Nearly Optimal Quantum Algorithm for Estimating Multiple Expectation Values

William J. Huggins,[1,*] Kianna Wan,[1,2] Jarrod McClean,[1] Thomas E. O'Brien,[1] Nathan Wiebe,[3,4,†] and Ryan Babbush[1,‡]
[1]Google Quantum AI, Mountain View, 94043 California, USA
[2]Stanford Institute for Theoretical Physics, Stanford University, Stanford, California 94305, USA
[3]University of Toronto, Toronto, Ontario ON M5S, Canada
[4]Pacific Northwest National Laboratory, Richland, 99354 Washington, USA

(Received 19 January 2022; revised 30 August 2022; accepted 30 September 2022; published 5 December 2022)

Many quantum algorithms involve the evaluation of expectation values. Optimal strategies for estimating a single expectation value are known, requiring a number of state preparations that scales with the target error ε as $\mathcal{O}(1/\varepsilon)$. In this Letter, we address the task of estimating the expectation values of M different observables, each to within additive error ε, with the same $1/\varepsilon$ dependence. We describe an approach that leverages Gilyén *et al.*'s quantum gradient estimation algorithm to achieve $\mathcal{O}(\sqrt{M}/\varepsilon)$ scaling up to logarithmic factors, regardless of the commutation properties of the M observables. We prove that this scaling is worst-case optimal in the high-precision regime if the state preparation is treated as a black box, even when the operators are mutually commuting. We highlight the flexibility of our approach by presenting several generalizations, including a strategy for accelerating the estimation of a collection of dynamic correlation functions.

DOI: 10.1103/PhysRevLett.129.240501

Introduction.—A fundamental task of quantum simulation is to perform an experiment *in silico*. Like traditional experimentalists, researchers using quantum computers will often be interested in efficiently measuring a collection of properties. For example, the electronic ground state problem is frequently cited as a motivation for quantum simulation of chemistry, but determining the ground state energy is only a starting point in most chemical applications. Depending on context, it may be essential to measure the dipole moment and polarizability, the electron density, the forces experienced by the classical nuclei, or various other quantities [1,2]. Similarly, in condensed matter physics and beyond, correlation functions play a central role in the theory of quantum many-body phenomena due to their interpretability and measurability in the lab [3,4].

In this Letter, we consider the problem of accurately and efficiently estimating multiple properties from a quantum computation. We focus on evaluating the expectation values of a collection of M Hermitian operators $\{O_j\}$ with respect to a pure state $|\psi\rangle$. We aim to evaluate each expectation value to within additive error ε using as few calls as possible to a state preparation oracle for $|\psi\rangle$ (or its inverse). One simple approach is to repeatedly prepare $|\psi\rangle$ and projectively measure mutually commuting subsets of $\{O_j\}$. Alternatively, strategies based on amplitude estimation achieve a quadratic speedup with respect to ε but entail measuring each observable separately [5–7]. A range of newer "shadow tomography" techniques use joint measurements of multiple copies of $|\psi\rangle$ to achieve polylogarithmic scaling with respect to M at the expense of an unfavorable $1/\varepsilon^4$ scaling [8–11]. In certain situations, randomized methods based on the idea of "classical shadows" of the state obtain $1/\varepsilon^2$ scaling while improving upon sampling protocols with deterministic measurement settings [12,13]. We review these existing approaches in Supplemental Material, Sec. I and compare them to our new strategy in Table I and Supplemental Material, Sec. II [14].

Our main contribution is an algorithm that achieves the same $1/\varepsilon$ scaling as methods based on amplitude estimation, but also improves the scaling with respect to M from $\tilde{\mathcal{O}}(M)$ to $\tilde{\mathcal{O}}(\sqrt{M})$, where the tilde in $\tilde{\mathcal{O}}(\cdot)$ hides logarithmic factors. Our approach is to construct a function f whose gradient yields the expectation values of interest and encode f in a parametrized quantum circuit. We can then apply Gilyén *et al.*'s quantum algorithm for gradient estimation [25] to obtain the desired scaling. The following theorem formalizes our result.

Theorem 1: Let $\{O_j\}$ be a set of M Hermitian operators on N qubits, with spectral norms $\|O_j\| \leq 1$ for all j. There exists a quantum algorithm that, for any N-qubit quantum state $|\psi\rangle$ prepared by a unitary U_ψ, outputs estimates $\widetilde{o_j}$ such that $|\widetilde{o_j} - \langle\psi|O_j|\psi\rangle| \leq \varepsilon$ for all j with probability at least 2/3, using $\tilde{\mathcal{O}}(\sqrt{M}/\varepsilon)$ queries to U_ψ and $U_\psi^\dagger$, along with $\tilde{\mathcal{O}}(\sqrt{M}/\varepsilon)$ gates of the form

TABLE I. A comparison of the (worst-case) complexities, in terms of state preparation oracle queries, of different approaches for measuring multiple observables. We consider three applications: estimating the expectation values of M commuting or non-commuting observables, and determining the fermionic k-RDM of an N-mode system. Here, ε denotes the additive error to which each quantity is estimated. We compare strategies based on naive sampling, amplitude estimation, and shadow tomography to our gradient-based approach. We cite the specific works used to determine these complexities, including the Pauli-specific shadow protocol of Ref. [11]. Note that methods based on sampling and shadow tomography also work under a weaker input model where only copies of the state are provided.

	Commuting	Non-commuting	k-RDM
Sampling	$\mathcal{O}(\log M/\varepsilon^2)$	$\tilde{\mathcal{O}}(M/\varepsilon^2)$	$\tilde{\mathcal{O}}(N^k/\varepsilon^2)$ [13]
Amplitude Estimation [6]	$\tilde{\mathcal{O}}(M/\varepsilon)$	$\tilde{\mathcal{O}}(M/\varepsilon)$	$\tilde{\mathcal{O}}(N^{2k}/\varepsilon)$
Shadow Tom. [11]	$\mathcal{O}(\log M/\varepsilon^4)$	$\mathcal{O}(\log M/\varepsilon^4)$	$\mathcal{O}(k \log N/\varepsilon^4)$
Gradient	$\tilde{\mathcal{O}}(\sqrt{M}/\varepsilon)$	$\tilde{\mathcal{O}}(\sqrt{M}/\varepsilon)$	$\tilde{\mathcal{O}}(N^k/\varepsilon)$

controlled-e^{-ixO_j} for each j, for various values of x with $|x| \in \mathcal{O}(1/\sqrt{M})$.

As we show in Corollary 3, this query complexity is worst-case optimal (up to logarithmic factors) in the high-precision regime where $\varepsilon \in [0, (1/3\sqrt{M})]$. After establishing this lower bound for our problem, we review the gradient algorithm of Ref. [25] and present the proof of Theorem 1. We then discuss several extensions of our approach, including a strategy for estimating multiple dynamic correlation functions and a method that handles observables with arbitrary norms (or precision requirements) based on a generalization of the gradient algorithm.

Lower bounds.—In Ref. [26], Apeldoorn proved a lower bound for a task that is essentially a special case of our quantum expectation value problem. We explain how a lower bound for our problem can be obtained as a corollary. Their results are expressed in terms of a particular quantum access model for classical probability distributions:

Definition 1 (Sample oracle for a probability distribution). Let $\mathbf{p}$ be a probability distribution over M outcomes, i.e., $\mathbf{p} \in [0,1]^M$ with $\|\mathbf{p}\|_1 = 1$. A *sample oracle* $U_\mathbf{p}$ for $\mathbf{p}$ is a unitary operator that acts as

$$U_\mathbf{p} : |0\rangle|0\rangle \mapsto \sum_{j=1}^{M} \sqrt{p_j}\,|j\rangle \otimes |\phi_j\rangle, \tag{1}$$

where the $|\phi_j\rangle$ are arbitrary normalized quantum states.

We rephrase Lemma 13 of Ref. [26] below. Here and throughout this Letter, we count queries to a unitary oracle U and to its inverse $U^\dagger$ as equivalent in cost.

Theorem 2: [Lemma 13, [26] (rephrased)] Let M be a positive integer power of 2 and let $\varepsilon \in [0, (1/3\sqrt{M})]$. There exists a known matrix $A \in \{-1,+1\}^{M\times M}$ such that the following is true. Suppose $\mathcal{A}$ is an algorithm that, for every probability distribution $\mathbf{p}$, accessed via a sample oracle $U_\mathbf{p}$, outputs (with probability at least 2/3) a $\tilde{\mathbf{q}}$ such that $\|A\mathbf{p} - \tilde{\mathbf{q}}\|_\infty \le \varepsilon$. Then $\mathcal{A}$ must use $\Omega(\sqrt{M}/\varepsilon)$ queries to $U_\mathbf{p}$ in the worst case.

We can use this Theorem to derive the following corollary, establishing the near-optimality of the algorithm in Theorem 1 in certain regimes.

Corollary 3: Let M be a positive integer power of 2 and let $\varepsilon \in [0, (1/3\sqrt{M})]$. Let $\mathcal{A}$ be any algorithm that takes as an input an arbitrary set of M observables $\{O_j\}$. Suppose that, for every quantum state $|\psi\rangle$, accessed via a state preparation oracle U_ψ, $\mathcal{A}$ outputs estimates of each $\langle\psi|O_j|\psi\rangle$ to within additive error ε (with probability at least 2/3). Then, there exists a set of observables $\{O_j\}$ such that $\mathcal{A}$ applied to $\{O_j\}$ must use $\Omega(\sqrt{M}/\varepsilon)$ queries to U_ψ.

Proof.—Assume for the sake of contradiction that for any $\{O_j\}$ and U_ψ, the algorithm $\mathcal{A}$ uses $o(\sqrt{M}/\varepsilon)$ queries to U_ψ to estimate every $\langle\psi|O_j|\psi\rangle$ to within error ε (with success probability at least 2/3). For any sample oracle $U_\mathbf{p}$ of the form in Eq. (1), consider the state

$$|\psi(U_\mathbf{p})\rangle := \sum_{j=1}^{M} \sqrt{p_j} \left(\bigotimes_{i=1}^{M} \left| \frac{1 - A_{ij}}{2} \right\rangle \right) \otimes |j\rangle \otimes |\phi_j\rangle. \tag{2}$$

A quick computation verifies that the ith entry of the vector $A\mathbf{p}$ is equal to $\langle\psi(U_\mathbf{p})|Z_i|\psi(U_\mathbf{p})\rangle$, where Z_i denotes the Pauli Z operator acting on the ith qubit. Since the matrix A is known, it is clear that $|\psi(U_\mathbf{p})\rangle = U_A(I \otimes U_\mathbf{p})|0\rangle$ for a known unitary U_A:

$$U_A = \sum_j \left(\bigotimes_{i=1}^{M} X_i^{\delta_{A_{ij},-1}} \right) \otimes |j\rangle\langle j| \otimes \mathbb{I}. \tag{3}$$

Therefore, we can apply algorithm $\mathcal{A}$ with $O_j = Z_j$ for $j \in \{1, \ldots, M\}$ and $U_\psi = U_A(\mathbb{I} \otimes U_\mathbf{p})$. By our assumption, this constitutes an algorithm that for every $U_\mathbf{p}$, estimates each entry of $A\mathbf{p}$ to within error ε using $o(\sqrt{M}/\varepsilon)$ queries to $U_\mathbf{p}$, contradicting Theorem 2, and completing the proof. ∎

Background on Gilyén et al.'s gradient algorithm.—Our framework for simultaneously estimating multiple expectation values uses the improved quantum algorithm for gradient estimation of Gilyén, Arunachalam, and Wiebe (henceforth, Gilyén *et al.*) [25]. Gilyén *et al.* built on earlier work by Jordan [27], which demonstrated an exponential quantum speedup for computing the gradient in a particular black-box access model. Specifically, Jordan's algorithm uses one query to a *binary oracle* (see Supplemental Material, Sec. III [14]) for a function f, along with phase kickback and the quantum Fourier transform, to obtain an approximation of the gradient ∇f.

While we defer a technical discussion of Gilyén *et al.*'s algorithm to Supplemental Material, Sec. III [14] (and we

refer the reader also to Ref. [25]), we give a brief, colloquial description of their algorithm here. It is helpful to review their definition for a *probability oracle*.

Definition 2 (Probability oracle). Consider a function $f:\mathbb{R}^M \to [0,1]$. A probability oracle U_f for f is a unitary operator that acts as

$$U_f:|\boldsymbol{x}\rangle|\mathbf{0}\rangle \mapsto |\boldsymbol{x}\rangle(\sqrt{f(\boldsymbol{x})}|1\rangle|\phi_1(\boldsymbol{x})\rangle + \sqrt{1-f(\boldsymbol{x})}|0\rangle|\phi_0(\boldsymbol{x})\rangle), \tag{4}$$

where $|\boldsymbol{x}\rangle$ denotes a discretization of the variable $\boldsymbol{x}$ encoded into a register of qubits, $|\mathbf{0}\rangle$ denotes the all-zeros state of a register of ancilla qubits, and $|\phi_0(\boldsymbol{x})\rangle$ and $|\phi_1(\boldsymbol{x})\rangle$ are arbitrary quantum states.

Gilyén *et al.* show how such a probability oracle can be used to encode a finite-difference approximation to a directional derivative of f in the phase of an ancilla register, e.g., a first-order approximation is implemented by

$$A_{f_1'}:|\boldsymbol{x}\rangle|\mathbf{0}\rangle \mapsto e^{i(f(\boldsymbol{x})-f(-\boldsymbol{x}))}|\boldsymbol{x}\rangle|\mathbf{0}\rangle. \tag{5}$$

As in Jordan's original algorithm, a quantum Fourier transform can then be used to extract an approximate gradient from the phases accumulated on an appropriate superposition of basis states. By using higher-order finite-difference formulas, Gilyén *et al.* are able to estimate the gradient with a scaling that is optimal (up to logarithmic factors) for a particular family of smooth functions. We restate the formal properties of their algorithm in the theorem below.

Theorem 4: [Theorem 25, Ref. [25] (rephrased)] Let ε, $c \in \mathbb{R}_+$ be fixed constants, with $\varepsilon \le c$. Let $M \in \mathbb{Z}_+$ and $\boldsymbol{x} \in \mathbb{R}^M$. Suppose that $f:\mathbb{R}^M \to \mathbb{R}$ is an analytic function such that for every $k \in \mathbb{Z}_+$, the following bound holds for all kth order partial derivatives of f at $\boldsymbol{x}$ (denoted by $\partial_{\boldsymbol{\alpha}} f(\boldsymbol{x})$): $|\partial_{\boldsymbol{\alpha}} f(\boldsymbol{x})| \le c^k k^{(k/2)}$. Then, there is a quantum algorithm that outputs an estimate $\tilde{\boldsymbol{g}} \in \mathbb{R}^M$ such that $\|\nabla f(\boldsymbol{x}) - \tilde{\boldsymbol{g}}\|_\infty \le \varepsilon$, with probability at least $1-\delta$. This algorithm makes $\tilde{\mathcal{O}}(c\sqrt{M}\log(M/\delta)/\varepsilon)$ queries to a probability oracle for f.

Expectation values via the gradient algorithm.—To construct our algorithm and prove Theorem 1, we build a probability oracle for a function whose gradient encodes the expectation values of interest and apply the quantum algorithm for the gradient.

Proof of Theorem 1.—We begin by defining the parametrized unitary

$$U(\boldsymbol{x}) \coloneqq \prod_{j=1}^{M} e^{-2ix_j O_j} \tag{6}$$

for $\boldsymbol{x} \in \mathbb{R}^M$. The derivative of this unitary with respect to x_ℓ is

$$\frac{\partial U}{\partial x_\ell} = -2i\left(\prod_{j=1}^{\ell} e^{-2ix_j O_j}\right) O_\ell \left(\prod_{k=\ell+1}^{M} e^{-2ix_k O_k}\right). \tag{7}$$

We are interested in the expectation of the O_j with respect to the state $|\psi\rangle$, so we define the following function f:

$$f(\boldsymbol{x}) \coloneqq -\frac{1}{2}\mathrm{Im}[\langle\psi|U(\boldsymbol{x})|\psi\rangle] + \frac{1}{2}. \tag{8}$$

Using Eq. (7), we have

$$\left.\frac{\partial f}{\partial x_\ell}\right|_{\boldsymbol{x}=\mathbf{0}} = \langle\psi|O_\ell|\psi\rangle. \tag{9}$$

Therefore, the gradient $\nabla f(\mathbf{0})$ is precisely the collection of expectation values of interest.

Now, we verify that f satisfies the conditions of Theorem 4. Observe that f is analytic and that the kth order partial derivative of f with respect to any collection of indices $\alpha \in \{1,\ldots,M\}^k$ takes the form

$$\partial_\alpha f(\boldsymbol{x}) = (-2)^{k-1}\mathrm{Im}(i^k\langle\psi|V(\boldsymbol{x},\boldsymbol{\alpha})|\psi\rangle), \tag{10}$$

for some operator $V(\boldsymbol{x},\alpha)$ which depends on both α and $\boldsymbol{x}$. Note that V is a product of terms which are either unitary, or from $\{O_j\}$. Since $\|O_j\| \le 1$ for all j, we have $\|V\| \le 1$, and therefore $|\partial_\alpha f(\mathbf{0})| \le 2^{k-1}$ for all k and α. By setting $c=2$, we satisfy the derivative conditions of Theorem 4.

To construct a probability oracle for f (see Definition 2), we need a quantum circuit that encodes $f(\boldsymbol{x})$ into the amplitudes of an ancilla. We construct such a circuit using the Hadamard test for the imaginary component of $\langle\psi|U(\boldsymbol{x})|\psi\rangle$ [28,29]. Let

$$F(\boldsymbol{x}) \coloneqq (H\otimes\mathbb{I})(\mathrm{C}\text{-}U(\boldsymbol{x}))(S^\dagger H\otimes U_\psi), \tag{11}$$

where H denotes the Hadamard gate, C-$U(\boldsymbol{x})$ the $U(\boldsymbol{x})$ gate controlled on the first qubit, and $S \coloneqq |0\rangle\langle 0| + i|1\rangle\langle 1|$ the phase gate. Applied to $|0\rangle\otimes|\mathbf{0}\rangle$, this circuit encodes $f(\boldsymbol{x})$ in the amplitudes with respect to the computational basis states of the first qubit:

$$F(\boldsymbol{x})|0\rangle\otimes|\mathbf{0}\rangle = \sqrt{f(\boldsymbol{x})}|1\rangle\otimes|\phi_1(\boldsymbol{x})\rangle + \sqrt{1-f(\boldsymbol{x})}|0\rangle\otimes|\phi_0(\boldsymbol{x})\rangle, \tag{12}$$

for some normalized states $|\phi_0(\boldsymbol{x})\rangle$ and $|\phi_1(\boldsymbol{x})\rangle$ (see Supplemental Material, Sec. IV [14] for more details). Note that $F(\boldsymbol{x})$ uses a single call to the oracle U_ψ.

All that remains is to add quantum controls to the rotations in $F(\boldsymbol{x})$, so that $F(\boldsymbol{x})$ is controlled on a register encoding $\boldsymbol{x}$. Specifically, we consider the unitary

$$U_f \coloneqq \sum_{\boldsymbol{k}\in G_n^M} |\boldsymbol{k}\rangle\langle\boldsymbol{k}|\otimes F(\boldsymbol{k}x_{\max}), \tag{13}$$

where G_n^M is a set of 2^{nM} points distributed in an M-dimensional unit hypercube, with $n = \mathcal{O}(\log(1/\varepsilon))$, and $x_{\max}$ is a rescaling factor. The values of $x_{\max}$ and n are chosen to satisfy the requirements of the gradient algorithm (see Supplemental Material, Sec. IV [14]). Here, $|\boldsymbol{k}\rangle = |k_1\rangle \ldots |k_M\rangle$ for $\boldsymbol{k} \in G_n^M$ denotes the basis state storing the binary representation of $\boldsymbol{k}$ in M n-qubit index registers. The controlled time evolution operator for each O_j can be implemented efficiently as a product of n controlled-e^{-ixO_j} gates with exponentially spaced values of x, each controlled on the appropriate qubit of the jth index register. We illustrate an example of such a U_f in Fig. 1.

U_f is a probability oracle for the function f, and each call to U_f involves a single call to the state preparation oracle U_ψ. Theorem 4 then implies that with probability at least $2/3$, every component of the gradient of f, and hence all of the expectation values $\langle\psi|O_j|\psi\rangle$, can be estimated to within an error ε using $\tilde{\mathcal{O}}(\sqrt{M}/\varepsilon)$ queries to U_f. The complexity in terms of the controlled time evolutions follows from multiplying the number of controlled time evolutions required for each query to U_f, i.e., $\mathcal{O}(\log(M/\varepsilon))$ per observable, by the total number of queries, i.e., $\tilde{\mathcal{O}}(\sqrt{M}/\varepsilon)$. As discussed in Supplemental Material, Sec. IV [14], we have $x_{\max} \in \mathcal{O}(1/\sqrt{M})$ as a consequence of the details of the proof of Theorem 4 in Ref. [25]. This completes the proof of Theorem 1. ∎

Furthermore (see Supplemental Material, Sec. IV [14]), the space complexity of the gradient algorithm is the same as that of the probability oracle up to an additive logarithmic factor [30]. Therefore, our algorithm uses $\mathcal{O}(M\log(1/\varepsilon) + N)$ qubits.

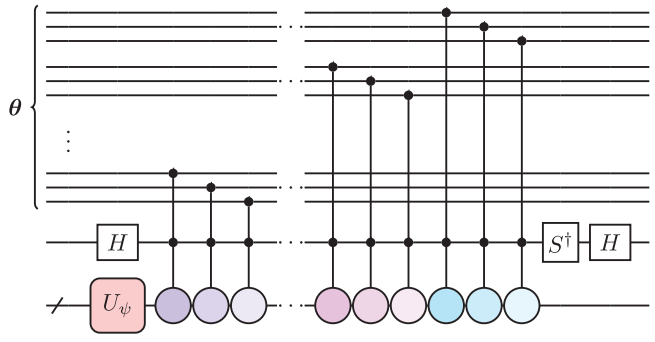

FIG. 1. Schematic depiction of the quantum circuit for U_f, the probability oracle for the function $f(\boldsymbol{x})$ defined in Eq. (13). The top registers encode the ($n = 3$ bit in this case) binary representations of $x_1, x_2, \ldots, x_M$. The ancilla qubit whose amplitudes encodes $f(\boldsymbol{x})$ [cf. Eq. (11)] is indicated below the $\boldsymbol{x}$ registers. The final line represents the N-qubit system register. The gates that act on the system register with colored circles represent the doubly controlled time evolution by the various observables. Estimating the expectation values of the M observables $\{O_j\}$ requires executing this circuit and its inverse $\tilde{\mathcal{O}}(\sqrt{M}/\varepsilon)$ times.

Discussion.—In this Letter, we considered the problem of simultaneously estimating the expectation values of multiple observables with respect to a pure state $|\psi\rangle$. We presented an algorithm that uses $\tilde{\mathcal{O}}(\sqrt{M}\varepsilon^{-1})$ applications of U_ψ and its inverse, where M denotes the number of observables and ε the target error, and U_ψ is a unitary that prepares $|\psi\rangle$. We explained how a lower bound on a closely related problem posed in Ref. [26] implies that, for algorithms given black-box access to U_ψ, this query complexity is worst-case optimal up to logarithmic factors when $\varepsilon \in [0, (1/3\sqrt{M})]$. In fact, our algorithm affirmatively resolves an open question from Ref. [26] regarding the achievability of this bound for the simultaneous estimation of classical random variables [31]. These results imply that the optimal cost for expectation value estimation can become exponentially worse with respect to M when one demands a scaling that goes as ε^{-1} instead of ε^{-2}. Furthermore, the instances used in establishing our lower bounds involve a set of mutually commuting observables, implying that commutativity is not necessarily helpful when one demands ε^{-1} scaling.

We presented a comparison with other approaches for the estimation of expectation values in Table I, which we elaborate on in Supplemental Material, Secs. I,II [14]. For example, we find that our algorithm is capable of estimating each element of the k-body fermionic reduced density matrix (k-RDM) of an N-mode system to within error ε using $\tilde{\mathcal{O}}(N^k/\varepsilon)$ state preparation queries. This offers an unconditional asymptotic speedup compared to existing methods when $\varepsilon = o(N^{-k/3})$. This may be particularly useful in practical applications where we wish to achieve a fixed error in extensive quantities by measuring the 1 or 2-RDM and summing $\Omega(N)$ elements.

Our gradient-based approach to estimating expectation values can be extended to other properties. For example, consider the task of evaluating a collection of two-point dynamic correlation functions. These functions take the form

$$C_{A,B}(t) \coloneqq \langle\psi|U(0,t)A^\dagger U(t,0)B|\psi\rangle, \tag{14}$$

where A and B are some simple operators and $U(t,t')$ is the time evolution operator that maps the system from time t' to time t. These correlation functions are often directly accessible in experiment, as in the case of angle-resolved photoemission spectroscopy [3], and are also central to hybrid quantum-classical methods based on dynamical mean-field theory [32–34]. In Supplemental Material, Sec. V [14], we explain how a generalization of our approach can reduce the number of state preparations required for estimating a collection of these correlation functions.

Although we focused on quantifying the number of state preparation oracle queries, we also considered two other complexity measures. Our approach requires time

evolution by each of the M observables. The total duration of time evolution required scales as $\tilde{\mathcal{O}}(M/\varepsilon)$. We also need an additional $\tilde{\mathcal{O}}(M\log(1/\varepsilon))$ qubits, although we can modify our approach to trade off between space and query complexities (see Supplemental Material, Sec. VI [14]). When we are interested in simultaneously estimating $O(N)$ expectation values, the asymptotic scaling of the space complexity is only logarithmically larger than that of storing the system itself. This is the case in a variety of contexts, for example, in the evaluation of the momentum distribution [35]. In other situations, the space overhead may be more substantial, though the capability of modern simulation algorithms to use so-called "dirty ancilla" (temporarily borrowing qubits in an arbitrary state) may offset this challenge in some contexts [36–38]. As a concrete example, we consider the double-factorized simulation of the electronic structure Hamiltonian proposed in Ref. [37]. Von Burg *et al.* find that the time complexity of their simulation algorithm can be minimized by using $\tilde{\mathcal{O}}(N^{3/2})$ qubits for data lookup. These same qubits could be used by our algorithm for expectation value estimation to parallelize the measurement of $\tilde{\mathcal{O}}(N^{3/2})$ observables, offering a $\tilde{\mathcal{O}}(N^{3/4})$ asymptotic speedup without any additional qubit overhead.

Another potential reason for modifying our approach arises when the observables of interest have different norms, or when the desired precision varies. In Supplemental Material, Sec. VII [14], we consider addressing this situation by measuring certain observables using our strategy and measuring others using a sampling-based method. In Supplemental Material, Sec. VIII [14], we take a different approach, and generalize Gilyén *et al.'s* gradient estimation algorithm to accommodate functions whose gradient components are not necessarily uniformly bounded. This allows us to simultaneously estimate the expectation values of observables $\{O_j\}$ with arbitrary norms $\|O_j\|$ (possibly greater than 1) using $\tilde{\mathcal{O}}(\sqrt{\sum_j \|O_j\|^2}/\varepsilon)$ queries. By rescaling the individual observables we can then also vary how precisely we estimate each expectation value, thereby extending Theorem 1 to the most general setting.

Our focus has been on the asymptotic scaling of our approach, but it will also be desirable to understand the actual costs. Performing a fault-tolerant resource estimate and a comparison against other measurement strategies in the context of a practical application would be a useful line of future work. It is possible that our approach could be modified to obtain a further speedup by taking advantage of the structure of the states and/or observables for particular problems of interest. Another potentially fruitful direction would be to explore extensions of the gradient algorithm to yield quantum algorithms for the Hessian or even higher-order derivatives.

Extracting useful information from a quantum computation, especially a quantum simulation, is a bottleneck for many applications. This is especially true in fields such as quantum chemistry and materials science, where it may be necessary to couple high-level quantum calculations with coarser approximations at other length scales in order to describe macroscopic physical phenomena. We expect that our gradient-based approach to the estimation of expectation values will be a useful tool and a starting point for related approaches to other problems.

The authors thank Bryan O'Gorman, Yuan Su, and Joonho Lee for helpful discussions and various referees for their constructive input. N. W. worked on this project under a research grant from Google Quantum AI and was also supported by the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers, Co-Design Center for Quantum Advantage under Contract No. DE-SC0012704. Some discussion and collaboration on this project occurred while using facilities at the Kavli Institute for Theoretical Physics, supported in part by the National Science Foundation under Grant No. NSF PHY-1748958.

*Corresponding author.
whuggins@google.com
†Corresponding author.
nawiebe@cs.toronto.edu
‡Corresponding author.
babbush@google.com

[1] Peter Pulay, Geza Fogarasi, Frank Pang, and James E. Boggs, Systematic ab initio gradient calculation of molecular geometries, force constants, and dipole moment derivatives, J. Am. Chem. Soc. **101**, 2550 (1979).
[2] J. K. Gregory, D. C. Clary, K. Liu, M. G. Brown, and R. J. Saykally, The water dipole moment in water clusters, Science **275**, 814 (1997).
[3] A Damascelli, Probing the electronic structure of complex systems by ARPES, Phys. Scr. **2004**, 61 (2004).
[4] G Rickayzen, *Green's Functions and Condensed Matter* (Dover Publications, Mineola, 2013).
[5] Gilles Brassard, Peter Hoyer, Michele Mosca, and Alain Tapp, Quantum amplitude amplification and estimation, arXiv:quant-ph/0005055.
[6] Emanuel Knill, Gerardo Ortiz, and Rolando D. Somma, Optimal quantum measurements of expectation values of observables, Phys. Rev. A **75**, 012328 (2007).
[7] Patrick Rall, Quantum algorithms for estimating physical quantities using block encodings, Phys. Rev. A **102**, 022408 (2020).
[8] Scott Aaronson, Shadow tomography of quantum states, SIAM J. Comput. **49**, STOC18-368 (2020).
[9] Fernando G. S. Brandão, Amir Kalev, Tongyang Li, Cedric Yen-Yu Lin, Krysta M. Svore, and Xiaodi Wu, Quantum SDP solvers: Large speed-ups, optimality, and applications to quantum learning, arXiv:1710.02581.

[10] Joran van Apeldoorn and András Gilyén, *Improvements in Quantum SDP-solving with Applications* (Schloss Dagstuhl—Leibniz-Zentrum fuer Informatik GmbH, Wadern/Saarbruecken, Germany, 2019).
[11] Hsin-Yuan Huang, Richard Kueng, and John Preskill, Information-Theoretic Bounds on Quantum Advantage in Machine Learning, Phys. Rev. Lett. **126**, 190505 (2021).
[12] Hsin-Yuan Huang, Richard Kueng, and John Preskill, Predicting many properties of a quantum system from very few measurements, Nat. Phys. **16**, 1050 (2020).
[13] Andrew Zhao, Nicholas C. Rubin, and Akimasa Miyake, Fermionic Partial Tomography via Classical Shadows, Phys. Rev. Lett. **127**, 110504 (2021).
[14] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.129.240501 for various extensions and applications of the results presented in the main text, which includes Refs. [15–24].
[15] Lin Lin and Yu Tong, Near-optimal ground state preparation, Quantum **4**, 372 (2020).
[16] Kianna Wan and Isaac Kim, Fast digital methods for adiabatic state preparation, arXiv:2004.04164.
[17] Vladyslav Verteletskyi, Tzu-Ching Yen, and Artur F. Izmaylov, Measurement optimization in the variational quantum eigensolver using a minimum clique cover, J. Chem. Phys. **152**, 124114 (2020).
[18] William J. Huggins, Jarrod R. McClean, Nicholas C. Rubin, Zhang Jiang, Nathan Wiebe, K. Birgitta Whaley, and Ryan Babbush, Efficient and noise resilient measurements for quantum chemistry on near-term quantum computers, Quantum Inf. **7**, 23 (2021).
[19] Senrui Chen, Wenjun Yu, Pei Zeng, and Steven T. Flammia, Robust shadow estimation, PRX Quantum **2**, 030348 (2021).
[20] Charles Hadfield, Sergey Bravyi, Rudy Raymond, and Antonio Mezzacapo, Measurements of quantum Hamiltonians with locally-biased classical shadows, arXiv:2006.15788.
[21] Ernest Davidson, *Reduced Density Matrices in Quantum Chemistry* (Elsevier, New York, 2012).
[22] Xavier Bonet-Monroig, Ryan Babbush, and Thomas E. O'Brien, Nearly Optimal Measurement Scheduling for Partial Tomography of Quantum States, Phys. Rev. X **10**, 031064 (2020).
[23] R. Somma, G. Ortiz, J. E. Gubernatis, E. Knill, and R. Laflamme, Simulating physical phenomena by quantum networks, Phys. Rev. A **65**, 042323 (2002).
[24] Michael A. Nielsen and Isaac L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, Cambridge, England, 2010).
[25] András Gilyén, Srinivasan Arunachalam, and Nathan Wiebe, Optimizing quantum optimization algorithms via faster quantum gradient computation, in *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms* (Society for Industrial and Applied Mathematics, Philadelphia, PA, 2019), pp. 1425–1444.
[26] J. van Apeldoorn, Quantum probability oracles & multidimensional amplitude estimation, *16th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2021)* (Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Dagstuhl, 2021).
[27] Stephen P. Jordan, Fast Quantum Algorithm for Numerical Gradient Estimation, Phys. Rev. Lett. **95**, 050501 (2005).
[28] A Yu. Kitaev, Quantum measurements and the Abelian stabilizer problem, arXiv:quant-ph/9511026.
[29] Dorit Aharonov, Vaughan Jones, and Zeph Landau, A polynomial quantum algorithm for approximating the jones polynomial, Algorithmica **55**, 395 (2009).
[30] To achieve this space complexity we actually need to compile the circuits for a logarithmic (in M and ε^{-1}) number of probability oracles across a series of hypercubes of varying sizes. Otherwise there would be additional multiplicative logarithmic factors in the space complexity.
[31] Specifically, any matrix $A \in [-1,1]^M$ can be encoded in a known unitary U_A along similar lines as in the proof of Corollary cor:lower_bounds; then $A\mathbf{p}$ can be estimated by applying our algorithm to the single-qubit Z operators $\{Z_j\}$, with state preparation oracle $U_A(\mathbb{I} \otimes U_{\mathbf{p}})$.
[32] Bela Bauer, Dave Wecker, Andrew J. Millis, Matthew B. Hastings, and Matthias Troyer, Hybrid Quantum-Classical Approach to Correlated Materials, Phys. Rev. X **6**, 031045 (2016).
[33] A. Georges and G. Kotliar, Hubbard model in infinite dimensions, Phys. Rev. B **45**, 6479 (1992).
[34] G. Kotliar, S. Y. Savrasov, K. Haule, V. S. Oudovenko, O. Parcollet, and C. A. Marianetti, Electronic structure calculations with dynamical mean-field theory, Rev. Mod. Phys. **78**, 865 (2006).
[35] M. Meckel, D. Comtois, D. Zeidler, A. Staudte, D. Pavicic, H. C. Bandulet, H. Pépin, J. C. Kieffer, R. Dörner, D. M. Villeneuve, and P. B. Corkum, Laser-induced electron tunneling and diffraction, Science **320**, 1478 (2008).
[36] Joonho Lee, Dominic W. Berry, Craig Gidney, William J. Huggins, Jarrod R. McClean, Nathan Wiebe, and Ryan Babbush, Even more efficient quantum computations of chemistry through tensor hypercontraction, PRX Quantum **2**, 030305 (2021).
[37] Vera von Burg, Guang Hao Low, Thomas Häner, Damian S. Steiger, Markus Reiher, Martin Roetteler, and Matthias Troyer, Quantum computing enhanced computational catalysis, Phys. Rev. Res. **3**, 033055 (2021).
[38] Guang Hao Low, Vadym Kliuchnikov, and Luke Schaeffer, Trading t-gates for dirty qubits in state preparation and unitary synthesis, arXiv:1812.00954.