.

The PEA [9] for a unitary operator W and initial state $|\psi'\rangle$ returns an estimate of the phase ϕ ("eigenphase") of an eigenvalue $e^{i\phi}$ of W, where the probability of ϕ is given by the probability amplitude of $|\psi'\rangle$ in the $e^{i\phi}$-eigenspace of W. In the limit of perfect precision, it acts as a von Neumann measurement of W on state $|\psi'\rangle$ in the sense that the final state is projected onto the $e^{i\phi}$-eigenspace of W. For finite precision, the eigenspaces may be decohered and the projection is incomplete, unless there are no other eigenvalues within the precision bound. The error in the projection is related to the confidence level with which the precision bound holds.

The original PEA is based on the binary-quantum Fourier transform [14]. It determines an eigenphase ϕ with precision $\frac{1}{2^n}$ with 2^n-1 uses of the conditional ${}^cW$ operator to obtain a phase kickback to ancilla qubits. The original PEA begins by preparing n qubits labeled $1,\ldots,\mathsf{n}$ in state $|+\rangle_1,\ldots,|+\rangle_\mathsf{n}$ and system S in state $|\psi'\rangle_\mathsf{S}$. Next, for each $m=1,\ldots,n$, ${}^cW$ is applied from qubit m to system S 2^{m-1} times. The binary-quantum Fourier transform is applied to the n qubits, and the qubits are measured in the logical basis $|0\rangle$, $|1\rangle$. The measurement outcomes give the first n digits of the binary representation of $\phi/(2\pi)+\epsilon/2^n$, where $|\epsilon|<1/2$ with probability at least 0.405 [9].

The PEA as outlined in the previous paragraph makes suboptimal use of quantum resources. We prefer a one-qubit version of the algorithm based on the measured quantum Fourier transform [16] and described in [17], which has been experimentally implemented on an ion trap quantum computer [18]. An advantage of this approach is that it does not require understanding the quantum Fourier transform and is readily related to more conventional approaches for measuring phases. To understand how the algorithm given below works, note that the eigenstates of W are invariant under ${}^cW$. The only interaction with S is via uses of ${}^cW$. Therefore, without loss of generality, we can assume that S is initially projected onto an $e^{i\phi}$ eigenstate of W with $0\leqslant\phi<2\pi$. The bits of an approximation of $\phi/(2\pi)$ are determined one by one, starting with the least significant one that we wish to learn. Given n, let $[.b_1\ldots b_n]_2=\Sigma_{i=1}^n b_i/2^i$ (with $b_i=0,1$) be a best n-digit binary approximation to $\phi/(2\pi)$, where the notation $[x]_2$ is used to convert a sequence of binary digits x to the number that it represents. Write $\epsilon=(\phi/(2\pi)-[.b_1\ldots b_n]_2)2^n$.

Phase estimation algorithm. Given are W, $|\psi'\rangle$ (as a state of a quantum system), and the goal precision p. An estimate of an eigenphase ϕ of W is to be returned, where the probability of ϕ is given by the population of $|\psi'\rangle$ in the corresponding eigenspace.

(1) Let n be the smallest natural number such that $2^n\geqslant 1/p$.

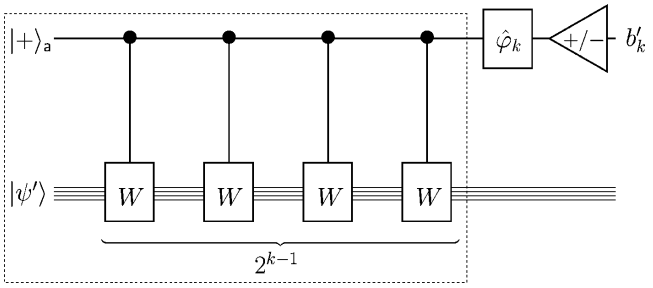

FIG. 6. Step (3) of the PEA to estimate bit k of the eigenphase, where $k=3$. The phase $\hat{\varphi}_k$ is computed according to previously obtained information about the eigenphase. By applying it before the measurement, the probability of obtaining the optimal value for bit k is maximized. The measurement is denoted by the triangle pointing left with +/− inside and is a measurement in the $|+\rangle/|-\rangle$ basis. The outlined part of the network is parallelized in Sec. V.

(2a) Prepare $|+\rangle_\mathsf{a}$ in an ancilla qubit a and apply ${}^cW^{(\mathsf{aS})}$ 2^{n-1} times. With the auxiliary assumption that $|\psi'\rangle$ is an $e^{i\phi}$-eigenstate of W, the effect is a phase kickback, changing $|+\rangle_\mathsf{a}$ to $(|0\rangle_\mathsf{a}+e^{i2^{n-1}\phi}|1\rangle_\mathsf{a})/\sqrt{2}$.

(2b) Measure a in the $|+\rangle$, $|-\rangle$ basis, so that measurement outcome 0 [1] is associated with detecting $|+\rangle$ $[|-\rangle]$. Let b'_n be the measurement outcome. With the auxiliary assumption, the probability that $b'_n=b_n$ is $\cos(\pi\epsilon/2)^2$.

(3) Do the following for each $k=(n-1),\ldots,1$:

(3a) Prepare $|+\rangle_\mathsf{a}$ in an ancilla qubit a and apply ${}^cW^{(\mathsf{aS})}$ 2^{k-1} times. With the auxiliary assumption, this changes $|+\rangle_\mathsf{a}$ to $(|0\rangle_\mathsf{a}+e^{i2^{k-1}\phi}|1\rangle_\mathsf{a})/\sqrt{2}$.

(3b) Compensate the phase of $|1\rangle_\mathsf{a}$ by changing it by $e^{-i\pi[.b'_{k+1}\ldots b'_n]_2}$. With the auxiliary assumption, this changes the state of the ancilla to $(|0\rangle_\mathsf{a}+e^{i(2^{k-1}\phi-\pi[.b'_{k+1}\ldots b'_n]_2)}|1\rangle_\mathsf{a})/\sqrt{2}$.

(3c) Measure a in the $|+\rangle$, $|-\rangle$ basis to obtain b'_k. With the auxiliary assumption and if $b'_l=b_l$ for $l>k$, the probability that $b'_k=b_k$ is $\cos(\pi\epsilon/2^{n-k+1})^2$.

(4) Estimate ϕ as $2\pi[.b'_1\ldots b'_n]_2$.

A step of the algorithm is depicted in Fig. 6.

The probability $P(\epsilon)$ that the value returned by the PEA is $2\pi[.b_1\ldots b_n]_2$ is the product of the probabilities $\cos(\pi\epsilon/2^l)^2$ for $l=1,\ldots,n$ and is bounded below by $\sin(\pi\epsilon)^2/(\pi\epsilon)^2$. This bound can be obtained by taking the limit $n\to\infty$ in $P(\epsilon)$. The worst case is given for $|\epsilon|=1/2$, leading to the bound $P(\epsilon)\geqslant 4/\pi^2\approx 0.405$ [9]. Since the goal precision is 2^{-n}, it is acceptable for the algorithm to obtain the next best binary approximation to ϕ. For this, the value obtained for b'_n may not be the one with maximum probability, but the subsequent bits b'_k are always the best possible given b'_n. Taking this into account, the probability that the phase returned is within 2^{-n} is given by $P(\epsilon)+P(1-\epsilon)\geqslant 8/\pi^2\approx 0.81$ (see Endnote [19]).

The key step of the one-qubit phase estimation procedure is to modify the phase kickback by the previously obtained phase estimate. This differentiates it from an adaptive phase measurement method that determines the bits of an approximation of $\phi/(2\pi)$ starting with the most significant bit and making sufficiently many measurements with different phase compensations for each bit to achieve high confidence level. This is the phase estimation method given in [20], mentioned

in [2] and used in [21]. To ensure that all the digits are correct with sufficiently high probability, the number of measurements needed for each bit is logarithmic in the number of digits. It approximates what is done in practice for the efficient determination of an unknown frequency or pulse time.

The PEA can be applied directly to the problem of estimating the parameter ϕ in a unitary operator $e^{-i\phi H}$ if an eigenstate of H with known eigenvalue can be prepared. In this application of the PEA it is often possible to avoid the use of conditional evolutions. For example, suppose that we can treat the subspace spanned by two eigenstates of H with known and different eigenvalues as the state space of a qubit. In this case, the phase kickback required for the PEA can be implemented by applying $e^{-i\phi H}$ directly on this qubit. This removes the need for an ancilla qubit and makes the algorithm useful for improving the efficiency of protocols such as that used for clock synchronization in [22] from $O(|\log(p)|/p)$ to $O(1/p)$.

The resources required by the PEA, AEA, and OEA can be summarized as follows.

PE$(W,|\psi'\rangle,p)$: This requires $N(p)=2^{\lceil \log_2(1/p)\rceil}-1$ uses of W. $|\psi'\rangle$ is prepared once. Here, $\lceil x\rceil$ denotes the least integer $m\geqslant x$.

AE$(U,|\psi\rangle,p)$: This calls PE once. It requires $N(2p)$ uses of $S=VP_0V^\dagger UVP_0V^\dagger U^\dagger$ and one use of V to prepare the initial state. We count this as being equivalent to $4N(2p)+1$ state preparations and $2N(2p)$ applications of U.

OE$(U,|\psi\rangle,p)$: This contains three calls to the AEA with higher precision. The total resource count is $8N(p/8)+4N(p/2)+3$ state preparations and $4N(p/8)+2N(p/2)$ uses of U.

Since $N(p)$ is of order $1/p$, each of these algorithms uses resources of order $1/p$.

III. CONFIDENCE BOUNDS

The PEA as described in the previous section obtains an estimate $\phi_{\rm est}$ of an eigenphase ϕ such that the prior probability that $|\phi_{\rm est}-\phi|<2^{-n+1}\pi$ is at least 0.81, regardless of the value of ϕ, where $n=\lceil \log_2(1/p)\rceil$. (The comparison of $\phi_{\rm est}$ to ϕ is modulo 2π, so that $|\phi_{\rm est}-\phi|$ is the angular distance between $e^{i\phi_{\rm est}}$ and $e^{i\phi}$.) Thus, after having obtained $\phi_{\rm est}$, we say that $\phi=\phi_{\rm est}\pm 2^{-n+1}\pi$ with confidence level 0.81 or $P[\phi_{\rm est}-2^{-n+1}\pi<\phi<\phi_{\rm est}+2^{-n+1}\pi]\geqslant 0.81$. The error bound of $2^{-n+1}\pi$ must not be confused with a standard deviation. Suppose that we use a single sample from a Gaussian distribution with standard deviation σ to infer the mean. We would expect that the confidence level increases as $1-e^{-\Omega((\Delta/\sigma)^2)}$ for an error bound of Δ. [The notation $\Omega(x)$ means a quantity asymptotically bounded below by something proportional to x; that is, there exists a constant $C>0$ such that the quantity is eventually bounded below by Cx.] In general, it is desirable to have confidence levels that increase at least exponentially as a function of distance Δ or as a function of additional resources used. Unfortunately, for a single instance of the PEA, we cannot do better than have confidence level $1-O(1/\Delta)$ for $\phi=\phi_{\rm est}\pm 2^{-n+1}\pi\Delta$ [9]. [Here, $O(x)$ denotes a quantity that is of order x; that is, a quantity that is eventually bounded above by Cx for some constant C. The meaning of "eventually" depends on context. Here it means "for sufficiently small x." If the asymptotics of the argument require that it go to infinity, it means "for sufficiently large x."] The method suggested in [9] for increasing the confidence level is to use the PEA with a higher goal precision of $p/2^l$. However, this improves the confidence level on $\phi=\phi_{\rm est}\pm 2^{-n+1}\pi\Delta$ to only $1-\Omega(1/(\Delta 2^l))$ and requires a 2^l resource overhead, which is not an efficient improvement.

A reasonable goal is to attain confidence level $c=1-e^{-\Omega(r)}$ for $\phi=\phi_{\rm est}\pm 2^{-n+1}\pi$ with a resource overhead of a factor of $O(r)$. This modifies the resource counts from the previous section from $O(1/p)$ to $O(|\log(1-c)|/p)$, where c is the confidence level achieved. To attain this goal, we modify each step of the PEA by including repetition to improve the confidence level that acceptable values for the bits are determined. Let the two nearest n-digit binary approximations to $\phi/(2\pi)$ be given by $\phi/(2\pi)=[.b_1\ldots b_n]_2+\delta/2^n$ and $\phi/(2\pi)=[.\tilde{b}_1\ldots\tilde{b}_n]_2+(\delta-1)/2^n$, where $0\leqslant\delta<1$. We wish to obtain one of these approximations with a high confidence level. For the first step of the PEA, we perform two sets of r experiments to obtain a good estimate of $\delta'=\pi(\delta+b_n)$. The first set consists of r $(|+\rangle_{\rm a},|-\rangle_{\rm a})$-measurements of the state ${}^cW^{2^{n-1}}|+\rangle_{\rm a}|\psi\rangle_{\rm S}$. The second consists of r $(|+\rangle_{\rm a},|-\rangle_{\rm a})$-measurements of the state ${}^cW^{2^{n-1}}(|0\rangle_{\rm a}-i|1\rangle_{\rm a})/\sqrt{2}|\psi\rangle_{\rm S}$. Let x_1,x_2 be the sample means of the measurement outcomes of the two sets of experiments. In the limit of large r, x_1 and x_2 approach $\sin(\delta'/2)^2$ and $\sin(\delta'/2-\pi/4)^2$, respectively. We have

$$\sin(\delta')=\cos(\delta'-\pi/2)=1-2\sin(\delta'/2-\pi/4)^2,$$

$$\cos(\delta')=1-2\sin(\delta'/2)^2, \tag{4}$$

so we can estimate δ' from x_1 and x_2 by letting $\delta'_{\rm est}$ be the phase of the complex vector $(1-2x_1)+i(1-2x_2)$. The probability of the event E that δ' differs from $\delta'_{\rm est}$ by more than $\pi/4$ modulo 2π can be bounded as follows. For this event, $|\sin(\delta')+i\cos(\delta')-[(1-2x_1)+i(1-2x_2)]|^2\geqslant 1/2$. It follows that either $|\sin(\delta'/2)^2-x_1|\geqslant 1/4$ or $|\sin(\delta'/2-\pi/4)^2-x_2|\geqslant 1/4$. The probability of each of these possibilities is bounded by the probability that the mean of r samples of the binomial distribution with probability p of outcome 1 differs from p by at least $x=1/4$. The probability of this event is bounded by $2e^{-2rx^2}=2e^{-r/8}$ (Hoeffding's bound [23]). This bound can now be doubled to obtain a bound of $4e^{-r/8}$ on the probability of E.

Let $a_n=1$ if $\delta'_{\rm est}$ is closer to π than to 0 and $a_n=0$ otherwise. Then $a_n=b_n$ or $a_n=\tilde{b}_n$. Which equality holds does not affect the subsequent arguments, so without loss of generality, assume that $a_n=b_n$. Suppose that event E did not happen and that we have correctly obtained $a_n=b_n,\ldots,a_{k+1}=b_{k+1}$. For the step of the algorithm that determines the kth bit, modify the original step by compensating the phase of $|1\rangle_{\rm a}$ by $e^{-i(\pi[.b_{k+1}\ldots b_{n-1}]_2+\delta'_{\rm est}/2^{n-k})}$ and repeating the measurement r times. We set $a_k=1$ if the majority of the measurement out-

comes are 1 and $a_k=0$ otherwise. For each measurement, the probability that the measurement outcome does not agree with b_k is at most $\sin((\delta'-\delta'_{\text{est}})/2^{n-k+1})^2$. Our assumptions imply that this is at most $\sin(\pi/2^{n-k+3})^2 \leqslant (\pi/2^{n-k+3})^2$. Using Hoeffding's bound again, the probability that $a_k \neq b_k$ is bounded by $2e^{-2r[1/2-(\pi/2^{n-k+3})^2]} < 2e^{-r/2}$ (for a loose upper bound).

Summing the probabilities, we find that the probability that we do not learn $b_1,\ldots,b_n$ or $\tilde{b}_1,\ldots,\tilde{b}_n$ is bounded by $x(n,r)=2(n-1)e^{-r/2}+4e^{-r/8}$. We can therefore say that the modified PEA yields the desired phase to within $\pi/2^{n-1}$ with confidence level $1-x(n,r)$, where $x(n,r)$ decreases exponentially in r. Note again that this confidence bound should not be confused with a similar confidence bound for a Gaussian random variable. Increasing the confidence bound does not result in the expected increase in confidence level. In order to have confidence level increasing exponentially toward 1 with increasing confidence bound and an additional overhead of at most $O(|\log(p)|)$, we can repeat the determination of the kth bit $2^{n-k}r$ instead of r many times.

For the purpose of having a high confidence level in the precision with which a quantity is estimated, our algorithms require the confidence level goal as an input. The modified PEA may be outlined as follows.

Modified phase estimation algorithm. Given are W, $|\psi'\rangle_{\mathsf{S}}$, a goal precision p, and a goal confidence level c. An eigenphase ϕ of W is to be returned, where the probability of ϕ is given by the population of $|\psi'\rangle$ in the corresponding eigenspace. The final state of S consists of states with eigenphases in the range $\phi\pm p$ with prior probability at least c.

(1) Let n be the smallest natural number such that $2^n \geqslant 1/p$. Let r be the smallest natural number such that $x(n,r)<(1-c)$.

(2) Obtain δ'_{est} with the two sets of r measurements described above. Let $a_n=1$ if δ'_{est} is closer to π than 0 and $a_n=0$ otherwise.

(3) Do the following for each $k=(n-1),\ldots,1$, in this order:

(3a) Obtain an estimate of the kth bit a_k of a binary approximation to $\phi/(2\pi)$ by r repetitions of the measurement of steps (3a)–(3c) given previously, but with a phase compensation that uses δ'_{est} as well as the previously obtained bits.

(4) Return $2\pi[.a_1\ldots a_n]_2$.

We define $\text{PE}(W,|\psi'\rangle,p,c)$ to be the value returned by the modified PEA.

The resources required grow by a factor of less than $2r$, where $r=O(|\log(1-c)|)$. The constant hidden by the order notation may be determined from the expression for r in step (1) and is not very large. To modify the AEA to attain confidence level c, it suffices to change the call to PE by including c as an argument. Because the OEA has three independent calls to the AEA, it needs to make these calls with confidence level arguments of $1-(1-c)/3$ to ensure that the final confidence level is c. The resource requirements of all three algorithms are $O(|\log(1-c)|/p)$, where this applies to the uses of U as well as the state preparation operator V in the case of the AEA and OEA.

IV. EXPECTATION ESTIMATION

Let A be an observable and assume that it is possible to evolve under $\pm A$ for any amount of time. This means that we can implement the unitary operator e^{-iAt} for any t. The traditional idealized procedure for measuring $\langle A\rangle=\text{tr}(A\rho)$ is to adjoin a system consisting of a quantum particle in one dimension with momentum observable $\hat{p}$ and apply the coupled evolution $e^{-iA\otimes\hat{p}}$ to the initial state $\rho\otimes|0\rangle\langle 0|$, where $|0\rangle$ is the position "eigenstate" with eigenvalue 0. Measuring the position of the particle yields a sample from the distribution of eigenvalues of A [5,24]. This procedure requires unbounded energy, both for preparing $|0\rangle$ and to implement the coupled evolution. Performing this measurement N times yields an estimate of $\langle A\rangle$ with precision of order $\text{var}(A)/\sqrt{N}$, where the variance is $\text{var}(A)=\langle(A-\langle A\rangle)^2\rangle$. It is desirable to improve the precision and to properly account for the resources required to implement the coupling.

We focus on measurement methods that can be implemented in a quantum information processor. In order to accomplish this, some prior knowledge of the distribution of eigenvalues of A with respect to ρ is required. Suppose we have an upper bound b on $|\text{tr}(A\rho)|$ and a bound on the tail distribution $F(\Delta)\geqslant\text{tr}([|A-\langle A\rangle|>\Delta]\rho)$, where $[|A-\langle A\rangle|>\Delta]$ denotes the projection operator onto eigenspaces of A with eigenvalues λ satisfying $|\lambda-\langle A\rangle|>\Delta$. That is, $F(\Delta)\geqslant\Sigma_{|\lambda-\langle A\rangle|>\Delta}\,p_\lambda$ with $p_\lambda=\text{tr}(|\lambda\rangle\langle\lambda|\rho)$. Without loss of generality, F is nonincreasing in Δ. An estimate of the tail distribution is needed to guarantee the confidence bounds on $\text{tr}(A\rho)$ derived from measurements by finite means. Here are some examples: If the maximum eigenvalue of A is $\lambda_{\max}$, we can set $b=\lambda_{\max}$ and use $F(\Delta)=1$ if $\Delta<\lambda_{\max}$ and $F(\Delta)=0$ otherwise. Suppose that we have an upper bound v on the variance $\text{var}(A)$. If we know that the distribution of eigenvalues of A is Gaussian, we can estimate $F(\Delta)$ by means of the error function for Gaussian distributions. With no such prior knowledge, the best estimate is $F(\Delta)=\min(1,v/\Delta^2)$. (Observe that $v\geqslant\Delta^2\Sigma_{|\lambda-\langle A\rangle|>\Delta}\,p_\lambda$.) Such "polynomial" tails result in significant overheads for measuring $\langle A\rangle$. "Good" tails should drop off at least exponentially for large Δ ("exponential tails").

We give an EEA based on overlap estimation. The relevant resources for the EEA are the number M of times a unitary operator of the form e^{-iAt} is used, the total time T that we evolve under A, and the number N of preparations of ρ. The total time T is the sum of the absolute values of exponents t in uses of e^{-iAt}. For applying the OEA, it is necessary to be able to evolve under $-A$ as well as A. If the evolution is implemented by means of quantum networks, this poses no difficulty. However, if the evolution uses physical Hamiltonians, this is a nontrivial requirement. The complexity of realizing e^{-iAt} may depend on t and the precision required. Since this is strongly dependent on A and the methods used for evolving under A, we do not take this into consideration and assume that the error in the implementation of e^{-iAt} is sufficiently small compared to the goal precision. In most cases of interest this is justified by results such as those in [25], which show that for a large class of operators A, e^{-iAt}

can be implemented with resources of order $t^{1+\alpha'}/\epsilon^{\alpha'}$, where ϵ is the error of the implementation and α' is arbitrarily small.

For exponential tails F, our algorithm achieves $M,N=O(1/p^{1+\alpha})$ and $T=O(1/p)$ for arbitrarily small α. The order notation hides constants and an initialization cost that depends on b and F. The strategy of the algorithm is to measure $\mathrm{tr}(e^{-iAt}\rho)$ for various t. In the limit of small t, $\mathrm{tr}(e^{-iAt}\rho)=1+O(t^2)-i[\langle A\rangle t+O(t^3)]$, so that $\langle A\rangle$ can be determined to $O(t^3)$ from the imaginary part of $\mathrm{tr}(e^{-iAt}\rho)$. The first problem is to make an initial determination of $\langle A\rangle$ to within a deviation of A as determined by F. This is an issue when b is large compared to the deviation. To solve the first problem, we can use phase estimation. We also give a more efficient method based on amplitude estimation. The second problem is to avoid excessive resources to achieve the desired precision while making t small. To solve this problem requires choosing t carefully and taking advantage of higher-order approximations of $\langle A\rangle$ by linear combinations of $\mathrm{tr}(e^{-iAt}\rho)$ for different times t.

To bound the systematic error in the approximation of $\langle A\rangle$ by $i\mathrm{tr}(e^{-iAt}\rho)$, note that $|\mathrm{Im}(e^{i\theta})-\theta|\leq\theta^3/6$. To see this it is sufficient to bound the Lagrange remainder of the Taylor series of $\sin(\theta)$. This bound suffices for achieving $\alpha=1/2$ in the bounds on M and N. Reducing α requires a better approximation, which we can derive from the Taylor series of the principal branch of $\ln(x+1)$. For $|x|<1$,

$$\left|\ln(x+1)-\sum_{k=1}^{K}(-1)^{k-1}x^k/k\right|\leq|x|^{K+1}/[(K+1)(1-|x|)^{K+1}]. \tag{5}$$

To apply these series to the problem of approximating $\langle A\rangle$, we compute

$$\sum_{k=1}^{K}(-1)^{k-1}(e^{-iBt}-1)^k/k=\sum_{l=0}^{K}C_l e^{-iBlt} \tag{6}$$

for real constants C_l satisfying $|C_l|\leq 2^K$. In particular, if B is an operator satisfying $|B|<x/t$, we can estimate

$$\left|t\,\mathrm{tr}(B\rho)+\sum_{l=0}^{K}C_l\,\mathrm{Im}\,\mathrm{tr}(e^{-iBlt}\rho)\right| \leq|x|^{K+1}/[(K+1)(1-|x|)^{K+1}). \tag{7}$$

Define $G_e(\Delta)=\Delta F(\Delta)+\int_\Delta^\infty F(s)ds$. Then $G_e(\Delta)$ is an upper bound on the contribution to the mean from eigenvalues of A that differ from the mean by more than Δ. That is, $G_e(\Delta)\geq\mathrm{tr}(|A-\langle A\rangle|[|A-\langle A\rangle|>\Delta]\rho)=\Sigma_{|\lambda-\langle A\rangle|>\Delta}|\lambda-\langle A\rangle|p_\lambda$. Like $F(\Delta)$, $G_e(\Delta)$ is nonincreasing. We assume that a nonincreasing bound $G(\Delta)\geq G_e(\Delta)$ is known and that $G(\Delta)\to 0$ as $\Delta\to\infty$. Because $F(\Delta)\leq G_e(\Delta)/\Delta$, we can use G to bound both G_e and F. For $x>0$, define $G^{-1}(x)=\inf\{\Delta\,|\,G(\Delta)\leq x\}$. The behavior of G^{-1} as x goes to 0 determines the resource requirements for the EEA. If A is a bounded operator with bound $\lambda_{\max}$, then we can use $G^{-1}(x)\leq\lambda_{\max}$ independent of $x>0$. If F is exponentially decaying, then so is G and $G^{-1}(x)=O(|\log(x)|)$. For polynomial tails with $F(\Delta)=O(1/\Delta^{2+\beta})$, we have $G(\Delta)=O(1/\Delta^{1+\beta})$ and $G^{-1}(x)=O(1/x^{1/(1+\beta)})$.

The EEA has two stages. The first is an initialization procedure to determine $\langle A\rangle$ with an initial precision that is of the order of a bound on the deviation of A from its mean, where the deviation is determined from F and G. This initialization procedure involves phase estimation to sample from the eigenvalue distribution of A. Its purpose is to remove offsets in the case where the expectation of A may be very large compared to the width of the distribution of eigenvalues as bounded by F and G. The second stage zooms in on $\mathrm{tr}(A\rho)$ by use of the overlap estimation procedure. As before, we can assume without loss of generality that ρ is pure, $\rho=|\psi\rangle\langle\psi|$. We first give a version of the EEA that achieves $M,N=O(1/p^{3/2})$ and then refine the algorithm to achieve better asymptotic efficiency.

Expectation estimation algorithm. Given are A, $|\psi\rangle$ (in terms of a preparation unitary $V:|0\rangle\mapsto|\psi\rangle$), a goal precision p, and the desired confidence level c. The returned value is within p of $\langle A\rangle=\mathrm{tr}(A|\psi\rangle\langle\psi|)$ with probability at least c.

Stage I.

(1) Choose Δ such that $F(\Delta/2)<1/4$ and $\Delta\geqslant p$. Δ should be chosen as small as possible. Let $t_i=\pi/[4(b+\Delta)]$. Let r be the minimum natural number such that $2e^{-r/8}\leq(1-c)/4$ and set c' according to the identity $r(1-c')=(1-c)/4$.

(2) Obtain $\Lambda_1,\ldots,\Lambda_r$ from r instances of the PEA, $\Lambda_k=\mathrm{PE}(e^{-iAt_i},|\psi\rangle,\Delta t_i/2,c')$, where 2π is subtracted for any return values between π and 2π to ensure that $-\pi\leq\Lambda_k<\pi$.

(3) Let Λ_m be the median of $\Lambda_1,\ldots,\Lambda_r$. We show below that the probability that $|\Lambda_m/t_i+\langle A\rangle|>\Delta$ is bounded by $2e^{-r/8}+r(1-c')\leq(1-c)/2$.

(4) Let $a_0=-\Lambda_m/t_i$. We expect a_0 to be within Δ of $\langle A\rangle$ with confidence level $1-(1-c)/2$.

Stage II. If $p=\Delta$, return a_0 and skip this stage.

(1) Choose $\theta_{\max}$ and t so that they satisfy

$$\begin{aligned}&\text{(A)}\quad \theta_{\max}^3/6\leq(t/2)p/4,\\ &\text{(B)}\quad G(\theta_{\max}/t)\leq\theta_{\max}p/8,\\ &\text{(C)}\quad \theta_{\max}\leq 1,\\ &\text{(D)}\quad t\Delta\leq\theta_{\max}.\end{aligned} \tag{8}$$

The constraints and how they can be satisfied are explained below. The parameter t should be chosen as large as possible to minimize resource requirements.

(2) Obtain $x=\mathrm{OE}(e^{-i(A-a_0)(t/2)},|\psi\rangle,(t/2)p/4,1-(1-c)/2)$.

(3) Return $-\mathrm{Im}(x)/(t/2)+a_0$.

Consider stage I of the algorithm. The probability that $|\Lambda_m/t_i+\langle A\rangle|>\Delta$ may be bounded as follows. The choice of t_i ensures that eigenvalues Λ of $-At_i$ within Δt_i of the mean are between $\pm\pi/4$ and do not get "aliased" by e^{-iAt_i} in the calls to the PEA. With probability at least $1-r(1-c')$, each Λ_k returned by these calls is within $t_i\Delta/2$ of an eigenvalue of $-At_i$ sampled according to the probability distribution induced by $|\psi\rangle$. Assume that the event described in the previ-

ous sentence occurred. The probability that $|\Lambda_m/t_i+\langle A\rangle|>\Delta$ is upper bounded by the probability that at least $\lceil r/2\rceil$ of the r samples fall outside the range $[-\langle A\rangle t_i-\Delta t_i,-\langle A\rangle t_i+\Delta t_i]$. The choice of Δ with respect to F implies that Hoeffding's bound can be applied to bound this probability by $2e^{-r/8}$. Thus, we can bound the overall prior probability P that $|\Lambda_m/t_i+\langle A\rangle|>\Delta$ by $P<2e^{-r/8}+r(1-c')\leq(1-c)/2$.

The resources required for stage I include $N=r=O(|\log(1-c)|)$ preparations of $|\psi\rangle$, $M=O(|\log(1-c)|(b+\Delta)/\Delta)$ uses of e^{-iAs} (specifically, M is within a factor of 2 of $2r/\Delta t_i$) and a total evolution time of $T=O(|\log(1-c)|/\Delta)$ (where T is within a factor of 2 of $2r\Delta$). Note that none of these resource bounds depend on the p and that Δ is a bound on a deviation of A from the mean with respect to $|\psi\rangle$. Also, if Δ is of the same order as b, the formulation of stage I of the algorithm is such that the uses of phase estimation require minimal precision. In fact, in this case, stage I of the algorithm could be skipped with minor adjustments to stage II. We show below that stage I can be modified so that the overhead as a function of b is logarithmic. The modification requires that the number of state preparations, N, is of the same order as M.

In the special case of parameter estimation (see the Introduction), $\Delta=p$. Consequently stage II is skipped and the resources of stage I are the total resources required. The algorithm therefore achieves the optimal $O(1/p)$ resource requirements for this situation.

Consider stage II of the algorithm. The error $|-\mathrm{Im}(x)/(t/2)+a_0-\langle A\rangle|$ may be bounded as follows. We assume that all the precision constraints of stages I and II are satisfied. The confidence level that this is true is c overall. With this assumption, $x/(t/2)$ is within $p/4$ (the "precision error") of $\mathrm{tr}(e^{-i(A-a_0)(t/2)}\rho)/(t/2)$. There are three contributions to the "approximation error," which is the difference between $-\mathrm{Im}\,\mathrm{tr}(e^{-i(A-a_0)(t/2)}\rho)/(t/2)$ and $\mathrm{tr}[(A-a_0)\rho]$. For all contributions, we have to consider the fact that a_0 approximates $\langle A\rangle$ to within only Δ, which is why we need constraint (D) of Eq. (8). The first arises from eigenvalues of $(A-a_0)(t/2)$ in $[-\theta_{\max},+\theta_{\max}]$ due to $|\mathrm{Im}(e^{i\theta})-\theta|$ not being zero and is bounded by $\theta_{\max}^3/[6(t/2)]=p/4$ [constraint (A) of Eq. (8)]. The second and third come from eigenvalues of $(A-a_0)(t/2)$ outside $[-\theta_{\max},+\theta_{\max}]$. Constraint (D) of Eq. (8) implies that $|(a_0-\langle A\rangle)(t/2)|\leq\theta_{\max}/2$. Constraints (B) and (C) of Eq. (8) imply that the contribution to $\langle A\rangle$ of eigenvalues differing from the mean by more than $\theta_{\max}/[2(t/2)]$ is at most $\theta_{\max}p/8\leq p/8$. However, the same eigenvalues still contribute to the measurement, each contributing at most 1 to x. Constraint (B) of Eq. (8) together with the inequality $F(\Delta)\leq G(\Delta)/\Delta$ imply that $F(\theta_{\max}/[2(t/2)])\leq tp/8$, so this contribution has probability at most $tp/8$ and therefore adds at most another $p/4$ (after dividing by $t/2$) to the approximation error. Thus, the combination of the approximation and precision error is less than p, as desired. Clearly these estimates are suboptimal; tighter choices of $\theta_{\max}$ and t could be made. However, this does not affect the asymptotics of the resource requirements.

To find good solutions $\theta_{\max}$ and t subject to the constraints given in Eq. (8), we can rewrite the constraints as follows:

$$(\mathrm{A}')\quad G^{-1}(\theta_{\max}p/8)\leq\theta_{\max}/t\leq(p/8)/(\theta_{\max}^2/6),$$

$$(\mathrm{B}')\quad \theta_{\max}\leq 1,\quad \theta_{\max}/t\geq\Delta. \tag{9}$$

The first inequality of (A′) is implied by constraint (B) and the second by constraint (A) of Eq. (8). To satisfy these constraints, we first find $\theta_{\max}\leq 1$ as large as possible so that

$$(\mathrm{A}'')\quad \Delta\leq G^{-1}(\theta_{\max}p/8)\leq(p/8)/(\theta_{\max}^2/6), \tag{10}$$

and then set $t=\theta_{\max}/G^{-1}(\theta_{\max}p/8)$. Consider the three examples of bounded, exponential, and polynomial tails. For the case of bounded tails, constraint (A″) of Eq. (10) can be solved by setting $\theta_{\max}$ according to $\lambda_{\max}=(p/8)/(\theta_{\max}^2/6)$, so that $\theta_{\max}=[3p/(4\lambda_{\max})]^{1/2}$. The parameter t is given by $\theta_{\max}/\lambda_{\max}=(3p/4)^{1/2}/\lambda_{\max}^{3/2}=\Omega(p^{1/2})$. For the case of exponential tails, we can use $G^{-1}(x)=O(|\log(x)|)$ to show that $\theta_{\max}=\Omega([p/|\log(p)|]^{1/2})$ and $t=\Omega(p^{1/2}/|\log(p)|^{3/2})$ (see Endnote [26]). For polynomial tails with $G^{-1}(x)=O(x^{-1/(1+\beta)})$, we get $\theta_{\max}=\Omega(p^{(2+\beta)/(1+2\beta)})$ and $t=\Omega(p^{(5+6\beta+\beta^2)/[(1+\beta)(1+2\beta)]})$ (see Endnote [27]).

The resource requirements for stage II of the EEA can be estimated as $M=O(|\log(1-c)|/tp)$ uses of an exponential of the form e^{-iAs}, $N=O(|\log(1-c)|/tp)$ state preparations, and a total time of $T=O(|\log(1-c)|/p)$, in terms of the parameter t computed in step (1) (of stage II). The dependence on G shows up in the value of t. With t as computed in the previous paragraph, for bounded A, M and N are $O(|\log(1-c)|/p^{3/2})$. For exponential tails, M and N are $O(|\log(1-c)|/[p/|\log(p)|]^{3/2})$. For polynomial tails, they are $O(|\log(1-c)|/p^{\gamma(\beta)})$, where $\gamma(\beta)$ is a polynomial satisfying $\gamma(\beta)\to 1+1/2$ for $\beta\to\infty$ and $\gamma(\beta)=1+5$ for $\beta=0$.

To reduce the resource requirements of stage II of the EEA, we use overlap estimation at multiple values of t and Eq. (6). Here is the modified stage. We assume that $K\geq 2$.

Stage II′.

(1) Choose $\theta_{\max}$ and t so that they satisfy

$$(\mathrm{A})\quad \theta_{\max}^{K+1}/[(K+1)(1-\theta_{\max})^{K+1}]\leq(t/2)p/4,$$

$$(\mathrm{B})\quad G(\theta_{\max}/t)\leq\theta_{\max}p/(8K2^K),$$

$$(\mathrm{C})\quad \theta_{\max}\leq 1,$$

$$(\mathrm{D})\quad t\Delta\leq\theta_{\max}. \tag{11}$$

The parameter t should be chosen as large as possible to minimize resource requirements.

(2) For $l=1,\ldots,K$, obtain $y_l=\mathrm{OE}(e^{-i(A-a_0)(lt/2)},|\psi\rangle,(t/2)p/(4K2^K),1-(1-c)/(2K))$. Let $y_0=1$.

(3) Return $-\mathrm{Im}(\Sigma_{l=0}^K C_l y_l)/(t/2)+a_0$.

The precisions and the confidence levels in the calls to the OEA have been adjusted so that the final answer has the correct precision and confidence level. The explanation for this is similar to that for the original stage II (see Endnote [28]).

The earlier method for finding $\theta_{\max}$ and t is readily adapted to the constraints in stage II′. Constraint (A″) of Eq. (10) now reads as

$$(A'') \quad \Delta \leqslant G^{-1}(\theta_{\max} p/(8K2^K)) \leqslant (p/8)(K+1)(1-\theta_{\max})^{K+1}/\theta_{\max}^K, \qquad (12)$$

and we can set $t=\theta_{\max}/G^{-1}(\theta_{\max}p/(8K2^K))$. To simplify the right-hand side of Eq. (12), we add the inequality $\theta_{\max} \leqslant 1/(K+1)$ and use the inequality $1/4 \leqslant (2/3)^3 \leqslant [1-1/(K+1)]^{K+1}$ (for $K \geqslant 2$) to replace the right-hand side by $(p/32)(K+1)/\theta_{\max}^K$. Thus for bounded tails, $\theta_{\max}=\Omega(\min(1/K,(Kp)^{1/K}))$ and $t=\Omega(\min(1/K,(Kp)^{1/K}))$, where we give the asymptotic dependence on K explicitly but suppress parameters not depending on K or p (see Endnote [29]). For exponential tails, $\theta_{\max}=\Omega(\min(1/K,[p/|\log(p)|]^{1/K}))$ and $t=\Omega(\min(1/\{K[|\log(p)|+K]\},p^{1/K}/\{|\log(p)|^{1/K}[|\log(p)|+K]\}))$ (see Endnote [30]). For polynomial tails with exponent β, $\theta_{\max}=\Omega(\min(1/K,p^{(2+\beta)/(K-1+K\beta)}))$ and $t=\Omega(\min(1/\{K[2^{K/(1+\beta)}\times(Kp^{-1})^{1/(1+\beta)}]\},2^{-K/(1+\beta)}p^{(2+\beta)^2/[(1+\beta)(K-1+K\beta)]+1/(1+\beta)}))$ (see Endnote [31]).

With the expressions from the previous paragraph, we can estimate the resources requirements of stage II′. In terms of t, M and N are $O(K^2 2^K|\log(1-c)|/tp)$ and $T=O(K^3 2^K|\log(1-c)|/p)$, where the powers of K account for the K calls to the OEA, the coefficient in the denominator of the precision, and in the case of T, the factor of l in the evolution time. For bounded tails, we obtain $M,N=O(|\log((1-c))|K^3 2^K/p^{1+1/(K)})$, where we have loosely increased the power of K by 1 to account for the upper bound of $O(1/K)$ on t. For exponential tails, $M,N=O(|\log((1-c)/K)|K^4 2^K/[p/|\log(p)|]^{1+1/(K+1)})$ (with appropriate increases in the power of K), and for polynomial tails, $M,N=O(|\log((1-c)/K)|K^4 2^{3K}/p^{\gamma(\beta,K)})$ (with conservative increases in the power of K and the exponent of 2), where $\gamma(\beta,K)$ approaches $1+1/(1+\beta)$ for large K. Note that for $\beta=0$, this approaches the "classical" resource bound as a function of precision.

The final task of this section is to modify stage I so that the dependence of the resource requirements on b is logarithmic rather than linear in b. The basic idea is to use logarithmic search to reduce the uncertainty in $\langle A\rangle$ to Δ. Define q by $b=q\Delta$.

Stage I′.

(1) Choose Δ minimal so that $G(\Delta)<\Delta/6$ and $F(\Delta)<1/18$. Set the initial estimate of $\langle A\rangle$ to $a=0$ and the initial precision to $p_a=b=q\Delta$.

(2) Repeat the following until $p_a \leqslant \Delta$:

(2a) Set $t=1/(p_a+\Delta)$ and obtain $x=\mathrm{OE}(e^{-i(A-a)t},|\psi\rangle,1/18,1-(1-c)/(2\lceil\log_2(q)\rceil))$.

(2b) Update a and p_a according to the assignments $a \leftarrow a-\mathrm{Im}(x)/t$ and $p_a \leftarrow \lceil\Delta/6+(5/18)(p_a+\Delta)\rceil$.

We claim that at the end of this stage, we have determined $\langle A\rangle$ to within Δ with overall confidence level $1-(1-c)/2$, so that we can continue with the second stage, as before. To verify the claim, it is necessary to confirm that at the end of step (2b), the updated estimate a of $\langle A\rangle$ has precision p_a. The error in a can be bounded as we have done for stage II. Let a_0 be the estimate of A used in the call to the OEA. There is an error of less than $1/(18t)=(p_{a_0}+\Delta)/18$ due to the precision of x in the call to the OEA. The remaining error is due to the approximation of $\mathrm{tr}[(A-a_0)t\rho]$ by $-\mathrm{Im}[\mathrm{tr}(e^{-i(A-a)t}\rho)]$. For eigenvalues λ of A within $1/t$ of a, this is bounded by $|\lambda t+\mathrm{Im}(e^{-i\lambda t})| \leqslant 1/6$, which translates into an approximation error of at most $1/(6t)=(p_{a_0}+\Delta)/6$. Eigenvalues of A further from a than $1/t=p_{a_0}+\Delta$ are at least Δ from $\langle A\rangle$. This requires the inductive assumption that $|a_0-\langle A\rangle| \leqslant p_{a_0}$. The contribution to the mean from such eigenvalues is bounded by $\Delta/6$, and the bias resulting from their contribution to x is at most $F(\Delta)/t=(p_{a_0}+\Delta)/18$. Adding up the errors gives the p_a computed in step (2b). The confidence levels in the calls to the OEA are chosen so that the final confidence level is $1-(1-c)/2$. To see this requires verifying that the number of calls of the OEA is at most $\lceil\log_2(q)\rceil$. It suffices to show that if $p_{a_0} \geqslant 2\Delta$, then $\Delta/6+(5/18)(\Delta+p_{a_0}) \leqslant p_{a_0}/2$. Rewrite the left-hand side as $(8/18)\Delta+(5/18)p_{a_0}$, which for $p_{a_0} \geqslant 2\Delta$ is less than $(4/18)p_{a_0}+(5/18)p_{a_0}=p_{a_0}/2$.

Each call to the OEA in stage II′ has constant precision, which implies that M and N are both $O(\log(q))=O(\log(b/\Delta))$ for large q. The total time T is $O(1/\Delta)$.

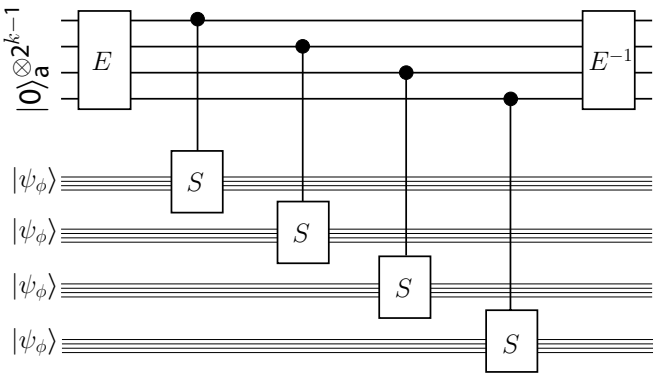

FIG. 7. Parallelization of the PEA algorithm to estimate the bit $k=3$ of the phase. This replaces the outlined parts of the network in Fig. 6. E is an entangler such that $E|0\rangle_{\mathrm{a}}^{\otimes 2^{k-1}}=(|0\ldots0\rangle_{\mathrm{a}}+|1\ldots1\rangle_{\mathrm{a}})/\sqrt{2}$ and $E|100\ldots0\rangle_{\mathrm{a}}=(|0\ldots0\rangle_{\mathrm{a}}-|1\ldots1\rangle_{\mathrm{a}})/\sqrt{2}$. E^{-1} is the decoding operation that maps $E^{-1}|0\ldots0\rangle_{\mathrm{a}}=|+0\ldots0\rangle_{\mathrm{a}}$, and $E^{-1}|1\ldots1\rangle_{\mathrm{a}}=|-0\ldots0\rangle_{\mathrm{a}}$, where $|\pm\rangle=(|0\rangle\pm|1\rangle)/\sqrt{2}$. The kth bit is estimated from the measurement outcome of the first ancilla qubit in the logical basis.

V. PARALLELIZABILITY

To what extent are the algorithms given in the previous sections parallelizable? Consider the OEA. At its core is the PEA with a unitary operator S that has two eigenvalues $e^{\pm i\phi}$ on the relevant state space. In the sequential implementation, one of the eigenvalues is eventually obtained with the desired precision. Which eigenvalue is returned cannot be predicted beforehand. The initial state is such that each one has equal probability. If it is possible to deterministically (or near-deterministically) prepare an eigenstate $|\psi_\phi\rangle$ with (say) eigenvalue $e^{i\phi}$ using sufficiently few resources, then we can use the entanglement trick in [11] to parallelize the algorithm (see Fig. 7). Instead of applying S sequentially 2^{k-1} many times to determine bit k of the phase, we prepare the en-

tangled state $(|0\ldots0\rangle_{\mathsf{a}}+|1\ldots1\rangle_{\mathsf{a}})/\sqrt{2}$ on 2^{k-1} ancilla qubits and 2^{k-1} copies of $|\psi_\phi\rangle$. We next apply ${}^cS$ between the jth ancilla and the jth copy of $|\psi_\phi\rangle$ and then make a measurement of $(|0\ldots0\rangle_{\mathsf{a}}\pm|1\ldots1\rangle_{\mathsf{a}})/\sqrt{2}$. On a quantum computer, the measurement requires decoding the superposition into a qubit, which can be done with $O(2^k)$ gates. The decoding procedure can be parallelized to reduce the time to $O(k)$ (see Endnote [32]). Using this trick reduces the time of the PEA to $O(\log(1/p))$ (the number of bits to be determined), counting only the sequential uses of S and ignoring the complexity of preparing the initial states $|\psi_\phi\rangle$ and the decoding overhead in the measurement. The repetitions required for achieving the desired confidence level are trivially parallelizable and do not contribute to the time. It is possible to reduce the time from $O(\log(1/p))$ to $O(1)$ by avoiding the feed-forward phase correction used in the algorithm and reverting to the algorithm in [20] and mentioned in [2].

Based on the discussion in the previous paragraph, the main obstacle to parallelizing the OEA is the preparation of $|\psi_\phi\rangle$. If $\phi=2\arccos[|\mathrm{tr}(S\rho)|]$ is not close to 0, $|\psi_\phi\rangle$ can be prepared near deterministically with relatively few resources as follows. Suppose we have a lower bound ϵ on ϕ. With the original initial state, use sequential phase estimation with precision $\epsilon/2$ and confidence level $1-(1-c)p/B$ to determine whether we have projected onto the eigenstate $|\psi_\phi\rangle$ with eigenvalue $e^{i\phi}$ or the one with $e^{-i\phi}$. The occurrence of p in the confidence level accounts for the total number of states that need to be prepared. The parameter B is a constant that provides an additional adjustment to the confidence level. It must be chosen sufficiently large, and other confidence level parameters must be adjusted accordingly, to achieve the desired overall confidence level. If we have projected onto $|\psi_\phi\rangle$, return the state. If not, either try again or adapt the parallel PEA to use the inverse operator $S^\dagger$ instead of S for this instance of the initial state. The (sequential) resources required are of the order of $|\log((1-c)p)|/\epsilon$, but all the needed states can be prepared in parallel. For ϵ constant, the time required by the parallel PEA is increased by a factor of $O(|\log((1-c)p)|)$. The parallel overlap estimation for a unitary operator U based on these variations of phase estimation thus requires $O(|\log((1-c)p)|)$ time, provided $|\langle\psi|U|\psi\rangle|$ is not too close to 1.

For $|\langle\psi|U|\psi\rangle|$ close to 1, the OEA is intrinsically not parallelizable without increasing the total resource cost by a factor of up to $O(\sqrt{p})$. This is due to the results in [33], where it is shown that Grover's algorithm cannot be parallelized without reducing the performance to that of classical search. For example, consider the problem of determining which unique state $|k\rangle$ of the states $|0\rangle,\ldots,|2^n\rangle$ has its sign flipped by a "blackbox" unitary operator U. This can be done with n many uses of the OEA by preparing the states $|\psi_b\rangle$ that are uniform superpositions of the $|i\rangle$ for which the number i has 1 as its bth bit. If $\langle\psi_b|U|\psi_b\rangle=1-1/2^{n-2}$, then the bth bit of k is 1. If $\langle\psi_b|U|\psi_b\rangle=1$, then it is 0. It suffices to use an unparametrized (Fig. 4) precision of $1/2^{n-1}$ and confidence level sufficiently much larger than $1-1/n$. Because $|1-\cos(\phi)|=O(\phi^2)$, the parametrized precision required is $\Theta(1/2^{n/2})$. [$\Theta(x)$ is a quantity that is both $O(x)$ and $\Omega(x)$.] Thus $O(n2^{n/2})$ sequential resources suffice, which is close to the optimum attained by Grover's algorithm. However, the results of [33] imply that implementing a quantum search with depth (sequential time) d requires $\Omega(2^n/d)$ uses of U for $d<2^{n/2}$. This implies that to achieve a parametrized precision of $\Theta(1/2^{n/2})$ for $1-|\langle\psi|U|\psi\rangle|=O(1/2^n)$ using time $O(2^{n/2}/P)$ requires $\Omega(2^{n/2}P)$ resources (P represents the amount of parallelism).

The EEA was described so that overlap estimation is used with small ϕ and therefore cannot be immediately parallelized without loss of precision or larger resource requirements. However, for the version of overlap estimation needed for stages I′ and II′, it is only the imaginary part of the overlap that is needed, and the parameters are chosen so that the overlap's phase is expected to be within 1 of 0 (because $\theta_{\max}\leqslant 1$). The actual precision required is absolute in the overlap, not the parametrization of the overlap in terms of the upper hemisphere in Fig. 4. This implies that we can call the parallel overlap algorithm with an intentionally suppressed overlap. If the desired overlap is $\langle\psi|U|\psi\rangle$, one way to suppress it is to replace $U^{(\mathsf{S})}$ by ${}^cU^{(\mathsf{aS})}$ and the initial state by $(\mathbf{I}^{(\mathsf{a})}/2)|\psi\rangle_{\mathsf{S}}^{\mathsf{S}}\langle\psi|$. The suppression ensures that the phases in the calls to the PEA are sufficiently distinguishable to allow the near-deterministic preparation of the appropriate eigenstates discussed above. This adds at most a constant overhead to the EEA due to the additional precision required to account for the scaling associated with the overlap suppression.

ACKNOWLEDGMENTS

We thank Ryan Epstein and Scott Glancy for their help in reviewing and editing the manuscript. This work was carried out under the auspices of the National Nuclear Security Administration of the U.S. Department of Energy at Los Alamos National Laboratory under Contract No. DE-AC52-06NA25396.

[1] V. Giovannetti, S. Lloyd, and L. Maccone, Science **306**, 1330 (2004).
[2] V. Giovannetti, S. Lloyd, and L. Maccone, Phys. Rev. Lett. **96**, 010401 (2006).
[3] R. P. Feynman, Int. J. Theor. Phys. **21**, 467 (1982).
[4] B. M. Terhal and D. P. DiVincenzo, Phys. Rev. A **61**, 022301 (2000).
[5] G. Ortiz, J. E. Gubernatis, E. Knill, and R. Laflamme, Phys. Rev. A **64**, 022319 (2001).
[6] R. Somma, G. Ortiz, E. Knill, and J. Gubernatis, Int. J. Quantum Inf. **1**, 189 (2003).
[7] R. Somma, G. Ortiz, J. E. Gubernatis, E. Knill, and R.

Laflamme, Phys. Rev. A **65**, 042323 (2002).
[8] C. Miquel, J. P. Paz, M. Saraceno, E. Knill, R. Laflamme, and C. Negrevergne, Nature (London) **418**, 59 (2002).
[9] R. Cleve, A. Ekert, C. Macchiavello, and M. Mosca, Proc. R. Soc. London, Ser. A **454**, 339 (1998).
[10] G. Brassard, P. Høyer, M. Mosca, and A. Tapp, in *Quantum Computation and Quantum Information: A Millennium Volume*, edited by J. S. J. Lomonaco AMS Contemporary Mathematics Series (American Mathematical Society, Providence, 2000).
[11] J. J. Bollinger, W. M. Itano, D. J. Wineland, and D. J. Heinzen, Phys. Rev. A **54**, R4649 (1996).
[12] J. Preskill, Proc. R. Soc. London, Ser. A **454**, 385 (1998).
[13] A. Barenco, C. H. Bennett, R. Cleve, D. P. DiVincenzo, N. Margolus, P. Shor, T. Sleator, J. Smolin, and H. Weinfurter, Phys. Rev. A **52**, 3457 (1995).
[14] P. W. Shor, SIAM J. Comput. **26**, 1484 (1997).
[15] For the moment, we identify the precision with the maximum error. The correct formulation of precision in terms of confidences is given later. Because $\langle\psi|U|\psi\rangle \leqslant 1$, the error in the value of a^2, is at most $p/2$. Similarly, the errors in $4b_0^2$ and $4b_{\pi/2}^2$ are at most $p/2$. Here we used the fact that the parametrized precision returned by the AEA is also an upper bound on the precision of the (unparametrized) value returned. Thus $\mathrm{Re}(y)$ and $\mathrm{Im}(y)$ are both off by at most $p/2$. This error affects only the component of the returned value perpendicular to $\langle\psi|U|\psi\rangle$ in the complex plane, to which it contributes at most $p/\sqrt{2}$. When these errors are lifted back to the upper unit hemisphere, they are bounded by $\sqrt{(p/2)^2+(p/\sqrt{2})^2}=p\sqrt{3}/2$, taking into account that the precision in a is actually with respect to the parametrization. Our choice of precisions in the calls to amplitude estimation can be improved.
[16] R. B. Griffiths and C.-S. Niu, Phys. Rev. Lett. **76**, 3228 (1996).
[17] A. M. Childs, J. Preskill, and J. Renes, J. Mod. Opt. **47**, 155 (2000).
[18] J. Chiaverini *et al.*, Science **308**, 997 (2005).
[19] Using n as an explicit parameter, we have $P_0(\epsilon)=1$ and $P_n(\epsilon)=P_{n-1}(\epsilon/2)\cos(\pi\epsilon/2)^2$. The probability P that the phase returned is within 2^{-n} is the probability $P_n(\epsilon)$ of getting the best approximation, plus the probability $[1-\cos(\pi\epsilon/2)^2]=\sin(\pi\epsilon/2)^2=\cos(\pi(1-\epsilon)/2)^2$ that the most significant bit is wrong times the probability $P_{n-1}((1-\epsilon)/2)$ that all subsequent bits are best possible. Thus $P=P_n(\epsilon)+P_{n-1}((1-\epsilon)/2)\cos(\pi(1-\epsilon)/2)^2=P_n(\epsilon)+P_n(1-\epsilon)\geqslant 0.81$.
[20] A. Y. Kitaev, e-print quant-ph/9511026.
[21] A. Aspuru-Guzik, A. D. Dutoi, P. J. Love, and M. Head-Gordon, Science **309**, 1704 (2005).
[22] M. de Burgh and S. D. Bartlett, Phys. Rev. A **72**, 042301 (2005).
[23] W. Hoeffding, J. Am. Stat. Assoc. **58**, 13 (1963).
[24] J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton University Press, Princeton, 1971).
[25] D. W. Berry, G. Ahokas, R. Cleve, and B. C. Sanders, e-print quant-ph/0508139.
[26] Let C be a constant so that $G^{-1}(x)\leqslant C|\log(x)|$ for x small enough. Since we are considering the asymptotic behavior of resources for small p, x is small and the first inequality in constraint (A″) of Eq. (10) may be assumed to be satisfied. To solve constraint (A″) maximize $\theta_{\max}$ subject to $C|\log(\theta_{\max}p/8)|\leqslant(p/8)/(\theta_{\max}^2/6)$. Rewrite this inequality as $|\log(\theta_{\max}p/8)|\theta_{\max}^2\leqslant(3p/4)/C$. For p small enough (and since $\theta_{\max}\leqslant 1$) this is the same as $[|\log(\theta_{\max})|+|\log(p/8)|]\theta_{\max}^2\leqslant(3p/4)/C$. We can impose the additional constraint that $|\log(\theta_{\max})|\leqslant|\log(p/8)|$ to write the inequality in the form $\theta_{\max}^2\leqslant[(3p/4)/C]/|\log(p/8)|$ or $\theta_{\max}\leqslant D'\times[p/|\log(p)|]^{1/2}$ for some constant D'. We can try a solution $\theta_{\max}=D''[p/|\log(p)|]^{1/2}=\Omega([p/|\log(p)|]^{1/2})$. For this solution $|\log(\theta_{\max})|\leqslant|\log(D'')|+|\log(p)|/2+\log(|\log(p)|)/2\leqslant|\log(p/8)|$ for sufficiently small p. Hence the additional constraint is asymptotically satisfied.
[27] Let C be a constant so that $G^{-1}(x)\leqslant Cx^{-1/(1+\beta)}$ for x small enough. As in [26], we are interested in the behavior for small p. To solve constraint (A″), maximize $\theta_{\max}$ subject to $C(\theta_{\max}p/8)^{-1/(1+\beta)}\leqslant(p/8)/(\theta_{\max}^2/6)$. Equivalently $\theta_{\max}^{2-1/(1+\beta)}=\theta_{\max}^{(1+2\beta)/(1+\beta)}\leqslant D'p^{1+1/(1+\beta)}=D'p^{(2+\beta)/(1+\beta)}$ for some constant D'. Hence $\theta_{\max}=\Omega(p^{(2+\beta)/(1+2\beta)})$ works. From this we get $t=\theta_{\max}/G^{-1}(\theta_{\max}p/8)=\Omega(p^{(2+\beta)/(1+2\beta)}\times(p^{(2+\beta)/(1+2\beta)+1})^{1/(1+\beta)})=\Omega(p^{(5+6\beta+\beta^2)/((1+2\beta)(1+\beta))})$.
[28] The prior probability that all the y_l are within $(t/2)p/(4K2^K)$ of the true overlap in the calls to the OEA is at least $1-K(1-c)/(2K)=1-(1-c)/2$. Thus the confidence for stage II′ matches that of stage II. Assuming that all the y_l have the stated precision, the difference $|\Sigma_{l=0}^K C_l y_l-\Sigma_{l=0}^K C_l\mathrm{tr}(e^{-i(A-a_0)(lt/2)}\rho)|$ is bounded by $B=\Sigma_{l=0}^K C_l(t/2)p/(4K2^K)$. Since $|C_l|\leqslant 2^K$, $B\leqslant(t/2)p/4$. As before, the approximation error $|-\mathrm{Im}(\Sigma_{l=0}^K C_l\times\mathrm{tr}(e^{-i(A-a_0)(lt/2)}\rho))/(t/2)-\mathrm{tr}[(A-a_0)\rho]|$ has three contributions. The first is due to the error term in Eq. (7) for eigenvalues of $(A-a_0)(t/2)$ in $[-\theta_{\max},+\theta_{\max}]$. This is bounded by $\theta_{\max}^{K+1}/[(K+1)(1-\theta_{\max})^{K+1}(t/2)]$. Constraint (A) implies that it is at most $p/4$. The second and third contributions are due to other eigenvalues of $(A-a_0)(t/2)$. The contribution to the mean of these eigenvalues can be bounded by using constraints (B), (C), and (D). These eigenvalues differ from a_0 by at least $2\theta_{\max}/t$. According to constraint (D) and the correctness of stage I, they therefore differ from the mean by at least $\theta_{\max}/t$. By use of (B) and (C), their contribution to the mean is bounded by $p/(8K2^K)$. Each such eigenvalue still contributes to the values returned by the calls to the OEA, changing each y_l by at most $F(\theta_{\max}/t)\leqslant tp/(8K2^K)$, which changes the returned value by at most $p/4$.
[29] In this case $G^{-1}(x)\leqslant\lambda_{\max}$. Thus we need $\theta_{\max}^K\leqslant(p/32)(K+1)/\lambda_{\max}$ and can set $t=\theta_{\max}/\lambda_{\max}$. Asymptotically $K+1$ is equivalent to K in the expressions obtained.
[30] Use $G^{-1}(x)\leqslant C|\log(x)|$ for sufficiently small x. We therefore need $C|\log(\theta_{\max}p/(8K2^K))|\leqslant(p/32)(K+1)/\theta_{\max}^K$. Because $\log(8K2^K)\leqslant 3K$, it is sufficient to satisfy $[|\log(\theta_{\max})|+|\log(p)|+3K]\theta_{\max}^K\leqslant Kp/32$. Add the additional constraint $|\log(\theta_{\max})|\leqslant|\log(p)|$ so that $\theta_{\max}^K\leqslant Kp/[|\log(p)|+3K]$, suffices. We can therefore set $\theta_{\max}=\{Kp/[|\log(p)|+3K]\}^{1/K}=\Omega([p/|\log(p)|]^{1/K})$. Observe that $|\log(\theta_{\max})|$ satisfies the additional constraint for sufficiently small p, independent of K. To obtain t, note that $G^{-1}(\theta_{\max}p/(8K2^K))=O(|\log(p)|+K)$ where we used the order notation to absorb constants.
[31] Here $G^{-1}(x)\leqslant Cx^{-1/(1+\beta)}$ for sufficiently small x, so we solve $[\theta_{\max}p/(8K2^K)]^{-1/(1+\beta)}\leqslant(p/32)(K+1)/\theta_{\max}^K$. It is sufficient to solve $\theta_{\max}^{K-1/(1+\beta)}\leqslant D2^{-K/(1+\beta)}p^{1+1/(1+\beta)}$

for some sufficiently small constant D. Here we used the fact that $(K+1)/K^{1/(1+\beta)}=\Omega(1)$. Thus we can set $\theta_{\max}=\Omega(2^{-K/(K-1+K\beta)}p^{(2+\beta)/(K-1+K\beta)})$. The first factor is $\Omega(1)$. To bound t, $G^{-1}(\theta_{\max}p/(8K2^K))=O(2^{K/(1+\beta)}\times p^{-(2+\beta)/[(1+\beta)(K-1+K\beta)]-1/(1+\beta)})$. Thus $t=\Omega(2^{-K/(1+\beta)}\times p^{(2+\beta)^2/[(1+\beta)(K-1+K\beta)]+1/(1+\beta)})$.

[32] It suffices to assign the qubits to the leaves of a binary tree. The decoding proceeds recursively by applying controlled NOT operations to pairs of leaves with a common parent, removing the target qubit, assigning the control qubit to the parent, and removing the leaves from the tree. The qubit that ends up at the root of the tree is measured in the $|+\rangle,|-\rangle$ basis.

[33] C. Zalka, Phys. Rev. A **60**, 2746 (1999).