Qubit-Efficient Randomized Quantum Algorithms for Linear Algebra

Samson Wang,[1,*] Sam McArdle,[2,3] and Mario Berta[4,5]

[1] *Department of Physics, Imperial College London, London SW7 2BW, United Kingdom*
[2] *AWS Center for Quantum Computing, Pasadena, CA 91106, USA*
[3] *Institute for Quantum Information and Matter, Caltech, Pasadena, CA 91125, USA*
[4] *Department of Computing, Imperial College London, London SW7 2BW, United Kingdom*
[5] *Institute for Quantum Information, RWTH Aachen University, 52056 Aachen, Germany*

(Received 17 May 2023; revised 15 December 2023; accepted 4 March 2024; published 30 April 2024; corrected 17 June 2024)

We propose a class of randomized quantum algorithms for the task of sampling from matrix functions, without the use of quantum block encodings or any other coherent oracle access to the matrix elements. As such, our use of qubits is purely algorithmic and no additional qubits are required for quantum data structures. Our algorithms start from a classical data structure in which the matrix of interest is specified in the Pauli basis. For $N \times N$ Hermitian matrices, the space cost is $\log(N) + 1$ qubits and, depending on the structure of the matrices, the gate complexity can be comparable to state-of-the-art methods that use quantum data structures of up to size $O(N^2)$, when considering equivalent end-to-end problems. Within our framework, we present a quantum linear system solver that allows one to sample properties of the solution vector, as well as algorithms for sampling properties of ground states and Gibbs states of Hamiltonians. As a concrete application, we combine these subroutines to present a scheme for calculating Green's functions of quantum many-body systems.

DOI: 10.1103/PRXQuantum.5.020324

I. INTRODUCTION

A. Overview

As improvements in hardware increase the number and quality of qubits, we seek quantum algorithms that are able to showcase practical quantum advantage in the earliest possible time frame. Looking beyond noisy intermediate-scale quantum (NISQ) technologies [1–3], it is reasonable to assume that, given continued progress in quantum hardware, so-called fault-tolerant algorithms will have an important place in the gamut of quantum computing applications. Thus, it is pertinent to ask how soon such algorithms can be useful for real-life applications and how much can we accelerate this time line by constructing algorithms with lower and more flexible quantum resource costs.

Quantum algorithms for manipulating matrices have been proposed for many problems, including factoring, linear systems, ground-state energy estimation, simulation, and beyond (see, e.g., Refs. [4–8] and references therein). These algorithms are often phrased in terms of a quantum oracle model from which elements of the matrix of interest can be coherently accessed. Since the seminal proposals, there have been extensive improvements and refinements to the asymptotic run time for each of these algorithms, which is usually measured in the number of required queries to the oracle. For many problems, the state-of-the-art query complexities are optimal or close to optimal according to known complexity lower bounds. Moreover, many of these recent techniques are unified under the so-called quantum singular-value transformation (QSVT) framework [9,10], in which polynomial approximations are applied to the singular values of the desired matrix. Here, the oracle embeds the matrix in a larger unitary, commonly known as a block encoding.

Despite this promise, when considering end-to-end implementations of such quantum algorithms, two major hurdles can arise [11]. First, the implementation of the quantum oracles can require costly additional quantum resources, both in the depth required for each call and the number of qubits consumed. Second, the quantum algorithm can come with certain conditions or caveats that need to be satisfied for efficient applications [8].

As an example, consider the linear-systems problem, for which the pioneering Harrow, Hassidim, and Lloyd (HHL) algorithm has been proposed [12]. The current

*samsonwang@outlook.com

state-of-the-art quantum linear-systems solver (QLSS) presented in Ref. [13] uses $\mathcal{O}(\log(N))$ algorithmic qubits and a block-encoding oracle to process $N \times N$ matrices. This algorithm only needs to make a small number of calls to the block encoding if the matrix is well conditioned, compared to the run time of classical linear-systems algorithms. However, for general matrices an implementation of the block encoding using quantum random access memory (QRAM) [14] in depth $\mathcal{O}(\log(N))$ requires $\mathcal{O}(N^2)$ qubits [15–18]. Thus, for general matrices, the exponential savings in space resources are nullified. In order to circumvent this burden, one should search for specific classes of matrices with structure for which access is less costly. For instance, for matrices that are L-sparse in an efficiently implementable unitary basis, block encodings can be implemented with substantially less quantum cost than in the general case [19–22]. However, this still leads to an additional qubit overhead that we argue could be minimized further when considering early implementations of fault-tolerant algorithms. For matrices sparse in the computational basis, up to $\mathcal{O}(N)$ qubit overhead is still required [23], unless one seeks additional structure such that the matrix entries can be efficiently coherently computed. Second, it is important to consider the exact problem that the quantum algorithm solves: the QLSS returns a quantum state in which the solution vector is encoded with some nonzero additive error, unlike textbook classical solvers that provide the full classical vector exactly. Thus, in order to assess the utility of quantum linear-systems solvers, full end-to-end applications including possible additional subroutines need to be carefully analyzed [24]. For instance, it may be more fair to compare the quantum algorithm to randomized classical solvers that allow for some error [25] or so-called "dequantized" approaches that operate with a classical data structure analogous to QRAM [26,27]. We discuss various approaches to linear systems in more detail in Sec. III C.

1. Our contribution

In this work, we present a framework for constructing algorithms that sample properties of matrix functions that do not use quantum oracles to provide coherent access to the matrix in question (see Fig. 1). Despite having no qubit overhead to implement quantum oracles, the asymptotic complexities of our algorithms can remain comparable with those of other algorithms in the literature whenever the considered matrices have an amenable structure in the Pauli basis (and when considering equivalent end-to-end problems). Hence, for physically motivated matrices, potential quantum advantages originally requiring QRAM could possibly be similarly obtained in our approach without using a quantum data structure, making them more applicable for the early fault-tolerant regime [28–33]. Specifically, given a Fourier-series approximation to a function f and an $N \times N$ Hermitian matrix A with known decomposition in the Pauli basis, we give algorithms to sample properties of $f(A)$ using a total of $\log(N) + 1$ qubits. These properties take the form $\mathrm{Tr}[f(A)\rho f(A)^\dagger O]$ and $\langle\psi|f(A)|\phi\rangle$ for some quantum states ρ, $|\psi\rangle\langle\psi|$, and $|\phi\rangle\langle\phi|$ and some measurement observable O. Using this framework, we present algorithms for sampling properties of the solution vector in the linear-systems problem, as well as from ground states and Gibbs states of a given Hamiltonian. We provide direct comparisons of the complexities of our algorithms with other classical and quantum algorithms for specific end-to-end tasks and we present an application of our algorithms for computing Green's functions in many-body physics.

As our starting point, we take inspiration from algorithms for quantum chemistry [29,30,34], where quantum

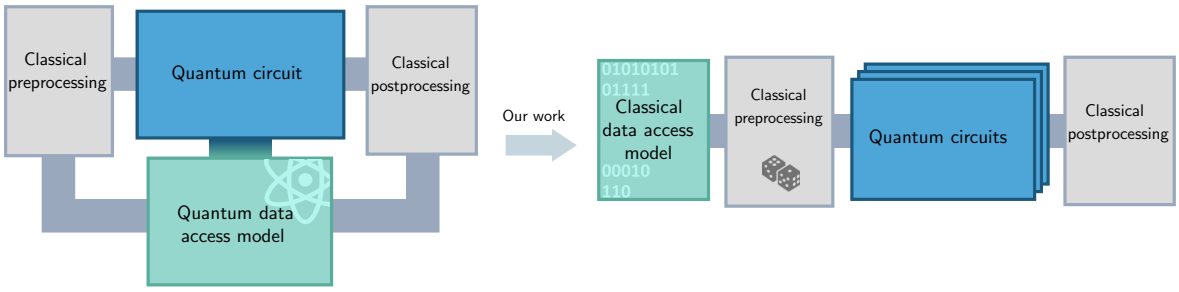

FIG. 1. The motivation of our work. We reduce quantum hardware requirements for quantum algorithms on classical data by removing the need for quantum data structures or quantum oracles. This is achieved by replacing coherent access to the data with a classical description of the data in the Pauli basis and utilizing a randomized algorithm that samples the outputs of many quantum circuits. These circuits are chosen independently and thus in theory can also be parallelized, trading reduced total run time for additional space cost in the form of many quantum processors. Our approach uses circuits with at most $\log N + 1$ qubits when processing data from $N \times N$ Hermitian matrices. This can be compared to other algorithms that utilize quantum data-access models that may have significantly greater qubit overhead overall.

data structures are often not needed. Instead of quantum (coherent) access to the matrix A, we ask for classical access to the coefficients $a_\ell \in \mathbb{R}$ in its decomposition in the Pauli basis,

$$A = \sum_{\ell=1}^{L} a_\ell P_\ell, \tag{1}$$

where the P_ℓ are multiqubit Pauli operators. We refer to this as the Pauli access model and note that it is a natural representation for data coming from physical problems, e.g., when the matrix comes from a Hamiltonian. The model also mathematically matches the physical intuition that running quantum subroutines for "quantumly structured data" has possible potential for quantum speed-ups. Indeed, our algorithms are faster for matrices with a small vector ℓ_1-norm of the Pauli coefficients $\lambda := \sum |a_\ell|$, which we refer to as the "Pauli weight." Moreover, we remark that there is no explicit dependence on the sparsity or the number of Pauli terms L (which we will call the "Pauli sparsity") in the quantum run times—both of which can be substantially larger than the Pauli weight and appear in the run time of other algorithms.

B. Related work

Applications for the early fault-tolerant era of quantum computing have recently begun to be explored, following the motivation to design algorithms that extract practical value out of fault-tolerant quantum algorithms as soon as possible [29–36]. In this spirit, algorithms have been designed to consume fewer quantum resources for Hamiltonian problems including phase estimation [29–32,36], ground-state preparation [31], and the computation of ground-state properties [33], by increasing the number of circuit samples required. Until now, these algorithms have predominantly aimed to reduce a proxy for the maximum circuit depth, in the form of the number of calls to a time-evolution oracle for a prescribed Hamiltonian in one coherent run of a circuit. It then remains to choose an appropriate time-evolution oracle for the intended setting, which can substantially affect the gate overhead or the number of qubits required. This is in contrast to our approach, where the first priority is to reduce qubit overhead. In Sec. IV B, we discuss further the implications for various choices of time-evolution oracle and how the resulting implementations compare to our results for the ground-state property-estimation problem. A key tool in the aforementioned algorithms is the use of *randomization*. Randomized approaches have also more generally found use in Hamiltonian simulation [34,37] and in simplifying quantum walk algorithms [38].

One distinctive approach is that of Ref. [30] for phase estimation, which uses the Pauli access model rather than a time-evolution oracle. In this work, an algorithm is proposed that randomly compiles the Heaviside function $H(A)$ via a quantity of the form $\mathrm{Tr}[H(A)\rho]$ by sampling from a Fourier-series approximation to the function. Reference [35] also presents an algorithm to perform randomized sampling of a given observable after a time evolution of a given Hamiltonian. We extend these ideas in our work to a more general class of properties corresponding to any function for which we have a Fourier approximation. The result of this approach is that an overhead of only one additional qubit is required to run the algorithm for Hermitian matrices. Moreover, as with the approach in Ref. [30], we sample from the outputs of many quantum circuits, rather than running one long coherent evolution.

We remark that near-term approaches for the quantum linear-systems problem have recently been proposed that use similar data-access assumptions to our Pauli access model [39–41]. Namely, these works assume that the matrix of interest has a known decomposition $A = \sum_t c_t U_t$, where the U_t are efficiently implementable unitaries (such as the Pauli-basis decomposition as in our work). These approaches use parametrized circuits the depth of which can be tuned to whatever a near-term implementation allows. However, despite showing promising numerical performance for small problem sizes, they lack generic run-time guarantees. On the other hand, our algorithms give prescriptive circuits with run-time guarantees.

C. Outline

The rest of the paper is structured as follows. In Sec. II, we present our general framework, including: our main result in Sec. II B; a note on the classical power of our access model in Sec. II C; and a discussion on sampling from other linear combinations of unitaries in Sec. II D. We then demonstrate applications of our framework: we present our algorithm for sampling properties of the solution vector in the linear-systems problem in Sec. III; we describe our algorithm for sampling properties of the ground state in Sec. IV; we discuss our algorithm for sampling properties of Gibbs states in Sec. V; and we show how these algorithms can be used together to estimate single-particle Green's functions in the context of many-body physics in Sec. VI. Finally, in Sec. VII we present our concluding discussions and outlook. In the appendixes, we present detailed analytical statements and all proofs thereof.

II. GENERAL APPROACH

A. Warm-up problem

We start by demonstrating how to extend the ideas from Refs. [30] and [35] to sample properties of a Fourier series of a given matrix. This gives intuition for the core routine that we will use for the rest of our results. Readers who

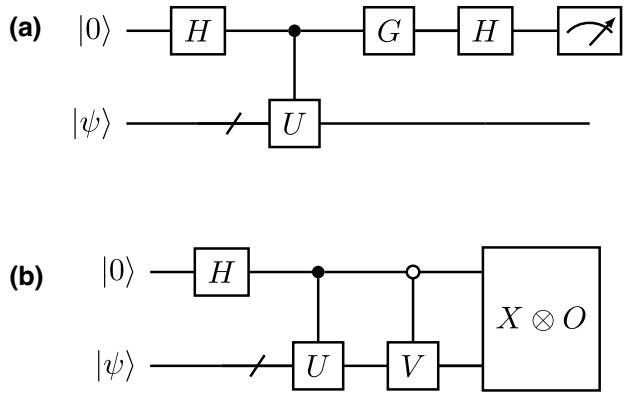

FIG. 2. The circuits used in our algorithms. We sample strings of quantum gates, consisting of specified Pauli operators and Pauli rotations, and perform quantum circuit runs with controlled versions of these gates. (a) The Hadamard test circuit. Measuring the expectation value of Z on the first register returns $\mathrm{Re}(\langle\psi|U|\psi\rangle)$ and $\mathrm{Im}(\langle\psi|U|\psi\rangle$ for choices of $G = \mathbb{1}$ and $G = S^\dagger := |0\rangle\langle 0| - i|1\rangle\langle 1|$, respectively. (b) The generalized Hadamard test circuit. Applying controlled-U and anticontrolled-V, followed by measurement of the observable $X \otimes O$ yields $\frac{1}{2}\left(\langle\psi|U^\dagger O V|\psi\rangle + \langle\psi|V^\dagger O U|\psi\rangle\right)$.

want a summary of the results may skip ahead to Sec. II B and the outlines of our specific algorithms in Secs. III, IV, V, and VI.

Our algorithms will make use of the Hadamard test circuit [see Fig. 2(a)], pioneered by Lin and Tong in Ref. [29] for use in the ground-state energy-estimation problem in the early fault-tolerant regime. We will also use the related circuit in Fig. 2(b), introduced by Childs and Wiebe in Ref. [42] to implement linear combinations of unitaries for Hamiltonian simulation. Our circuits will ask for elementary controlled unitary operations in the form of controlled Pauli gates and Pauli rotations.

Proposition 1 (Sampling from Fourier series). Suppose that we have a Fourier series

$$s(A) = \sum_{k\in F} \alpha_k \exp(it_k A), \tag{2}$$

$$\text{with } \ell_1\text{-norm of coefficients } \alpha := \sum_{k\in F} |\alpha_k|, \tag{3}$$

for an $N \times N$ Hermitian matrix A with known Pauli decomposition $A = \sum_\ell a_\ell P_\ell$ and Pauli weight $\lambda = \sum_\ell |a_\ell|$. Then, we find:

(a) Given a procedure to prepare the pure states $|\psi\rangle$, $|\phi\rangle$ with respective unitaries U_ψ, U_ϕ and respective gate depths d_ψ, d_ϕ, we have a randomized quantum algorithm that uses $\log(N) + 1$ qubits to

$$\text{approximate } \langle\phi|s(A)|\psi\rangle \text{ up to additive error } \varepsilon, \tag{4}$$

with arbitrary constant success probability, using

$$\mathcal{C}^{\phi}_{sample} = \mathcal{O}\left(\alpha^2/\varepsilon^2\right) \text{ circuit samples}, \tag{5}$$

where each circuit takes the form in Fig. 2(a) with

$$\mathcal{C}^{\phi}_{gate} = \mathcal{O}\left(\lambda^2 t^2_{\max} + d_\psi + d_\phi\right) \text{ gate depth}, \tag{6}$$

where we denote $t_{\max} := \max_{k\in F} t_k$.

(b) Given a procedure to prepare the quantum state ρ in depth d_ρ and perform measurements with measurement operator O, we have a randomized quantum algorithm that uses $\log(N) + 1$ qubits to

$$\text{approximate } \mathrm{Tr}[s(A)\rho s(A) O] \text{ up to additive error } \varepsilon, \tag{7}$$

with arbitrary constant success probability, using

$$\mathcal{C}^{O}_{sample} = \mathcal{O}\left(\|O\|^2\alpha^4/\varepsilon^2\right) \text{ circuit samples}, \tag{8}$$

where each circuit takes the form in Fig. 2(b) with

$$\mathcal{C}^{O}_{gate} = \mathcal{O}\left(\lambda^2 t^2_{\max} + d_\rho\right) \text{ gate depth}. \tag{9}$$

We see that various properties of the Fourier series determine the complexity of the quantum algorithm. Namely, the weight of the Fourier series α determines the sample complexity, while the time parameter $t_{\max}$ determines the gate complexity. The gate complexity also depends on the Pauli weight λ of the matrix A. In certain cases, the Pauli weight of a matrix can be much smaller than its dimension, despite there being many nonzero Pauli terms. In these cases, we expect the above algorithms to be efficient. Note that, similar to the linear combinations of unitaries (LCU) [42–45] and QSVT [9,10] frameworks for quantum algorithms, our framework enacts a general class of functions, which we will apply to different approximation problems in the rest of this paper. Unlike the aforementioned approaches, our framework does not need access to quantum oracles or any additional coherent resources.

We provide the proof of Proposition 1 in Appendix B 1. A key technical tool that we use is the *random compiler lemma* of Ref. [30, Lemma 2], which decomposes fractional time-evolution operators into a probabilistic mixture of Pauli matrices and Pauli rotations. This leads to a decomposition of the full time-evolution operator simply by taking the product of such fractional operators. Namely, for some time parameter $t \in \mathbb{R}$, for any choice $r \in \mathbb{N}$ and for Hermitian matrix A with known Pauli decomposition as in Eq. (1), Ref. [30] shows how one can obtain the

decomposition

$$e^{iAt/r} = \gamma \sum_{\ell \in T'} p_\ell \cdot u_\ell P_\ell \ \exp\left(i\theta_\ell P'_\ell\right), \tag{10}$$

for some index set T', where p_ℓ are probabilities, $u_\ell \in \{\pm 1, \pm i\}$ are phases, and P_ℓ, P'_ℓ are ($\lceil \log N \rceil$-qubit) Pauli operators, all of which implicitly depend on t/r. Moreover, the weight of the mixture can be shown to satisfy $\gamma \leqslant \exp(\lambda^2 t^2/r^2)$. This implies that one can express a full time-evolution operator as a linear combination

$$e^{iAt} = \left(e^{iAt/r}\right)^r = \sum_{m \in T} \beta_m^{(r)} U_m^{(r)}, \tag{11}$$

where the weights of the coefficients satisfy $\sum_{m \in T} |\beta_m^{(r)}| = \mathcal{O}(1)$ for $r = \Theta(\lambda^2 t^2)$, meaning that one can sample from the distribution with bounded variance. Each $U_m^{(r)}$ is a string of gates consisting of r pairs of controlled (multi-qubit) Pauli rotations and a series of controlled Pauli gates. Thus, r can be considered to control the quantum run time of the algorithm. Our algorithms sample from the strings of gates according to the linear combination in Eq. (11).

We make two brief remarks on compilation. First, we note that n-qubit Pauli-rotation gates can be compiled into a single single-qubit Pauli-rotation gate and $\mathcal{O}(n)$ Clifford gates. Thus, in our algorithmic framework, compilation of each layer of gates results in a single non-Clifford gate, i.e., the number of non-Clifford gates and the total gate depth go hand in hand. Second, any series of n-qubit Pauli gates can be classically compiled into a single n-qubit Pauli gate up to a phase in $\mathcal{O}(n)$ classical time. From hereon, we will generally refer to quantum gate depth while stressing that this is equivalent to the number of non-Clifford gates up to a logarithmic factor.

We provide pseudocode for the algorithm presented in Proposition 1 to prepare $\langle\phi|s(A)|\psi\rangle$ in Algorithm 1. The algorithm for $\mathrm{Tr}\left[s(A)\rho s(A)^\dagger O\right]$ is very similar and we present the pseudocode for this in full in Algorithm 2 in Appendix B 1. These two algorithms will form the core quantum routine for the rest of our results.

We remark that in the pseudocode in Algorithm 1, we have left the run-time vector $\vec{r}$ as a freely chosen algorithm parameter. In practice, one would likely choose each element of the run-time vector to be $r_k \propto \lambda^2 t_k^2$ for all k, which leads to the run-time guarantees specified in Proposition 1. However, we note that one can still freely choose a proportionality constant, which will trade a constant-factor improvement in the sample complexity for increased gate depth, or vice versa.

Finally, in Appendix B 2, we detail the classical overheads required to run our core algorithm in Algorithms 1 and 2. These amount to essentially linear prepossessing overhead in the number of Pauli terms L and the number of Fourier terms $|F|$. Each quantum sample comes with logarithmic classical overhead in problem parameters.

ALGORITHM 1. Fourier sampling of $\langle\phi|s(A)|\psi\rangle$.

Problem input:

- $N \times N$ Hermitian matrix $A = \sum_\ell a_\ell P_\ell$, with $a_\ell \in \mathbb{R}\ \forall \ell$, and known Pauli weight $\lambda = \sum_\ell |a_\ell|$
- Fourier parameters $\{\alpha_k\}_{k\in F}$, $\{t_k\}_{k\in F}$ for series $s(A)$ according to Eq. (2)
- Unitary gates U_ψ, U_ϕ which respectively prepare $|\psi\rangle, |\phi\rangle$.

Parameters: (1) Approximation error $\varepsilon \in (0,1)$;; (2) Success probability $1-\delta \in (0,1)$;; (3) Runtime vector $\vec{r} \in \mathbb{N}^{|F|}$.

Output: Approximation to $\langle\phi|s(A)|\psi\rangle$ with additive error at most ε and success probability at least $1-\delta$.

1. Compute coefficients $\{\beta_{km}^{(\vec{r})}\}_{(k,m)\in F\times T}$ as in Eq. (11) (see Appendix A) for each time parameter $\{t_k\}_{k\in F}$.
2. $R(\vec{r}) \leftarrow \sum_{(k,m)\in F\times T} \left|\alpha_k \beta_{km}^{(r_k)}\right|$.
3. $M(\vec{r}) \leftarrow \left\lceil 4\ln(2/\delta)\left(R(\vec{r})/\varepsilon\right)^2 \right\rceil$.
4. **for** $j \in \left[M(\vec{r})\right]$:
   (i) Sample indices (k', m') according to probability distribution $\{|\alpha_k \beta_{km}^{(r_k)}|/R(\vec{r})\}_{(k,m)\in F\times T}$ and select corresponding string of gates $U_{k'm'}^{(r_{k'})}$.
   (ii) Run $n+1$ qubit circuit in Figure 2(a) with input state $|0\rangle^{\otimes \lceil \log N\rceil + 1}$, $G = \mathbb{1}$, controlled $U_\phi U_{k'm'}^{(r_{k'})} U_\psi$ gates, and measurement operator Z on first register. Record measurement outcome o_j.
   (iii) Run same circuit again with $G = S^\dagger := |0\rangle\langle 0| - i|1\rangle\langle 1|$. Record measurement outcome o'_j.
   (iv) $z_j \leftarrow R(\vec{r})(o_j + i o'_j)$.
5. **end for**
6. $\bar{z}^{(M)} \leftarrow \sum_j z_j / M(\vec{r})$. **return** $\bar{z}$.

B. Main result

Using the results of Sec. II A, we can now demonstrate how to sample properties of matrix functions, starting from a sufficiently good Fourier-series approximation to the function, and a decomposition of the matrix in the Pauli basis.

Given a real-valued function $f : \mathbb{R} \to \mathbb{R}$, we consider a scenario in which we have a Fourier-series approximation $s(\varepsilon, D_A) : \mathbb{R} \to \mathbb{R}$ that is ε-close to f on the domain D_A.

More precisely, we suppose that

$$|f(x) - s(\varepsilon, D_A, x)| \leqslant \varepsilon\,, \quad \forall x \in D_A. \tag{12}$$

We note that the condition in Eq. (12) is dependent on the maximum deviation of the Fourier approximation over the entire domain of interest—this is the property that will determine the rigorous worst-case complexity of our algorithms—thus, it is important to find Fourier-series approximations that are accurate even on the extremities of a domain, not just on average. The Fourier series can always be expressed as

$$s(\varepsilon, D_A, x) = \sum_{k \in F_{\varepsilon, D_A}} \alpha_k(\varepsilon, D_A) \exp\left(i t_k(\varepsilon, D_A) x\right), \tag{13}$$

over some index set F_{ε,D_A}, where $\{\alpha_k(\varepsilon, D_A)\}_k$ and $\{t_k(\varepsilon, D_A)\}_k$ are Fourier parameters that, in general, have a dependence on the approximation error ε and the domain of approximation D_A. This setting can be simply translated to matrix functions, if one considers a Hermitian matrix A the spectrum of which lies in D_A. In this case, ε then corresponds to closeness of matrix functions in operator norm.

Theorem 1 (Generalized sampling from Fourier approximations). Suppose that we have a matrix function $f(A)$ of an $N \times N$ Hermitian matrix A that is approximated by a Fourier series $s(\tilde{\varepsilon}, A)$ with tunable error parameter $\tilde{\varepsilon}$ and has a known Pauli decomposition of A with Pauli weight λ. Suppose further that we have unitary oracles U_ψ, U_ϕ, and U_ρ to prepare $|\psi\rangle$, $|\phi\rangle$, and ρ, respectively. Then, we give explicit randomized algorithms to

$$\text{approximate } \frac{\langle\phi|f(A)|\psi\rangle}{q} \text{ and } \frac{\mathrm{Tr}\left[f(A)\rho f(A)^\dagger O\right]}{q^2} \tag{14}$$

to a given small enough additive error ε, using Algorithms 1 and 2, respectively, on $\log(N)+1$ qubits, where q is some arbitrary normalization factor.

We provide a more precise statement with exact complexities, and accompanying proofs accounting for constant-factor terms, in Appendix B 3. Theorem 1 provides a general recipe for sampling from functions of $N \times N$ Hermitian matrices by using $\log(N)+1$ qubits, given a Fourier-series approximation and a Pauli decomposition of the matrix. In particular, it specifies how to tune the Fourier-approximation parameter $\tilde{\varepsilon}$ such that we can directly use Algorithms 1 and 2. We stress that as with the warm-up problem, we do not use any hidden quantum oracles and specify circuits explicitly in terms of controlled Pauli gates and controlled Pauli rotations.

We note that a property of our algorithms in Theorem 1 is that the output is a number, rather than a quantum state. We envision in the majority of potential applications for quantum algorithms that the goal is to extract classical information out of a quantum state [8]. In many cases, this would be captured by the quantities in Eq. (14). If the application of a matrix function is to be used as a subroutine as part of a larger quantum computation, our framework allows further quantum processing by appending fixed controlled unitaries in the case of Algorithm 1 or simply appending fixed unitaries on the second register in Algorithm 2. Finally, some algorithmic frameworks ask instead to sample from a quantum state in the computational basis [46]. We highlight in the following remark that we can similarly statistically recover the output vector in the computational basis. A full exposition is provided in Appendix B 6.

Remark (Sampling from output vector). By modifying the measurement in Algorithm 2 to computational-basis measurements, we give an unbiased estimator for the vector the entries of which are (approximately) $\langle\vec{z}_n|f(A)|\psi\rangle$ for each $\vec{z}_n \in \{0,1\}^n$.

Our randomized scheme allows for generic normalization of the answer by some factor q, with all complexities accounted for. So far, we have assumed that q is exactly given. However, the desired q may in general not be exactly known. One salient example is if we wish for $f(A)|\psi\rangle$ to be a normalized quantum state for some input state $|\psi\rangle$. In this case, the relevant normalization quantity is $\|f(A)|\psi\rangle\|$. In Appendix B 4, we present a randomized subroutine to estimate quantities of this form, which can be directly integrated with our core result in Theorem 1. We summarize this with the following proposition.

Proposition 2 (Sampling normalization constant). Under the conditions of Theorem 1, the quantities

$$\frac{\langle\phi|f(A)|\psi\rangle}{\|f(A)|\psi\rangle\|} \text{ and } \frac{\mathrm{Tr}\left[f(A)|\psi\rangle\langle\psi|f(A)^\dagger O\right]}{\|f(A)|\psi\rangle\|^2} \tag{15}$$

can be approximated to a given desired additive error ε, with the addition of a subroutine on $\log(N)+1$ qubits to approximate $\|f(A)|\psi\rangle\|$.

Proposition 2 specifies how to approximate quantities in Eq. (15) by first running a subroutine to approximate $\|f(A)|\psi\rangle\|$ to a specified error, followed by the core algorithm using Theorem 1.

C. Classically easy functions

We briefly remark on the implications of our techniques for the power of randomized classical approaches using the Pauli access model. Namely, we show in the following proposition that sufficiently low-degree polynomials of

matrices (such as for matrix multiplication) can be sampled from, if these matrices have low Pauli weight.

Proposition 3 (Classical polynomial sampling). For stabilizer states $|s\rangle$ and $|t\rangle$, Hermitian matrix A with known Pauli decomposition and Pauli weight λ, and p_d, a d-degree polynomial with coefficients of magnitude $\mathcal{O}(1)$, we give a classical randomized scheme to

$$\text{approximate } \langle t|p_d(A)|s\rangle \text{ up to additive error } \varepsilon, \tag{16}$$

using

$$\mathcal{C}_{sample} = \widetilde{\mathcal{O}}\left(\lambda^{2d}/\varepsilon^2\right) \tag{17}$$

independent classical subroutines each with time and space complexity

$$\mathcal{C}_{time} = \widetilde{\mathcal{O}}\left(d\log^2(N)\right) \quad \text{and} \quad \mathcal{C}_{bits} = \widetilde{\mathcal{O}}\left(\log^2(N)\right), \tag{18}$$

respectively, where λ is the Pauli weight of A.

A more detailed statement, along with the proof thereof, can be found in Appendix B 7. Proposition 3 works by observing that the task in Eq. (16) can be obtained statistically via measurement outcomes of depth-$\mathcal{O}(d)$ Clifford circuits, which are efficiently simulable. We note that, as with our previous specified algorithms, each sample can be parallelized, thus moving $\mathcal{C}_{\text{sample}}$ into the space complexity. Proposition 3 also trivially extends to polynomials of multiple matrices. Thus, our result implies that classically sampling from certain primitives such as matrix multiplication can be efficient, if the matrices have low Pauli weight.

Proposition 3 implies similar efficiency for quantum or classical algorithms for functions with low-degree polynomial approximations. However, in general we do not expect such low-degree approximations to always exist. For instance, we can investigate some implications for the linear-systems problem and the ground-state property-estimation problem. Reference [47] gives a polynomial approximation to the inverse function with degree linear in the condition number. This implies a classical algorithm to sample an element of the solution vector of the linear-systems problem with exponential sample complexity in the condition number. Likewise, one could use the power-law method to approximately project to the ground state for Hamiltonians with negative spectra. However, this results in a classical algorithm with sampling complexity exponential in the inverse spectral gap (for more details, see Appendix C 2 a). It thus remains to see if there are problems of interest with low-degree polynomial approximations.

Similar results have also been shown for matrix powers for the sparse-access model [48,49], where the base of the exponential in Eq. (17) is different. Interestingly, the problem of evaluating powers to additive error $\mathcal{O}(\|A\|^d\varepsilon)$ (rather than $\mathcal{O}(\lambda^d\varepsilon)$) has been shown to be BQP-complete [50] and classically hard [51], again in the sparse-access model. The QSVT framework [9,10] can apply a more general class of polynomial transformations with cost scaling only linearly in the degree. We leave it as an open question as to whether there are more efficient randomized *early fault-tolerant* quantum algorithms for high-degree polynomials.

D. Sampling from other linear combinations of unitaries

In our main result, we present Monte Carlo sampling algorithms for a specific decomposition of functions into linear combinations of implementable unitaries—first, by decomposing the function into time-evolution operators via a Fourier decomposition and, second, by decomposing those time-evolution operators into Pauli gates and Pauli rotations. The resulting run-time complexities depend on the properties of the Fourier approximation chosen for the matrix function of interest, $f(A)$. One pertinent question is then: when searching for appropriate Fourier approximations, what sample complexity would we expect at best with these techniques?

Our algorithms sample from a weighted probability distribution of unitaries, with the weight directly factoring into the complexity. Assuming that we require an ε-close approximation on all states, this weight is lower bounded by $\|f(A)\|$, for any chosen Fourier decomposition. Thus, under our presented sampling framework, Hoeffding's inequality gives sufficient conditions for to approximate $\langle\phi|f(A)|\psi\rangle$ and $\mathrm{Tr}[f(A)\rho f(A)^\dagger O]$ using $\mathcal{O}\left(\|f(A)\|^2/\varepsilon^2\right)$ and $\mathcal{O}\left(\|f(A)\|^4/\varepsilon^2\right)$ respective samples at best. We remark that this argument holds when sampling from *any* decomposition of $f(A)$ into a linear combination of unitaries, not just the one we consider specifically in this work. As we shall see in the following applications sections, we achieve this for the linear-systems problem (Sec. III) and the ground-state property-estimation problem (Sec. IV) up to logarithmic factors.

We stress that this discussion does not constitute generic sample-complexity lower bounds (e.g., one could in practice obtain the desired result using fewer samples than Hoeffding's inequality specifies). Additionally, we do not rule out that better sample-complexity guarantees can be obtained for other randomized schemes beyond Monte Carlo here. Finally, one can achieve problem-dependent improvements to the sample complexity by approximating $f(A)$ on the relevant subspace for the problem at hand. This is how a better sample-complexity guarantee is achieved for our Gibbs-state algorithm in Sec. V.

III. LINEAR SYSTEMS

A. Randomized quantum linear system solver

In this section, we show how to apply Theorem 1 to sample from the inverse of a matrix, with applications for linear-systems problems. We use the Fourier-series approximation for the inverse function found in Ref. [47]. With this, we establish the following result [52].

Corollary 1 (Linear systems). Consider a Hermitian matrix A with known Pauli decomposition as in Eq. (1) with Pauli weight λ. Denote q as a freely chosen normalization parameter. Finally, suppose that we have ability to prepare state $|\vec{b}\rangle$ in $\mathcal{O}(d_{\vec{b}})$ depth. Then, we find:

(a) Given the ability to implement $|\psi\rangle$ via unitary U_ψ in gate depth d_ψ, we have a randomized quantum algorithm to

$$\text{approximate } \frac{\langle\psi|A^{-1}|\vec{b}\rangle}{q} \text{ up to additive error } \varepsilon, \tag{19}$$

with arbitrary constant success probability, utilizing $\mathcal{C}^{\psi}_{sample}$ quantum circuits of the form in Fig. 2(a), each consisting of $\mathcal{C}^{\psi}_{gate}$ layers of gates, where

$$\mathcal{C}^{\psi}_{sample} = \widetilde{\mathcal{O}}\left(\frac{\|A^{-1}\|^2}{\varepsilon^2 q^2}\right),$$
$$\mathcal{C}^{\psi}_{gate} = \widetilde{\mathcal{O}}\left(\|A^{-1}\|^2\lambda^2 + d_\psi + d_{\vec{b}}\right). \tag{20}$$

(b) Given the ability to measure observable O; $\|O\| \leqslant 1$, we have a randomized quantum algorithm to

$$\text{approximate } \frac{\langle\vec{b}|A^{-1}OA^{-1}|\vec{b}\rangle}{q^2} \text{ up to additive error } \varepsilon, \tag{21}$$

with arbitrary constant probability, utilizing $\mathcal{C}^{O}_{sample}$ quantum circuits of the form in Fig. 2(b), each consisting of $\mathcal{C}^{O}_{gate}$ layers of gates and one measurement of O, where

$$\mathcal{C}^{O}_{sample} = \widetilde{\mathcal{O}}\left(\frac{\|A^{-1}\|^4}{\varepsilon^2 q^4}\right),$$
$$\mathcal{C}^{O}_{gate} = \widetilde{\mathcal{O}}\left(\|A^{-1}\|^2\lambda^2 + d_{\vec{b}}\right). \tag{22}$$

We detail a proof of this corollary in Appendix C 1, where we specify polylogarithmic contributions and account for constant factors. Corollary 1 provides an oracle-free quantum linear-systems algorithm that only uses one ancillary qubit for Hermitian matrices. We note that we can consider non-Hermitian matrices for linear systems by embedding them in a larger Hermitian matrix via an additional qubit (details are provided in Appendix C 1).

Remark (Non-Hermitian matrices). We can extend Corollary 1 to include non-Hermitian matrices by embedding the matrix in a larger Hermitian matrix with the aid of a single additional qubit. The algorithm then uses $\log(N) + 2$ qubits with a factor-of-2 increase in sample complexity.

Further, we note that Proposition 2 allows the evaluation of the norm of the output vector $\|A^{-1}|\vec{b}\rangle\|$ via a subroutine with complexities equivalent to Eq. (22) up to logarithmic terms.

B. Statistical encoding of input vector

So far, we have supposed that $|\vec{b}\rangle$ is provided via a quantum oracle. With the following proposition, we demonstrate an approach to statistically encode the vector at a cost of classical preprocessing overhead and minimal gate overhead.

Proposition 4 (Statistical encoding of input vectors). Given a classical vector $\vec{b}$ with sparsity s, we give a scheme to replace the quantum oracle for the input state $|\vec{b}\rangle$ in Theorem 1 and Corollary 1. This has overhead:

(i) $\mathcal{O}(s)$ time for classical preprocessing.
(ii) sample-complexity-factor increase for problem (a) of $\mathcal{O}(\|\vec{b}\|_1^2/\|\vec{b}\|_2^2)$ and for problem (b) of $\mathcal{O}(\|\vec{b}\|_1^4/\|\vec{b}\|_2^4)$.
(iii) one or two single-qubit controlled $\lceil\log N\rceil$-qubit Pauli gates each circuit sample.

More details are provided in Appendix B 5. A key distinction is that this approach can return an encoding of $\vec{b}$ that represents the true magnitude of the vector rather than a normalized quantum state that a standard quantum oracle would provide, simply by increasing the sample complexity. Moreover, this is a general scheme for our randomized approach that is applicable beyond the linear-systems setting. We remark that our representation for the quantum state $|\vec{b}\rangle$ is similar to the so-called classical combination of variational quantum states considered in the near-term quantum linear-systems approach of Ref. [41]. Finally, we stress that in this scheme, we never explicitly prepare the quantum state. A comparison with explicit state preparation schemes [53] shows that this avoids $\mathcal{O}(Ns)$ cost in qubit overhead and trades it for classical preprocessing overhead and sample overhead.

C. Comparison

1. Quantum linear-systems solvers

The quantum linear-systems solver (QLSS) is an algorithm that takes an encoding of a matrix A and vector $\vec{b}$ and returns the quantum state proportional to $A^{-1}|\vec{b}\rangle$, where $|\vec{b}\rangle := \sum_i b_i|i\rangle / \|\vec{b}\|$. Since the proposal of the original QLSS, known as the Harrow, Hassidim, and Lloyd (HHL) algorithm [12], many improved schemes have been proposed [9,10,13,47,54–57] utilizing $\mathcal{O}(\log(N))$ algorithmic qubits. Early algorithms made use of a time-evolution oracle in the matrix of interest [47,54,57,58]. More recently, approaches that use a block encoding have been proposed [9,10,13].

We remark that, in contrast to other fault-tolerant approaches to the QLSS, our algorithms do not require coherent access to the matrix of interest A and thus there are no additional hidden dimension dependencies. Thus, for a general $N \times N$ Hermitian matrix, only $\log(N) + 1$ logical qubits are required to carry out the full algorithm, with a run time that does not explicitly depend on the dimension. We further note that our algorithm has the following additional distinctive features:

(i) In contrast to the QLSS, the error parameter ε in Corollary 1 is specified in terms of measurement errors in extracting information out of states of the form $A^{-1}|\vec{b}\rangle$. Thus, the QLSS is distinct from the problem that our algorithm solves, as our algorithm returns a number rather than a state. If one wishes to extract out classical information from the QLSS, one requires $\mathcal{O}(\varepsilon^{-2})$ circuit samples with incoherent measurements or an additional factor of $\mathcal{O}(\varepsilon^{-1})$ run time with coherent approaches [59,60].

(ii) Our complexities are determined by the operator norm $\|A^{-1}\|$, instead of the condition numbers $\kappa := \|A^{-1}\|\|A\|$ or $\kappa_F := \|A^{-1}\|\|A\|_F$. Moreover, there is no sparsity or explicit dimension dependence in the run time. Instead, the run time depends on the Pauli weight λ, which for certain problems can be much smaller than the sparsity or dimension.

(iii) We allow for arbitrary normalization by any normalization factor q, which rescales the asymptotic complexities by $\|A^{-1}\| \to 1/q\|A^{-1}\|$. We recall that in the QLSS, the solution is given as a normalized quantum state, which may be undesired. We discuss the role of normalization further in Sec. III C 2.

2. Complexity comparison

In Table I, we compare the complexities of our linear-systems algorithm with other classical and quantum algorithms in the literature for a specific task [61]. The task that we consider in Table I is to

$$\text{approximate the element } \frac{1}{c}\cdot\left(A^{-1}\vec{b}\right)_i \text{ to additive error } \varepsilon \text{ for some choice of } i \in [N] \text{ and normalization } c \in \mathbb{R}. \tag{23}$$

We assume that the normalization constant c is whatever is natural to the algorithm at hand and that we are querying the vector $A^{-1}\vec{b}$ in the computational basis. Thus, $c = 1$ for in the usual classical setting, and $c = \|A^{-1}\vec{b}\|$ for quantum solvers and quantum inspired classical solvers. This is to allow a "middle-ground" comparison with the other quantum and classical solvers, as relaxing these two assumptions incurs additional overhead for other quantum algorithms and classical algorithms, respectively.

We compare the space resources required to create the data access to A and to perform the algorithm. We also compare the maximum coherent quantum run time required (if applicable) and the total run time. For our randomized scheme, we assume no parallelization; thus the total run time is simply the product of the gate depth and the sample complexity. For all algorithms, we assume that the data starts in the most amenable *classical* format for that algorithm, such as in a sparse row representation, or in our case the classical Pauli access model. From there, we keep track of any additional *classical* or *quantum* overheads (both space and time) needed to provide any required data structures, e.g., the resources required to provide a block encoding.

Taking into account the aforementioned features, in Table I we consider the following algorithms:

(i) The Gaussian elimination "textbook" classical method, which returns the full exact solution vector. This has a run time that is dependent on the matrix multiplication exponent $\omega < 2.372$.

(ii) A randomized classical Kaczmarz method [25], which returns an ε-approximation to the full solution vector.

(iii) A "quantum inspired" approach [27]. This starts from a classical data structure intended to mimic QRAM, which allows sampling from probability distributions with probabilities proportional to the magnitude of elements in a given row of A. The algorithm returns a classical data structure that allows one to sample from individual elements of an ε-approximation to the solution vector. In order to estimate the norm, one needs to pay an extra $1/\varepsilon^2$ factor [27,63].

(iv) The HHL "textbook" quantum algorithm [12], which requires a time-evolution oracle. In Table I, we assume that this oracle is provided in a run-time-efficient manner via a block encoding [62]. We assume a Frobenius-norm block encoding, with

TABLE 1. The end-to-end complexities for the task in Eq. (23) of querying one element of the solution vector in the linear-systems problem to additive error ε. We consider a general $N \times N$ complex coefficient matrix A, with row sparsity s, which has Pauli decomposition [as in Eq. (1)] into up to N^2 Pauli terms with weight λ. We denote $\kappa_F := \|A^{-1}\|\|A\|_F$, where $\|\cdot\|$ denotes the operator norm and $\|\cdot\|_F$ denotes the Frobenius matrix norm, and note that $\kappa_F \geqslant \sqrt{\mathrm{rk}(A)}$. We also denote $\kappa := \|A^{-1}\|\|A\|$. Under "Data-access space," we have listed the dominant classical or quantum resources needed in order to represent A in the required format for the algorithm at hand. We have marked whether a given scheme requires approaches based on amplitude estimation (AE) to estimate the target quantity. Finally, each algorithm returns a different normalization c according to Eq. (23), which can factor into the complexity of each algorithm; we indicate this normalization in the final column. One can also consider different schemes that may be advantageous for Pauli-sparse matrices or matrices that are sparse in the computational basis; we discuss this later in this section and in Table II.

Algorithm	Access model	Data-access space	Algorithmic space	AE?	Quantum gate depth	Total run time	Norm
Textbook classical	Computational-basis elements	$\mathcal{O}(N^2)$ bits	$\mathcal{O}(N^2)$ bits	⋯	⋯	$\mathcal{O}(N^\omega)$	1
Randomized classical [25]	Sparse row	$\mathcal{O}(Ns)$ bits	$\mathcal{O}\left(s\kappa_F^2 \log(1/\varepsilon)\right)$ bits	⋯	⋯	$\mathcal{O}\left(s{\kappa_F}^2 \log(1/\varepsilon)\right)$	1
Quantum inspired [27]	Sampling and query	$\mathcal{O}(N^2)$ bits	$\mathcal{O}\left(\kappa_F^2 \log(1/\varepsilon)\right)$ bits	⋯	⋯	$\widetilde{\mathcal{O}}\left(\kappa^2\kappa_F^6/\varepsilon^2\right)$	$\|A^{-1}\vec{b}\|$
Textbook quantum [12]	Block-encoding time evolution [17,62]	$\mathcal{O}(N^2)$ qubits	$\mathcal{O}(\log N)$ qubits	Yes	$\widetilde{\mathcal{O}}\left(\kappa\kappa_F \log(N)/\varepsilon^2\right)$	$\widetilde{\mathcal{O}}\left(\kappa\kappa_F \log(N)/\varepsilon^2\right)$	$\|A^{-1}\vec{b}\|$
				No	$\widetilde{\mathcal{O}}\left(\kappa\kappa_F \log(N)/\varepsilon\right)$	$\widetilde{\mathcal{O}}\left(\kappa\kappa_F \log(N)/\varepsilon^3\right)$	$\|A^{-1}\vec{b}\|$
Quantum query-optimal [13]	Frobenius-norm block encoding [17]	$\mathcal{O}(N^2)$ qubits	$\log N + 6$ qubits	Yes	$\widetilde{\mathcal{O}}\left(\kappa_F \log(N)/\varepsilon\right)$	$\widetilde{\mathcal{O}}\left(\kappa_F \log(N)/\varepsilon\right)$	$\|A^{-1}\vec{b}\|$
				No	$\widetilde{\mathcal{O}}\left(\kappa_F \log(N)\log(1/\varepsilon)\right)$	$\widetilde{\mathcal{O}}\left(\kappa_F \log(N)/\varepsilon^2\right)$	$\|A^{-1}\vec{b}\|$
Randomized quantum (this work)	Pauli coefficients	$\mathcal{O}(N^2)$ bits	$\log N + 2$ qubits	No	$\widetilde{\mathcal{O}}\left(\|A^{-1}\|^2\lambda^2 \log^2(1/\varepsilon)\right)$	$\widetilde{\mathcal{O}}\left(\|A^{-1}\|^4\lambda^2/\varepsilon^2\right)$	$\|\vec{b}\|$

explicit construction via a QRAM in minimal depth as detailed in Ref. [17]. We note there are also other approaches that are more space efficient in quantum resources but more costly in quantum run time. Further, there is a choice of how to extract the classical information from the output of the HHL algorithm. We include two rows in Table I, which quantifies resources required for performing coherent approaches for extracting expectation values [59,60] (first row) and standard incoherent measurements assuming no parallelization, which have worse overall run time but use substantially less gate depth (second row).

(v) A state-of-the-art quantum linear-systems solver that achieves optimal query complexity [13], using a block-encoding access model. As with HHL we have assumed a Frobenius-norm block encoding implemented in low depth via a QRAM and we detail complexities both for incoherent measurements (first row) and for coherent approaches [59, 60] (second row).

(vi) Our algorithm using Corollary 1, assuming no parallelization. We remark that a particular artifact of the simplified task in Eq. (23) is that the run time of our algorithm can be reduced by an arbitrary factor of q^2 as in Theorem 1. However, we do not include this additional freedom in Table I, as we do not expect it to arise in more practical tasks. For instance, if one wished to compare two elements of $A^{-1}\vec{b}$ by estimating their ratio, q would not appear in the complexity.

The goal of our work is to reduce the quantum resources required for quantum approaches. This can be quantified by the three columns highlighted in Table I: "Data-access space," "Algorithmic space," and "Quantum gate depth." Our algorithm is clearly more efficient in space requirements compared to other quantum algorithms and, in particular, it uses no qubits for data access. We further note that if the Pauli weight λ is small, there is scope that our algorithm has competitive gate depth with the best alternative option, which is to use the adiabatic quantum algorithm of Ref. [13] with incoherent measurements (we remark that there exist matrices for which λ is much smaller than L, which in general can grow with the dimension of the system). If a coherent approach is used to extract the vector element, then our algorithm has exponentially better gate depth in the error parameter but worse overall run time.

The role of normalization. We note that in some practical tasks, one may care about the true magnitude of vector elements. In this case, the dependence of other quantum algorithms on ε^{-1} should be scaled up by a factor $\|A^{-1}|b\rangle\|$ in order to match our randomized algorithm or our dependence on ε^{-1} should be scaled down by a factor of $\|A^{-1}|b\rangle\|$ in order to match other quantum approaches. In addition, if the normalization is not given, computing it may incur significant overhead, which should be accounted for. When starting in the block-encoding model, Chakraborty *et al.* demonstrate how to obtain the state normalization to multiplicative error in $\widetilde{\mathcal{O}}(\kappa\mu\,\mathrm{polylog}(N)/\varepsilon)$ queries to the block encoding, where μ is the block-encoding subnormalization [64]. When searching for potential applications for quantum algorithms, the desired normalization for the problem of interest is an important consideration when benchmarking different algorithms.

Criteria for quantum advantage. For our algorithm to potentially be useful, it must also compete with classical algorithms. We remark that all the classical run times stated have at least linear-dimension dependence for generic matrices, as $\kappa_F \geqslant \sqrt{\mathrm{rk}(A)}$, where we denote $\mathrm{rk}(A)$ as the rank of A. In theory, a block encoding with a subnormalization of $\|A\|$ rather than $\|A\|_F$ may be possible, but at present an explicit construction for generic matrices is not known to the best of the authors' knowledge. For any of the quantum algorithms to display superpolynomial advantage for generic matrices, one requires at minimum $\kappa = \|A^{-1}\|\|A\| = \mathcal{O}(\mathrm{polylog}(N))$. Block encodings with stronger subnormalizations also need to be found (for generic matrices), as the scope for superpolynomial advantage is constrained to high-rank matrices. Further, one must consider a setting where the significant difference in error dependence is also accounted for. While our algorithm has significantly worse overall run time in $\|A^{-1}\|$, we note that the requirement for advantage in our algorithm is similar ($\|A^{-1}\| = \mathcal{O}(\mathrm{polylog}(N))$), although for practical advantages the degree of this polynomial may need to be kept small. Moreover, there is no rank condition yet to be overcome here. Thus, we expect our algorithm to potentially show quantum advantage if there is a setting in which other quantum algorithms also show advantage, given that the Pauli weight λ is small. We remark that it is possible for matrices to have a small λ but large computational-basis sparsity s, Pauli sparsity L, or rank $\mathrm{rk}(A)$.

Two known domains in which block encodings with subnormalizations that do not lead to rank dependence are sparse matrices in the computational basis and the Pauli basis. As discussed above, this is a requirement for any hope of superpolynomial quantum advantage. Our algorithmic construction can be thought of as extending the scope of the second category to also include nonsparse matrices in the Pauli basis that have a low Pauli weight λ. We now compare our algorithm to other approaches in these two settings in further detail.

Pauli-sparse matrices. So far, we have compared our algorithm against other approaches presuming that the Pauli sparsity L is large (note that $L = \mathcal{O}(N^2)$ in

general). If L is small and the Pauli decomposition is known, then the SELECT + PREPARE oracles (previously studied in algorithms for quantum chemistry [19–22]) can implement a more efficient block encoding than the general construction with a QRAM considered in Table I. We present the complexities for this block encoding in Table II. The qualitative conclusions are similar; the block-encoded quantum algorithm requires more qubits and our algorithm can have comparable gate depth if the Pauli weight λ is small compared to the Pauli sparsity L.

Matrices sparse in the computational basis. For the quantum query-optimal algorithm, one can also consider the quantum sparse-access model [65], which can be efficiently converted into a block-encoding model using $\mathcal{O}(\text{polylog}(N/\varepsilon))$ qubits and $\mathcal{O}(\text{polylog}(N/\varepsilon))$ elementary gates [9]. Thus, if this access model is naturally available, a significant space and time saving can be made. However, this access will typically arise because of inherent structure in the matrix, which enables the computation of entries, given their indices. If this structure is not present and we just have a generic sparse matrix, we still require QRAM for the most efficient block encoding of the matrix. In this case, implementation of the quantum sparse-access structure in minimal depth implementation uses $\widetilde{\mathcal{O}}(Ns)$ qubits and $\mathcal{O}(\log N)$ overhead [23], where s is the minimization over the row sparsity and column sparsity. Thus, the quantum space complexities could be up to quadratically better for sparse matrices than what is quoted in Table I but it still is linear in N. The corresponding run time is $\widetilde{\mathcal{O}}(s\kappa\,\log(N)/\varepsilon)$ using coherent approaches for expectation value estimation [59,60]. We remark that for specific classes of matrices there may also be more efficient ways to directly enact the sparse data-access model [66] but we leave the comparison for such special cases to be beyond the scope of this work.

Finally, we remark that in Tables I and II we presume that the relevant data structure for the input vector $\vec{b}$ is efficiently provided. In general, providing the input state $|\vec{b}\rangle$ also requires a QRAM. We recall that with our randomized scheme, we can circumvent this quantum resources cost by providing additional classical preprocessing overhead (see Proposition 4).

In this section, we have compared resource costs only for one specific task. For other applications, various classical and quantum overheads need to be carefully considered, which can vastly change the complexities. One end-to-end problem in which one does not expect to achieve superpolynomial quantum advantage is where one needs to read off the entire solution vector. In this case, the additional $\Omega(N)$ tomographic overhead can be compared to the polynomial scaling of the classical run times in $\widetilde{\kappa}(A) \geqslant \sqrt{\text{rk}(A)}$. We remark that in this setting, one would only hope to achieve approximately at most a quadratic speed-up in total run time to the best randomized classical solver in dimension dependence, ignoring dependencies on the approximation error ε.

IV. GROUND-STATE SAMPLING

A. Randomized quantum algorithm

In this section, we consider the problem of sampling properties of the ground state of a given Hamiltonian. In order to approximately project to the ground state, we use the Gaussian function $e^{-(1/2)\tau^2 x^2}$ for Hamiltonians with positive spectra, which has been proposed in previous works [67,68]. A Fourier-series representation has been found for this function in Ref. [68], which we use to establish the following corollary of our main result in Theorem 1.

Corollary 2 (Ground-state property estimation). Consider a Hamiltonian $H = \sum_l E_l |E_l\rangle\langle E_l|$ with all eigenvalues $E_l \geqslant 0$ and known Pauli decomposition $H = \sum_{\ell=1}^{L} a_\ell P_\ell$; $\lambda = \sum_\ell |a_\ell|$. Assume that the spectral gap of H is lower bounded by $\Delta \leqslant E_1 - E_0$. Additionally, we suppose that we have an initial trial state $|\psi_0\rangle\langle\psi_0|$ with overlap with the ground state $\gamma := |\langle\psi_0|E_0\rangle|$. Finally, we assume that $E_0 \leqslant \Delta/\sqrt{2\log\frac{\|O\|}{\varepsilon\gamma}}$ for some error parameter ε. Then, given the ability to measure observable O, we give a randomized algorithm to

$$\text{approximate } \langle E_0|O|E_0\rangle \text{ up to additive error } \varepsilon, \tag{24}$$

with arbitrary constant success probability, using

$$\mathcal{C}^{O}_{sample} = \widetilde{\mathcal{O}}\left(\frac{\|O\|^2}{\varepsilon^2\gamma^4}\right) \text{ circuit runs}, \tag{25}$$

each of the form in Fig. 2(b), each using one instance of $|\psi_0\rangle$ with

$$\mathcal{C}^{O}_{gate} = \widetilde{\mathcal{O}}\left(\frac{\lambda^2}{\Delta^2}\right) \text{ gate depth}. \tag{26}$$

Moreover, the assumption of the positivity of the spectrum can be dropped in place of a lower bound on the ground-state energy, with no change to the computational complexity of the algorithm.

We provide a proof of this corollary, including the extension to nonpositive spectra and accounting of polylogarithmic or constant-factor contributions, in Appendix C 2. Similar to our algorithms for linear systems, we emphasize that we do not ask for any quantum oracle or quantum subroutines encoding information about H. Additionally, if the trial state $|\psi_0\rangle$ is accessible as a classical description of vector amplitudes, or from a linear combination of gates, then the trial state can be prepared statistically

TABLE II. The end-to-end complexities for the task in Eq. (23) of querying one element of the solution vector in the linear-systems problem to additive error ε. Here, for all algorithms, we have assumed that we start from a Pauli description of the matrix of interest A, as in Eq. (1), where the number of terms L is small (e.g., a low-degree polynomial in $\log N$). In this case, it can be advantageous to use a SELECT + PREPARE block encoding for the query-optimal quantum algorithm. We compare complexities where the quantity in Eq. (23) is obtained using amplitude estimation- (AE) based approaches as well as incoherent sampling. In the final column, we indicate the implicit normalization c of each approach [see Eq. (23)].

Algorithm	Access model	Data-access space	Algorithmic space	AE?	Quantum gate depth	Run time	Norm
Quantum query-optimal [13]	SELECT + PREPARE [19–22]	$\mathcal{O}(L)$ bits, $\mathcal{O}(\log L)$ qubits	$\log N + 6$ qubits	Yes	$\widetilde{\mathcal{O}}\left(\lVert A^{-1}\rVert \lambda L/\varepsilon\right)$	$\widetilde{\mathcal{O}}\left(\lVert A^{-1}\rVert \lambda L/\varepsilon\right)$	$\lVert A^{-1}\vec{b}\rVert$
				No	$\widetilde{\mathcal{O}}\left(\lVert A^{-1}\rVert \lambda L \log(1/\varepsilon)\right)$	$\widetilde{\mathcal{O}}\left(\lVert A^{-1}\rVert \lambda L/\varepsilon^2\right)$	$\lVert A^{-1}\vec{b}\rVert$
Randomized quantum (this work)	Pauli coefficients	$\mathcal{O}(L)$ bits	$\log N + 2$ qubits	No	$\widetilde{\mathcal{O}}\left(\lVert A^{-1}\rVert^2 \lambda^2 \log^2(1/\varepsilon)\right)$	$\widetilde{\mathcal{O}}\left(\lVert A^{-1}\rVert^4 \lambda^2/\varepsilon^2\right)$	$\lVert \vec{b}\rVert$

as part of the randomized algorithm, with minimal depth requirements (see Proposition 4 and Appendix B 5). Our algorithm is particularly efficient when the Pauli weight λ is small, which can be the case for certain physically motivated problems. We discuss more detailed analysis of the complexities in Sec. IV B. Finally, we remark that if a trial state is instead provided with bounded positive overlap $\bar{\gamma} := \langle \psi_0 | E_0 \rangle$, then quantities of the form $\langle \phi | E_0 \rangle$ (given preparation of $|\phi\rangle\langle\phi|$) can be approximated with a reduced $\bar{\gamma}^{-2}$ dependence in the sample complexity.

B. Comparison

Different approaches to quantum algorithms for ground-state preparation have previously been studied in Refs. [31,33,67–72]. Recently, Ref. [33] has also established an explicit algorithm for ground-state property estimation. In Table III, we compare the complexities of various algorithms for the task of estimating a given observable with respect to the ground state. Namely, we compare the following approaches:

(i) A linear combination of unitaries (LCU) ground-state preparation algorithm [68], which requires a time-evolution oracle. There are many approaches to time evolution in the literature [19,34,37,44,45,62,73,74]. In Table III, we detail the complexities for two schemes that are in some sense the two ends of the spectrum: for first-order Trotter time evolution, which has no additional quantum space requirements, as well as the quantum run-time-efficient block-encoding approach [62].

(ii) The block-encoding approaches of Ref. [69] to prepare the ground state, which is query optimal in $\gamma, \Delta, \varepsilon$ up to logarithmic factors. We note that Ref. [69] gives two approaches; one in which there is an *a priori* bound for the ground-state energy and spectral gap and one in which there is no such assumption, which uses more ancillary qubits and has equivalent run time up to dominant order.

(iii) The early fault-tolerant approach of Ref. [33] to sample properties of the ground state, which again requires a time-evolution oracle. As a caveat, this approach can only estimate expectation values of unitary observables with respect to the desired state. Nonunitary observables can be taken into account with a block encoding, at the expense of more qubits and complexities augmented by the block-encoding factor. For the sake of easy comparison, in Table III we assume that the observable is unitary.

(iv) Our approach using Corollary 2, assuming no parallelization.

For the ground-state preparation algorithms considered in Table III, we assume that the observable is measured incoherently, in order to establish the most competitive gate depth with regard to our approach. We note that as in our discussion in Sec. III C, coherent estimation of observables yields a faster overall run time (linear scaling in $\mathcal{O}(\varepsilon^{-1})$), at the expense of $\mathcal{O}(\varepsilon^{-1})$ scaling in the gate depth, which can be exponentially worse [59,60]. For all algorithms, we assume that the Hamiltonian is provided classically in the form of its Pauli coefficients [as in Eq. (1)] and any subsequent resources required for data access are noted. If the total number of Pauli terms L is small, this allows for an efficient block encoding via the SELECT and PREPARE oracles, which is what we consider in Table III. We also consider Trotterized time evolution, which gives additional dependencies on commutators between terms [73], which here we bound with the magnitude of the largest Pauli coefficient $\Lambda := \max_\ell a_\ell$ for simplicity. One can also opt for higher-order Trotterized evolution, which gives improved scaling in ε^{-1} but worst scaling in L and λt.

TABLE III. An end-to-end comparison of the complexities of approximating the expectation value of an observable with respect to the ground state, up to ε-additive error. These algorithms presume access to a trial state approximating the ground state with overlap γ, a spectral-gap lower bound Δ, and possible further assumptions on the ground state energy (GSE) and spectral gap: (B1) given $\mu_1 \leqslant E_0$ such that $E_0 - \mu_1 \leqslant \Delta/\sqrt{\log \frac{1}{\gamma\varepsilon}}$; and (B2) given $\mu_2 \geqslant E_0$ such that $\mu_2 - \Delta/2 \geqslant E_0$, $\mu_2 + \Delta/2 \leqslant E_1$. In all cases, we assume that the Hamiltonian is given in terms of its Pauli decomposition (see Eq. (1)) with L Pauli terms, largest coefficient magnitude Λ, and coefficient weight λ.

	Quantum oracle	GSE and gap assumptions	Data-access space	Ancillary algorithmic space	Quantum gate depth	Run time
LCU [68]	First-order Trotter time evolution	(B1)	$\mathcal{O}(L)$ bits	$\mathcal{O}\left(\log\left(\frac{1}{\Delta}\log\frac{1}{\gamma\varepsilon^2}\right)\right)$ qubits	$\widetilde{\mathcal{O}}\left(\frac{L^3\Lambda^2}{\gamma\Delta^2}\frac{1}{\varepsilon}\right)$	$\widetilde{\mathcal{O}}\left(\frac{L^3\Lambda^2}{\gamma\Delta^2}\frac{1}{\varepsilon^3}\right)$
	Time evolution via qubitization [62]		$\mathcal{O}(L)$ bits, $\mathcal{O}(\log L)$ qubits		$\widetilde{\mathcal{O}}\left(\frac{L\lambda}{\gamma\Delta^2}\log^{3/2}\frac{1}{\varepsilon}\right)$	$\widetilde{\mathcal{O}}\left(\frac{L\lambda}{\gamma\Delta^2}\frac{1}{\varepsilon^2}\right)$
Near query optimal [69]	Block encoding H	(B2)	$\mathcal{O}(L)$ bits,	1 qubit	$\widetilde{\mathcal{O}}\left(\frac{L\lambda}{\gamma\Delta}\log\frac{1}{\varepsilon}\right)$	$\widetilde{\mathcal{O}}\left(\frac{L\lambda}{\gamma\Delta}\frac{1}{\varepsilon^2}\right)$
		None	$\mathcal{O}(\log L)$ qubits	$\mathcal{O}(\log(\frac{1}{\gamma}))$ qubits		
Early fault tolerant [33]	First-order Trotter time evolution	None	$\mathcal{O}(L)$ bits	1 qubit	$\widetilde{\mathcal{O}}\left(\frac{L^3\Lambda^2}{\Delta^2}\frac{1}{\varepsilon}\log\frac{1}{\gamma}\right)$	$\widetilde{\mathcal{O}}\left(\frac{L^3\Lambda^2}{\gamma^4\Delta^2}\frac{1}{\varepsilon^3}\right)$
	Time evolution via qubitization [62]		$\mathcal{O}(L)$ bits, $\mathcal{O}(\log L)$ qubits		$\widetilde{\mathcal{O}}\left(\frac{L\lambda}{\Delta}\log\frac{1}{\gamma\varepsilon}\right)$	$\widetilde{\mathcal{O}}\left(\frac{L\lambda}{\gamma^4\Delta}\frac{1}{\varepsilon^2}\right)$
This work (randomized quantum)	⋯	(B1)	$\mathcal{O}(L)$ bits	1 qubit	$\mathcal{O}\left(\frac{\lambda^2}{\Delta^2}\log^2\frac{1}{\gamma\varepsilon}\right)$	$\widetilde{\mathcal{O}}\left(\frac{\lambda^2}{\gamma^4\Delta^2}\frac{1}{\varepsilon^2}\right)$

As with our linear-systems algorithm, we focus on the quantum hardware requirements, which consist of the total number of logical qubits required for data access and to run the algorithm, as well as the gate depth. Similar to the linear-systems task considered in Sec. III C, our approach has the smallest space requirement and there is scope for our algorithm to have competitive gate depth if the Pauli weight λ is smaller than number of Pauli terms L (see, e.g., the values given in Ref. [75, Tables IX and X]).

Finally, we note that in Table III, we have not accounted for the run time to prepare the trial state. Other coherent approaches require coherent calls to the unitary that prepares the trial state. Thus, if the run time of this subroutine is significant, this would also cause discrepancies in the total run times listed.

V. GIBBS-STATE SAMPLING

A. Randomized quantum algorithm

We can also consider randomizing prior approaches to Gibbs-state preparation. Here, the task is usually to recover an approximation to the Gibbs state $e^{-\beta H}/\mathcal{Z}$ (also known as the thermal state) given some information about the Hamiltonian H, where $\mathcal{Z} := \mathrm{Tr}[e^{-\beta H}]$ is the partition function and β is a parameter physically corresponding to the inverse temperature. In our setting, we will recover observables with respect to the Gibbs state.

Under standard complexity-theoretic assumptions, this is expected to be hard in general [76]. However, one prominent line of work has aimed to emulate the thermalization process, following the intuition that there should exist efficient algorithms for certain physical systems [77–80] (for the current state-of-the-art approach, see Ref. [81]). This leads to complexity dependent on the mixing time, which while small for physical systems, may be difficult to estimate or may require additional assumptions to bound. Similar to classical Monte Carlo approaches for classical Hamiltonians, this nevertheless could be a promising approach to achieve efficient practical performance for interesting problems.

In this section, we follow a different line of work that aims to approximate the exponential operator directly for generic Hamiltonians [9,82–86]. When applied to one half of the maximally entangled state (with the correct normalization), this corresponds to the purification of the Gibbs state. Reflecting the expected hardness of the general problem, prior approaches have exponential run time in the inverse temperature β and operator norm $\|H\|$. A perturbative approach has been proposed in Ref. [86], in which one assumes access to the purification of the Gibbs state of an intermediate Hamiltonian H_0. In this case, the dominant contribution to the complexity instead depends

on the operator norm of the perturbation $V := H - H_0$, which could in principle be much smaller. Additionally, this dependence can be further reduced with the action of a so-called nonequilibrium unitary, though we omit this discussion for simplicity here. We note that, in the trivial case $H_0 = \mathbf{0}$, this starting assumption reduces to starting with the maximally entangled state, which is the same as the other aforementioned approaches. Here, the relevant function of interest is the exponential function of the work operator, defined as $W := H \otimes \mathbb{1} - \mathbb{1} \otimes H_0^*$. In this section, we import the insights of Ref. [86] into our randomized framework, which allows us to state the following corollary.

Corollary 3 (Gibbs-state property estimation). Suppose access to the Gibbs state of an intermediate Hamiltonian H_0 via its purification $|\Psi_0\rangle$ and the ability to measure the observable O. Further, assume that the Pauli decompositions of H_0 and H are known, as in Eq. (1), and that $[H_0, H] = 0$. Then, we give a randomized quantum algorithm to

$$\text{approximate } \frac{\mathrm{Tr}[e^{-\beta H} O]}{\mathcal{Z}} \text{ to additive error } \varepsilon, \tag{27}$$

with arbitrary constant success probability, utilizing

$$\mathcal{C}^{O}_{sample} = \widetilde{\mathcal{O}}\left(e^{2\beta\|V\|} \frac{\mathcal{Z}_0^2}{\mathcal{Z}^2} \frac{e^{\sqrt{\ln(\|O\|/\varepsilon)}}}{\varepsilon^2} \right) \text{ circuit runs}, \tag{28}$$

where $\mathcal{Z}_0 := \mathrm{Tr}[e^{-\beta H_0}]$ and $\mathcal{Z} := \mathrm{Tr}[e^{-\beta H}]$ are the partition functions of H_0 and H, respectively, and where we denote $V = H - H_0$. Each circuit is of the form in Fig. 2(b) and uses one instance of $|\Psi_0\rangle$ with at most

$$\mathcal{C}^{O}_{gate} = \widetilde{\mathcal{O}}\left(\beta^3 \lambda_W^2 \|W\|\right) \text{ gate depth}, \tag{29}$$

where λ_W is the Pauli weight of the work operator $W := H \otimes \mathbb{1} - \mathbb{1} \otimes H_0^*$.

Similar to before, these complexities can be established via Propositions 1 and 2, along with error bounds established in Ref. [86] (for details, see Appendix C 3).

B. Comparison

In Table IV, we compare the complexities of our randomized approach against other algorithms for the task in Eq. (27) with an operator O with $\|O\| \leqslant 1$. The other approaches that we consider are as follows:

(i) An LCU approach based on the Hubbard-Stratonovich transform, in which one requires access to time evolution of an operator $\widetilde{H}$ which, when squared, recovers the action of H conditioned on an ancillary register. We detail resources required to obtain a block encoding of $\widetilde{H}$ and using a qubitization approach for the time evolution [62]. Specifically, this will incur run-time overhead in the number of Pauli terms L_H and $\lambda_{\widetilde{H}} := \sum_{\ell=1}^{L_H} \sqrt{a_\ell}$, where a_ℓ are the Pauli coefficients of H.

(ii) A QSVT approach to implement polynomial approximations of the Gaussian and exponential function. We consider block encodings of $\widetilde{H}$ and H for these approaches, respectively, which again incurs overhead in L_H and $\lambda_{\widetilde{H}}$, or λ_H, which denote as the Pauli weight of H.

(iii) The perturbative approach of Ref. [86], starting from the purified Gibbs state of an intermediate Hamiltonian H_0. An LCU approximation of $e^{-\beta W/2}$ is implemented. As with the above approaches, the exponentially costly step comes from the number of rounds of amplitude amplification required. As we compare complexities with our randomized version of this algorithm, we add a second row in Table IV to demonstrate complexities using incoherent sampling rather than amplitude amplification, in which case the required gate depth is greatly reduced. Here, the run times are dependent on the number of Pauli terms constituting W and its Pauli weight, which we denote as L_W and λ_W, respectively.

(iv) Our approach, given in Corollary 3, based on the insights of Ref. [86].

Similar to our ground-state property-estimation comparison, in all the above settings we have considered incoherent sampling of the observable O in order to give the most competitive gate depth. In all cases, we assume classical access to the Pauli decomposition of H (and H_0) as a starting point. From then on, the classical and quantum resources required to process these data are recorded in the table. Specifically, all block encodings are presumed to be constructed via SELECT and PREPARE oracles. The weight of $\widetilde{H}$ can be found to satisfy $\sqrt{\lambda_H} \leqslant \lambda_{\widetilde{H}} \leqslant \sqrt{L_H \lambda_H}$. We also remark that $\|W\| \leqslant \|H\| + \|H_0\|$ and in the trivial case $H_0 = \mathbf{0}$, we have $\|W\| = \|V\| = \|H\|$. Additionally, the number of Pauli terms and the Pauli weight simply follow as $L_W = L_{H_0} + L_H$ and $\lambda_W = \lambda_{H_0} + \lambda_H$.

In Table IV, we see that our randomized approach offloads the exponential complexity in $\beta\|V\|$ from the gate depth onto the sample overhead, compared to the fully coherent approach. However, compared to replacing amplitude amplification with incoherent sampling, our approach does not incur any explicit dependence on L_W in the gate complexity or overall run time. In addition, it uses fewer qubits than any other approach. Finally, we recall that in our framework, our stated sample complexities serve as sufficient conditions and, in practice, an ε-approximation of the observable could be achieved

TABLE IV. The end-to-end complexities of estimating the expectation value of an operator O; $\|O\| = \mathcal{O}(1)$ to additive error ε with respect to the Gibbs state of Hamiltonian H on a system of n qubits. In the cells, we have omitted subpolynomial factors to simplify expressions. Certain approaches require access to a matrix $\widetilde{H}$ that recovers the action of $\sqrt{H}$ and has weight $\lambda_{\widetilde{H}}$, defined in the text, which satisfies $\sqrt{\lambda_H} \leqslant \lambda_{\widetilde{H}} \leqslant \sqrt{L_H \lambda_H}$. The perturbative approach assumes access to a purification of the Gibbs state of an intermediate Hamiltonian H_0 with complexity dependent on the perturbation $V := H - H_0$ and the work operator W defined in the text, which satisfies $\|W\| \leqslant \|H_0\| + \|H\|$, $L_W = L_{H_0} + L_H$, $\lambda_W = \lambda_{H_0} + \lambda_H$. For the algorithm of Ref. [86] we have included two rows: first, the more run-time-efficient approach, which uses amplitude amplification, and the measure-until-success approach, which uses exponentially less gate depth in β.

Algorithm	Access model	Data-access space	Algorithmic space	Quantum gate depth	Run time
LCU [82]	Time evolution $\widetilde{H}$ via qubitization [62]	$\mathcal{O}(L_H)$ bits, $\mathcal{O}(\log L_H)$ qubits	$2n + \mathcal{O}(\log \beta\|H\| + \log\log \frac{1}{\varepsilon}))$ qubits	$\widetilde{\mathcal{O}}\left(L_H \lambda_{\widetilde{H}} \sqrt{\beta} \sqrt{\frac{2^n}{\mathcal{Z}}} \log \frac{1}{\varepsilon}\right)$	$\widetilde{\mathcal{O}}\left(L_H \lambda_{\widetilde{H}} \sqrt{\beta} \sqrt{\frac{N}{\mathcal{Z}}} \frac{1}{\varepsilon^2}\right)$
QSVT [9]	Block encoding $\widetilde{H}$	$\mathcal{O}(L_H)$ bits, $\mathcal{O}(\log L_H)$ qubits	$2n + 2$ qubits	$\widetilde{\mathcal{O}}\left(L_H \lambda_{\widetilde{H}} \sqrt{\beta} \sqrt{\frac{2^n}{\mathcal{Z}}} \log \frac{1}{\varepsilon}\right)$	$\widetilde{\mathcal{O}}\left(L_H \lambda_{\widetilde{H}} \sqrt{\beta} \sqrt{\frac{2^n}{\mathcal{Z}}} \frac{1}{\varepsilon^2}\right)$
	Block encoding H	$\mathcal{O}(L_H)$ bits, $\mathcal{O}(\log L_H)$ qubits	$2n + 2$ qubits	$\widetilde{\mathcal{O}}\left(L_H \lambda_H e^{\beta/2} \sqrt{\frac{2^n}{\mathcal{Z}}} \log \frac{1}{\varepsilon}\right)$	$\widetilde{\mathcal{O}}\left(L_H \lambda_H e^{\beta/2} \sqrt{\frac{2^n}{\mathcal{Z}}} \frac{1}{\varepsilon^2}\right)$
Work operator LCU [13]	Block encoding W	$\mathcal{O}(L_W)$ bits, $\mathcal{O}(\log L_W)$ qubits	$2n + 3$ qubits	$\widetilde{\mathcal{O}}\left(L_W \lambda_W \sqrt{\|W\|} \sqrt{\frac{\mathcal{Z}_0}{\mathcal{Z}}} e^{\beta\|V\|/2} e^{\sqrt{\log 1/\varepsilon}}\right)$	$\widetilde{\mathcal{O}}\left(L_W \lambda_W \sqrt{\|W\|} \sqrt{\frac{\mathcal{Z}_0}{\mathcal{Z}}} e^{\beta\|V\|/2} \frac{1}{\varepsilon^2}\right)$
				$\widetilde{\mathcal{O}}\left(L_W \sqrt{\|W\|} \beta^{3/2} \left(\lambda_W + \|W\| \log \frac{1}{\varepsilon}\right)\right)$	$\widetilde{\mathcal{O}}\left(L_W \lambda_W \sqrt{\|W\|} \frac{\mathcal{Z}_0}{\mathcal{Z}} e^{\beta\|V\|} \frac{1}{\varepsilon^2}\right)$
Randomized quantum (this work)	Pauli coefficients of W	$\mathcal{O}(L_W)$ bits	$2n + 1$ qubits	$\widetilde{\mathcal{O}}\left(\lambda_W^2 \|W\| \beta^3 + \log \frac{1}{\varepsilon}\right)$	$\widetilde{\mathcal{O}}\left(\lambda_W^2 \|W\| \frac{\mathcal{Z}_0^2}{\mathcal{Z}^2} e^{2\beta\|V\|} \frac{1}{\varepsilon^2}\right)$

with a smaller circuit sample count. This could be especially relevant here, where our required gate depths are only polynomially large, in contrast to other approaches with exponentially large gate depth. Thus, given an efficient verifier, in our setting one could potentially collect samples from an efficient circuit and achieve convergence much faster than the stated run-time bounds.

VI. APPLICATION: ESTIMATION OF GREEN'S FUNCTIONS

So far, we have seen that our algorithms are naturally suited to settings in which the matrix of interest is already given in the Pauli basis. Thus, our algorithms are suited to physically motivated problems. One natural application is the evaluation of single-particle Green's functions in the context of many-body physics. Green's functions can be used to calculate single-particle expectation values such as the kinetic energy, as well as to determine the many-body density of states. For a more detailed background, see Refs. [87,88]. Previous works have proposed quantum approaches for preparing Green's functions in both the frequency and time domains [68,85,89–95]. One particular idea has been to use a quantum algorithm to evaluate Green's functions for the computationally expensive subroutine of the quantum impurity problem in dynamical mean-field theory (DMFT) calculations, in order to potentially extend their scope [89–92].

We define the advanced and retarded Green's function in the frequency domain (denoted as $G^{(+)}(\omega)$ and $G^{(-)}(\omega)$, respectively) as the matrix-valued functions with elements

$$G_{ij}^{(+)}(\omega) := \left\langle E_0 \middle| \hat{a}_i \left(\hbar\omega - (H - E_0) + i\eta\right)^{-1} \hat{a}_j^\dagger \middle| E_0 \right\rangle, \quad (30)$$

$$G_{ij}^{(-)}(\omega) := \left\langle E_0 \middle| \hat{a}_i^\dagger \left(\hbar\omega + (H - E_0) - i\eta\right)^{-1} \hat{a}_j \middle| E_0 \right\rangle, \quad (31)$$

where E_0 is the ground-state energy of H, η is a broadening parameter that determines the resolution of the Green's function, and $\hat{a}_i^\dagger, \hat{a}_i$ are fermionic single-particle creation and annihilation operators. We note that these quantities are expectation values of an operator that contains the inverse of some matrix, where the expectation value is taken with respect to the ground state. Thus, our earlier results in Secs. III and IV can be readily applied. The creation and annihilation operators can be expressed using

Pauli operators via the Jordan-Wigner transformation

$$\hat{a}_i = Z^{\otimes(i-1)} \otimes \frac{1}{2}(X + \mathrm{i}Y) \otimes I^{\otimes(N-i)}, \tag{32}$$

$$\hat{a}_i^\dagger = Z^{\otimes(i-1)} \otimes \frac{1}{2}(X - \mathrm{i}Y) \otimes I^{\otimes(N-i)}. \tag{33}$$

We now present our result for Green's-function estimation. From hereon, we denote $\Gamma^{(+)} = \hbar\omega - (H - E_0) + i\eta$ and $\Gamma^{(-)} = \hbar\omega + (H - E_0) - i\eta$. We remark that while it is possible to use the algorithms of Secs. III and IV as separate subroutines to evaluate the Green's functions, it is beneficial to compile the ground-state projection and matrix inversion all at once. This is the scheme that we present in the following proposition.

Proposition 5 (Green's-function estimation). Consider a Hamiltonian $H = \sum_l E_l |E_l\rangle\langle E_l|$ with all eigenvalues $E_l \geqslant 0$ and known Pauli decomposition $H = \sum_\ell a_\ell P_\ell$; $\lambda_H := \sum_\ell |a_\ell|$. Assume that the spectral gap is lower bounded as $\Delta \leqslant E_1 - E_0$. Additionally, we suppose that we can freely prepare an initial trial state $|\psi_0\rangle$ with overlap with the ground state $\gamma := |\langle\psi_0|E_0\rangle|$. Given parameters ω, η, and the ground-state energy E_0, we give a random compiler to

$$\text{approximate } G_{ij}^{(+)}(\omega) \text{ and } G_{ij}^{(-)}(\omega) \text{ up to additive error } \varepsilon, \tag{34}$$

and arbitrary constant success probability, each utilizing

$$\widetilde{\mathcal{O}}\left(\frac{\|(\Gamma^{(\pm)})^{-1}\|^2}{\gamma^4\varepsilon^2}\right) \text{ circuit runs}, \tag{35}$$

respectively, each consisting of at most

$$\widetilde{\mathcal{O}}\left(\frac{\lambda_H^2}{\Delta^2} + (|\hbar\omega \pm E_0| + \lambda_H + \eta)^2 \|\Gamma^{(\pm)-1}\|^2\right) \text{ gate depth}. \tag{36}$$

We provide a proof of this result in Appendix C 4. As with our previous algorithms, the scheme that we use has an advantage over other algorithms in that it does not use any additional ancillary qubits and it does not have any explicit dependence on the number of Pauli terms L. We remark that in the algorithm of Proposition 10, the ground-state energy is given exactly as an input. If this is not available, the ground-state energy may be approximated via the techniques of other early fault-tolerant schemes [29–32,36]. Moreover, if the ground-state energy is approximated to sufficiently small precision, the error contribution to the Green's functions can be constrained (see Appendix C 4 for more details). Finally, we remark that Green's functions in the (real-)time domain consist of expectation values of time-evolved creation and annihilation operators. Thus, again using the Jordan-Wigner transformation, this can be directly evaluated via Theorem 1 and the tools of our ground-state property-estimation algorithm, with the same asymptotic complexities as stated in Corollary 2.

VII. CONCLUSIONS AND OUTLOOK

We have presented a framework for sampling properties of general matrix Fourier series and applied this to give explicit algorithms for the linear-systems problem and the ground-state property-estimation problem. By starting with a (classical) description of the matrix in the Pauli basis, we circumvent the need for coherent data structures, which adds to the hardware burden of other quantum algorithms. Another distinct feature of our approach is that there is no explicit dimension or sparsity dependence in our complexities; instead, the run time depends on the norm of the Pauli coefficients for the matrix, which in principle can be much smaller than number of Pauli terms or the dimension of the system. As such, our framework is particularly suited to physically motivated matrices, where the Pauli description is readily available and of low weight.

There are immediate open questions that have yet to be explored:

(i) Is it possible for certain special classes of matrices to efficiently obtain the Pauli decomposition starting from a description in the computational basis? If the number of Pauli terms is known and small, up to a quadratic saving in the dimension can be made compared to the naive approach, though this would still present a barrier to any possible superpolynomial quantum run-time advantage compared to classical schemes directly working in the computational basis. The question of obtaining the Pauli decomposition also has implications for near-term schemes for linear systems that start with similar assumptions [39–41]. More broadly, this also leads into the question of what data structures or data sources are amenable to possible quantum speed-ups.

(ii) We have considered worst-case performance in two senses. First, bounds on sample complexities are constructed essentially by bounding the variances of estimators. Unlike some other approaches, we can halt our algorithm at any time if the solution is good enough with the obtained samples. It remains open whether there exists an efficient verifier for which solution quality can be checked, which could enable faster time to solution heuristically. Second, we have quoted maximum non-Clifford gate complexities, whereas in reality in our algorithm with each sample

we implement circuits with varying gate complexities. Thus, a more refined measure of unparallelized total run time should instead use the *expected* gate complexity, which can lead to significant savings in asymptotic run-time bounds [96]. We leave it as an open question as to how far this quantity can be optimized for the problems of interest discussed in this work.

(iii) A more precise numerical analysis of finite resource costs for concrete problems (such as the calculation of Green's functions for a particular problem of interest) would be illuminating for the feasibility of our schemes in the early fault-tolerant regime. One can also explore numerically obtained Fourier approximations to functions.

(iv) Our linear-systems algorithm could possibly undergo further refinements, inspired by classical algorithms. For instance, it is general practice to use preconditioners for linear-systems solvers to effectively reduce the effect of the condition number on the run time (see, e.g., Refs. [97–99]) and analogous quantum preconditioners have been studied [85,100,101]. We leave it as future work to investigate whether such techniques and beyond can be efficiently transported to the early fault-tolerant setting.

(v) We have discussed in Sec. II A various properties of the Fourier series that influence the complexities of our algorithms. In particular, in order to constrain the sample complexity, the variance of the randomized schemes should be constrained. It remains to be seen whether there are other functions of interest beyond the inverse function and Gaussian or exponential functions which have Fourier-series approximations that can lead to algorithms with favorable complexities.

Note added.—Chakraborty [102] has recently proposed a randomized scheme where, starting from a Fourier decomposition of a function, one constructs a sampling algorithm that queries a controlled time-evolution oracle.

ACKNOWLEDGMENTS

This project was started when M.B. and S.W. were additionally affiliated with the AWS Center for Quantum Computing, Pasadena, California, USA. We thank Earl Campbell for collaboration in the early stages of this project. We thank Leandro Aolita, Fernando Brandão, Gian Camilo, Chi-Fang (Anthony) Chen, Alexander M. Dalzell, Allan David Cony Tosta, Steven T. Flammia, Zoë Holmes, Thais de Lima Silva, Thiago O. Maciel, and Kianna Wan for discussions (in alphabetical order). S.W. and M.B. are supported by the Engineering and Physical Sciences Research Council (EPSRC) through Grant No. EP/W032643/1. S.W. is additionally supported by the Samsung GRP grant.

APPENDIX A: OUTLINE OF APPENDICES

In these appendixes, we present detailed statements for our theoretical results, as well as the proofs thereof.

In Appendix B, we introduce the main technical results of the paper. Specifically, in Appendix B 1, we first show how to prove the result in the warm-up problem (Proposition 1) for sampling properties of a given Fourier series, before proving our main result (Theorem 1) for sampling properties of matrix functions in Appendix B 3. In Appendix B 4, we discuss the complexities of evaluating normalized quantities that correspond to normalized quantum states. Next, in Appendix B 5, we discuss how to statistically encode classical vectors as part of our randomized scheme (Proposition 4). We then discuss the classical power of our randomized scheme for sampling properties of low-degree polynomials in Appendix B 7 (Proposition 3).

In Appendix C, we demonstrate how our algorithms for linear systems (Corollary 1), ground-state property estimation (Corollary 2), and Gibbs-state property estimation (Corrolary 3) follow from our main results. We also show how combining the linear-systems algorithm with the ground-state property-estimation algorithm allows for a scheme to evaluate Green's functions in many-body physics (Proposition 5).

APPENDIX B: FOURIER SAMPLING

1. Sampling from a given Fourier series—proof of Proposition 1

We start by introducing the random compiler lemma of Ref. [30, Lemma 2], which demonstrates how to decompose time-evolution operators into Pauli gates and Pauli rotations.

Lemma 1 (Random compiler lemma—adapted from Lemma 2 of Ref. [30]). Let $A = \sum_\ell a_\ell P_\ell$ be a Hermitian operator that is specified as a linear combination of Pauli operators with Pauli weight $\lambda := \sum_\ell |a_\ell|$ and real coefficients $a_\ell \in \mathbb{R}$. For any $t \in \mathbb{R}$ and any choice of $r \in \mathbb{N} := \{1, 2, \ldots\}$, there exists a linear decomposition

$$e^{iAt} = \sum_{m \in T} \beta_m^{(r)} U_m^{(r)} , \tag{B1}$$

for some index set T, unitaries $\{U_m^{(r)}\}_m$, and real numbers $\{\beta_m^{(r)}\}_m$ such that $\sum_{m\in T} \beta_m^{(r)} \leqslant \exp(\lambda^2 t^2/r)$. For all $m \in T$, the non-Clifford cost of controlled-$U_m^{(r)}$ is that of r controlled single-qubit Pauli rotations.

Proof. We note that Ref. [30] has provided an explicit proof for operators with normalized Pauli weight and positive coefficients (i.e., the coefficients are probabilities). We

leave the core proof to Ref. [30, Appendix C] and for completeness here we explicitly demonstrate how the proof extends to the more general setting.

First, we consider the scenario in which $\lambda = 1$ but a_ℓ may not necessarily be positive. Reference [30] expresses an rth of the time-evolution operator as

$$e^{iAt/r} = \sum_{n \text{ even}} \gamma_n(t/r) \sum_{\ell_1,\ldots,\ell_n,\ell'} a_{\ell_1} \ldots a_{\ell_n} a_{\ell'} P_{\ell_1} \ldots P_{\ell_n} \times \exp\left(i\theta_n P_{\ell'}\right) \tag{B2}$$

$$= \sum_{n \text{ even}} \gamma_n(t/r) \sum_{\ell_1,\ldots,\ell_n,\ell'} |a|_{\ell_1} \ldots |a|_{\ell_n} |a|_{\ell'} \cdot \text{sgn} \times \left(a_{\ell_1} \ldots a_{\ell_n} a_{\ell'}\right) P_{\ell_1} \ldots P_{\ell_n} \exp\left(i\theta_n P_{\ell'}\right), \tag{B3}$$

where $\gamma_n(t/r)$ are coefficients satisfying $\sum_{(n\,\text{even})} \gamma_n(t/r) \leqslant \exp(t^2/r^2)$. In the second line, we see that we can proceed as if the coefficients were positive (probabilities), by absorbing their sign into the string of unitaries to implement $P_{\ell_1} \ldots P_{\ell_n} \exp\left(i\theta_n P_{\ell'}\right)$. By considering the product of such fractional time-evolution operators $(e^{iAt/r})^r$, we can then sample strings of Pauli operators and r Pauli rotations (with an absorbed phase), with total weight $\leqslant (\exp(t^2/r^2))^r = \exp(t^2/r)$. This consists of r pairs of controlled (multiqubit) Pauli rotations and a string of controlled Pauli gates, the number of which is in theory unbounded. We remark, however, that the number of Pauli gates is with high probability zero, with exponentially decaying probability for increasing gate number [30]. Further, these gates can be efficiently compiled together classically into a single controlled (multiqubit) Pauli gate using standard Pauli product rules.

We can now consider how to deal with nonunit Pauli weight. For H with nonunit Pauli weight λ, we can simply consider the equality $e^{iAt} = e^{i\hat{A}\lambda t}$, where $\hat{A} = A/\lambda$ now has unit Pauli weight. As Lemma 1 holds for all $t \in \mathbb{R}$, the steps of the original proof can follow, by considering an extended time parameter $t \to \lambda t$. ■

With the above lemma, we can now demonstrate how to solve our warm-up problem of sampling properties of a given Fourier series. In the following, we present a more precise version of Proposition 1 from the main text.

Proposition 6 (Sampling from Fourier series—detailed version). Suppose that we have a Fourier series

$$s(A) = \sum_{k \in F} \alpha_k \exp\left(it_k A\right), \tag{B4}$$

in some $N \times N$ Hermitian matrix A, and denote the ℓ_1-norm of the coefficients as $\alpha := \sum_{k\in F} |\alpha_k|$. Suppose further that A has known Pauli decomposition $A = \sum_\ell a_\ell P_\ell$ with Pauli weight $\lambda = \sum_\ell |a_\ell|$. Then:

(a) Given a procedure to prepare the pure states $|\psi\rangle$, $|\phi\rangle$ with respective unitaries U_ψ, U_ϕ with respective gate depths d_ψ and d_ϕ, we have a randomized quantum algorithm that uses $\log(N) + 1$ qubits to approximate $\langle\phi|s(A)|\psi\rangle$ up to additive error ε with probability at least $1 - \delta$, using Algorithm 1 with

$$\mathcal{C}^{\phi}_{sample} = \mathcal{O}\left(\log\left(\frac{2}{\delta}\right)\frac{\alpha^2}{\varepsilon^2}\right), \tag{B5}$$

circuit samples, where each circuit takes the form in Fig. 2(a) and has depth

$$\mathcal{C}^{\phi}_{gate} = \mathcal{O}\left(\lambda^2 t^2_{\max} + d_\psi + d_\phi\right), \tag{B6}$$

where we denote $t_{\max} = \max_{k\in F} t_k$.

(b) Given a procedure to prepare the quantum state ρ in depth d_ρ and perform measurements with measurement operator O, we give a randomized quantum algorithm that uses $\log(N) + 1$ qubits to approximate $\text{Tr}\left[s(A)\rho s(A)^\dagger O\right]$ up to additive error ε with probability at least $1 - \delta$ using Algorithm 2 with

$$\mathcal{C}^{O}_{sample} = \mathcal{O}\left(\log\left(\frac{2}{\delta}\right)\frac{\|O\|^2\alpha^4}{\varepsilon^2}\right), \tag{B7}$$

circuit samples, where each circuit takes the form in Fig. 2(b) and has depth

$$\mathcal{C}^{\phi}_{gate} = \mathcal{O}\left(\lambda^2 t^2_{\max} + d_\psi + d_\phi\right). \tag{B8}$$

Proof of Proposition 6. We note that from Lemma 1, $s(\varepsilon, A)$ can be decomposed via a linear combination as

$$s(A) = \sum_{k\in F} \alpha_k \exp\left(it_k A\right) = \sum_{(k,m)\in F\times T} \alpha_k \beta^{(r_k)}_{km} U^{(r_k)}_{km}, \tag{B9}$$

where $\sum_{m\in T} \beta^{(r_k)}_{km} \leqslant \exp(\lambda^2 t_k^2/r_k)$ and $U^{(r_k)}_{km}$ consists of r_k non-Clifford operations. The above expression can further be seen as a quantity that is proportional to a probability distribution over unitaries

$$s(A) = R(\vec{r}) \sum_{(k,m)\in F\times T} \frac{|\alpha_k \beta^{(r_k)}_{km}|}{R(\vec{r})} \widetilde{U}^{(r_k)}_{km} = R(\vec{r}) \sum_{(k,m)\in F\times T} p^{(r_k)}_{km} \widetilde{U}^{(r_k)}_{km}, \tag{B10}$$

ALGORITHM 2. Fourier sampling of $\mathrm{Tr}\left[s(A)\rho s(A)^\dagger O\right]$.

Problem input:

- $N \times N$ Hermitian matrix $A = \sum_\ell a_\ell P_\ell$, with $a_\ell \in \mathbb{R}\ \forall \ell$, and known Pauli weight $\lambda = \sum_\ell |a_\ell|$
- Fourier parameters $\{\alpha_k\}_{k\in F}$, $\{t_k\}_{k\in F}$ for series $s(A)$ according to Eq. (2)
- Input state ρ and measurement operator O.

Parameters: (1) Approximation parameter ε;;
(2) Success probability $1-\delta$;;
(3) Runtime vector $\vec{r} \in \mathbb{N}^{|F|}$.

Output: Approximation to $\mathrm{Tr}\left[s(A)\rho s(A)^\dagger O\right]$ with additive error at most ε and probability at least $1-\delta$.

1. Compute coefficients $\{\beta_{km}^{(\vec{r})}\}_{(k,m)\in F\times T}$ as in Eq. (A1) for each time parameter $\{t_k\}_{k\in F}$.
2. $R(\vec{r}) \leftarrow \sum_{(k,m)\in F\times T} \left|\alpha_k \beta_{km}^{(r_k)}\right|$.
3. $M(\vec{r}) \leftarrow \left\lceil 4\ln(2/\delta)\left(R(\vec{r})/\varepsilon\right)^2 \right\rceil$.
4. **for** $j \in \left[M(\vec{r})\right]$:
   (i) Sample two strings of gates independently as:
      (a) Sample indices (k', m') according to probability distribution $\{|\alpha_k \beta_{km}^{(r_k)}|/R(\vec{r})\}_k$ and select corresponding string of gates $U_{k'm'}^{(r_{k'})}$
      (b) Repeat with new independent sample (k'', m''), select $U_{k'm'}^{(r_{k''})}$.
   (ii) Run $n+1$ qubit circuit in Figure 2(b) with input state $|0\rangle\langle 0| \otimes \rho$, controlled $U_{k'm'}^{(r_{k'})}$ and anti-controlled $U_{k''m''}^{(r_{k''})}$, and measurement $X \otimes O$. Record measurement outcome o_j.
   (iii) $z_j \leftarrow R(\vec{r})^2 o_j$.
5. **end for**
6. $\bar{z}^{(M)} \leftarrow \sum_j z_j / M(\vec{r})$. **return** $\bar{z}^{(M)}$.

where $p_{km}^{(r_k)} = |\alpha_k \beta_{km}^{(r_k)}|/R(\vec{r})$ are probabilities, $\widetilde{U}_{km}^{(r_k)} = U_{km}^{(r_k)}(\alpha_k \beta_{km}^{(r_k)})/|\alpha_k \beta_{km}^{(r_k)}|$ are unitaries that absorb the phase of the coefficients, and $R(\vec{r})$ is the weight of the linear combination, which satisfies

$$R(\vec{r}) = \sum_{(k,m)\in F\times T} \left|\alpha_k \beta_{km}^{(r_k)}\right| \tag{B11}$$

$$\leqslant \sum_{k\in F} |\alpha_k| \exp(\lambda^2 t_k^2/r_k) \tag{B12}$$

$$= \alpha \exp(\lambda^2 t_k^2/r_k), \tag{B13}$$

where in the last line we have denoted $\alpha := \sum_{k\in F} |\alpha_k|$. We note that one is free to tune $\vec{r} = (r_1, \ldots, r_{|F|})$, and in doing so, change the gate depth of the circuit to apply each unitary, while also changing $R(\vec{r})$, which will feed into the sample complexity. One simple choice is to set $r_k = \lambda^2 t_k^2$ for all k, which gives $R(\vec{r}) \leqslant \alpha e$, where we recall that α is the weight of the coefficients in the Fourier series. We now consider our two settings separately.

(i) *Preparation of* $\mathrm{Tr}\left[s(\varepsilon, A)\rho s(\varepsilon, A)^\dagger O\right]$. Using Eq. (B10) can now express this quantity as

$$\mathrm{Tr}\left[s(A)\rho s(A)^\dagger O\right] = R(\vec{r})^2 \sum_{(k,m),(k',m')\in F\times T} p_{km}^{(r_k)} p_{k'm'}^{(r_{k'})} \times \mathrm{Tr}\left[\widetilde{U}_{km}^{(r_k)} \rho \left(\widetilde{U}_{k'm'}^{(r_{k'})}\right)^\dagger O\right]. \tag{B14}$$

As described in Ref. [35], given controlled access to $\widetilde{U}_{km}^{(r_k)}$ and $\widetilde{U}_{k'm'}^{(r_{k'})}$, we can collect measurement shots corresponding to the quantity $1/2\left(\mathrm{Tr}[\widetilde{U}_{km}^{(r_k)}\ \rho(\widetilde{U}_{k'm'}^{(r_{k'})})^\dagger O] + \mathrm{Tr}[\widetilde{U}_{k'm'}^{(r_{k'})}\rho(\widetilde{U}_{km}^{(r_k)})^\dagger O]\right)$ with the quantum circuit in Fig. 2(b). This leads to an unbiased estimator for $1/(R(\vec{r})^2)\mathrm{Tr}\left[s(A)\rho s(A)^\dagger O\right]$. Moreover, according to Born's rule, the individual measurement outcomes o_j take values in the interval $[-\|O\|, \|O\|]$. We now propose the following (informal) algorithm. For some choice of $\vec{r}$: (1) sample according to probability distribution $\{p_{km}^{(r_k)} p_{k'm'}^{(r_{k'})}\}_{kmk'm'}$; (2) prepare the circuit for $\mathrm{Tr}[\widetilde{U}_{km}^{(r_k)}\rho(\widetilde{U}_{k'm'}^{(r_{k'})})^\dagger O]$ and collect measurement result z_j;

(3) multiply result by $R(\vec{r})^2$ to obtain z_j; (4) repeat the procedure and average over M samples. We present the full formal steps in Algorithm 2. As we are effectively sampling numbers in the interval $[-\|O\|R(\vec{r})^2, \|O\|R(\vec{r})^2]$, Hoeffding's inequality specifies that $\bar{z}^{(M)} := 1/M \sum_{i=1}^{M} z_j$ satisfies

$$\text{Prob}\left(\left| \bar{z}^{(M)} - \text{Tr}\left[s(A)\rho s(A)^\dagger O \right] \right| \geqslant \varepsilon \right) \leqslant 2\exp\left(-\frac{M\varepsilon^2}{2\|O\|^2 R(\vec{r})^4} \right). \tag{B15}$$

This implies that, in order to guarantee $\left|\bar{z}^{(M)} - \text{Tr}\left[s(A)\rho s(A)^\dagger O\right]\right| \leqslant \varepsilon$ with probability at least $1-\delta$, it is sufficient to perform

$$M \geqslant \log\left(\frac{2}{\delta}\right) \frac{2\|O\|^2 R(\vec{r})^4}{\varepsilon^2} \tag{B16}$$

circuit samples.

We now impose the choice $r_k = \lambda^2 t_k^2$ for all k, which sets $R = \mathcal{O}(\alpha)$. As specified by Lemma 1, this means that each $\widetilde{U}_{km}^{(r_k)}$ consists of layers of controlled Pauli gates in between $\lambda^2 t_k^2$ single-qubit controlled Pauli rotations. Thus, in order to obtain an ε-close approximation with probability at least $1-\delta$, it is sufficient to take a number of samples and non-Clifford gates satisfying

$$\mathcal{C}^O_{\text{sample}} = 2e^4 \log\left(\frac{2}{\delta}\right) \frac{\|O\|^2\alpha^4}{\varepsilon^2}, \quad \mathcal{C}^O_{\text{gate}} = 2\lambda^2 t_{\max}^2 + d_\rho, \tag{B17}$$

respectively, where we have denoted $t_{\max} = \max_{k\in F} t_k$.

(ii) *Preparation of* $\langle\phi|s(A)|\psi\rangle$. We can express this quantity as

$$\langle\phi|s(A)|\psi\rangle = R(\vec{r}) \sum_{(k,m)\in F\times T} p_{km}^{(r_k)} \langle 0|U_\phi^\dagger \widetilde{U}_{km}^{(r_k)} U_\psi|0\rangle. \tag{B18}$$

The real and imaginary parts of the quantity $\langle 0|U_\phi^\dagger \widetilde{U}_{km}^{(r_k)} U_\psi|0\rangle$ can be recovered separately by the circuit in Fig. 2(a). In both cases, individual measurement outcomes lie in the interval $[-1, 1]$. As before, Hoeffding's inequality tells us that

$$M \geqslant \log\left(\frac{2}{\delta}\right) \frac{4R(\vec{r})^2}{\varepsilon^2} \tag{B19}$$

shots are required to recover both the imaginary and real parts with probability at least $1-\delta$, to additive error ε. We can again choose $r_k = \lambda^2 t_k^2$ for all k, giving respective sample and gate counts:

$$\mathcal{C}^\phi_{\text{sample}} = 4e^2 \log\left(\frac{2}{\delta}\right) \frac{\alpha^2}{\varepsilon^2}, \quad \mathcal{C}^\phi_{\text{gate}} = \lambda^2 t_{\max}^2 + d_\phi + d_\psi. \tag{B20}$$

■

a. Constant-factor trade-offs

In the above, we set the elements of the run-time vector to the value $r_k = \lambda^2 t_k^2$ for all k. However, we remark that in general, by tuning $\vec{r}$ slightly, one can make small constant-factor trade-offs between the sample and gate complexity. Namely, by instead setting $r_k = a\lambda^2 t_k^2$ for some constant a, one can reduce the sample complexity by a factor $e^{2-2/a}$ for problem (a) and a factor $e^{4-4/a}$ for problem (b), at a cost of a increase in gate depth by a factor a.

2. Classical preprocessing cost

The focus of this work is to characterize the cumulative quantum run time. This is motivated by the fact that we expect quantum clock speeds to be significantly slower than classical ones. However, it is still useful to quantify any classical preprocessing costs in order to carry out full end-to-end comparisons with classical algorithms, which is what we do in this section.

To this end, we indicate the complexities for the classical preprocessing in Algorithm 1, which is equivalent to those for Algorithm 2. As in the analysis above in Sec. B 1, we assume that we set all elements of the run-time vector to be $r_k = \lambda^2 t_k^2$. The two preprocessing overheads that we need to characterize are, first, to evaluate the weight $R(\vec{r}) = \sum_{(k,m)\in F\times T} |\alpha_k \beta_{km}^{(r_k)}|$ of the sampling protocol (step (2) of Algorithm 1) and, second, any further preprocessing in order to construct sampling access to our desired probability distribution [allowing step (4) of Algorithm 1].

a. Evaluating $R(\vec{r})$

In Ref. [30, Appendix C], it is given that the weight of the coefficients in Eq. (B1) explicitly satisfy

$$\sum_{m\in T} \beta_m^{(r_k)} = \sum_{n=0}^{\infty} \frac{1}{(2n)!} \left(\frac{1}{\lambda t_k}\right)^{2n} \sqrt{1 + \left(\frac{(1/\lambda t_k)}{2n+1}\right)^2}. \tag{B21}$$

If there are $K = |F|$ Fourier terms in the Fourier series in Eq. (B4), then K such quantities need to be evaluated. If there is no efficient way to evaluate the sum in Eq. (B21), we remark that the sum has exponentially vanishing terms and can also be truncated, which we discuss more below. We denote the truncation degree to the sum in Eq. (B21),

which leads to additive error ε_k as $J(\varepsilon_k)$. In order to keep the total additive error ε, for each Fourier term k it is sufficient to pick $\varepsilon_k = \mathcal{O}(\varepsilon/\alpha_k) \geqslant \mathcal{O}(\varepsilon/\alpha)$. Thus, it is sufficient to approximate $R(\vec{r})$ in time $\mathcal{O}(K \cdot J(\varepsilon/\alpha))$.

b. Sample access preprocessing

In order to sample from a discrete probability distribution with S unique probabilities in $\mathcal{O}(1)$ time, it is sufficient to spend $\mathcal{O}(S)$ memory and preprocessing time. This can be achieved with, e.g., the alias method [103]. In step (4) of Algorithm 1, we must sequentially sample the index $k \in F$ followed by $m \in T$. In order to sample the index m, we sample from the probability distribution proportional to the linear combination in Eq. (B1), which consists of sampling a Taylor-series order n followed by sampling a Pauli term $n+1$ times. Setting up sampling access to the Pauli coefficients of A costs $\mathcal{O}(L)$ classical overhead. Setting up sampling overhead to the Taylor-series order has linear overhead in the truncation order. Again denoting the truncation order of Eq. (B21), which leads to additive error ε/α, as $J(\varepsilon/\alpha)$, we can write the classical preprocessing overhead to construct sample access as $\mathcal{O}(K + L + J(\varepsilon/\alpha))$. Additionally, each time the loop in step (4) repeats (corresponding to one quantum sample), there is at most $\mathcal{O}(J(\varepsilon/\alpha))$ overhead corresponding to sampling from the probability distribution according to the Pauli coefficients of A $\mathcal{O}(n)$ times.

It is shown in Ref. [30, Theorem 4] that it is sufficient to pick a truncation order $J(\varepsilon_k) = \mathcal{O}(\log(\lambda^2 t_k^2/\varepsilon_k))$ to approximate Eq. (B21) to a given error ε_k. Thus, in order to preserve the complexity guarantees specified by Proposition 6 with error in solution ε, it is sufficient to choose truncation order $J(\varepsilon/\alpha) = \mathcal{O}(\log(\alpha\lambda^2 t_k^2/\varepsilon))$ for each Fourier term labeled by k. Putting everything together, we have preprocessing overhead $\mathcal{O}(K + L + J(\varepsilon/\alpha) + K\log(\alpha\lambda^2 t_{\max}^2/\varepsilon)) = \mathcal{O}(L + K\log(\alpha\lambda^2 t_{\max}^2/\varepsilon))$. Additionally, each quantum sample comes with a classical overhead.

We comment that in Fourier-approximation approaches to quantum algorithms, a classical complexity depending on K is generic, as at minimum one must store K values and process them. For the algorithms we consider in the rest of this paper, K is polynomial in all problem parameters. One thing that is beneficial with the randomized framework is that K only appears in the classical preprocessing overhead. In contrast, for instance, in the LCU approach, K would appear logarithmically in the qubit overhead and linearly in the gate overhead. Similarly, a classical overhead of $\mathcal{O}(L)$ is generic for our Pauli access model. In order to obtain a better complexity, one may hope for additional structure in the Pauli basis, for instance, if the Pauli coefficients only take a small number of unique values. This can be thought of as analogous to the sparse-access model, where in order to avoid $\mathcal{O}(N)$ data-access overhead, it is not sufficient with known techniques simply to have a sparse matrix; the values must also be efficiently computable.

3. Generalized Fourier sampling—proof of Theorem 1

We first establish a lemma relating the closeness of operators to the closeness of expectation values constructed from them.

Lemma 2 (Tightness of expectation values, part 1). Consider two operators Y and D that satisfy the closeness relation $\|Y - D\| \leqslant \varepsilon \leqslant 1$, $\|Y\| \geqslant 1$. Then, we have

$$\left|\mathrm{Tr}[OY\rho Y^\dagger] - \mathrm{Tr}[OD\rho D^\dagger]\right| \leqslant 3\|O\|\|Y\|\varepsilon. \tag{B22}$$

Proof. We have

$$\begin{aligned}
&\left|\mathrm{Tr}[OY\rho Y^\dagger] - \mathrm{Tr}[OD\rho D^\dagger]\right| \\
&\quad = \left|\mathrm{Tr}[O(Y-D)\rho Y^\dagger] + \mathrm{Tr}[O(D-Y)\rho(Y^\dagger - D^\dagger)]\right. \\
&\qquad \left. + \mathrm{Tr}[OY\rho(Y^\dagger - D^\dagger)]\right| && \text{(B23)} \\
&\quad \leqslant \|O\|\left(\|Y-D\|\|Y^\dagger\| + \|Y-D\|^2 + \|Y\|\|Y-D\|\right) && \text{(B24)} \\
&\quad \leqslant \|O\|\left(2\|Y\|\varepsilon + \varepsilon^2\right) && \text{(B25)} \\
&\quad \leqslant 3\|O\|\|Y\|\varepsilon\,, && \text{(B26)}
\end{aligned}$$

where in the first line we have added and subtracted $\mathrm{Tr}[OD\rho Y^\dagger] + \mathrm{Tr}[OY\rho Y^\dagger] + \mathrm{Tr}[OY\rho D^\dagger]$; the first inequality is due to the triangle inequality, Hölder's tracial-matrix inequality and the submultiplicativity of the operator norm; and the second and third inequalities are due to our starting assumptions. ■

We can now proceed with the proof of the main theorem (Theorem 1). We suppose that we have a matrix function $f(A)$ that is approximated by a Fourier series as

$$\begin{aligned}
&\|f(A) - s(\varepsilon, A)\| \leqslant \varepsilon\,; \\
&\quad s(\varepsilon, A) = \sum_{k \in F_{\varepsilon,A}} \alpha_k(\varepsilon, A)\exp\left(it_k(\varepsilon, A)A\right), && \text{(B27)}
\end{aligned}$$

for some set of Fourier parameters $\alpha_k(\varepsilon, A)$ and $t_k(\varepsilon, A)$, for any $\varepsilon \leqslant 1$. Denote the ℓ_1-norm of the coefficients as $\alpha(\varepsilon) := \sum_k |\alpha_k(\varepsilon)|$. Suppose further that A has known Pauli decomposition $A = \sum_\ell a_\ell P_\ell$ with Pauli weight $\lambda = \sum_\ell |a_\ell|$ and that we are given some normalization constant q.

Theorem 2 (Generalized sampling from Fourier approximations—detailed version). Suppose that we have a

matrix function $f(A)$ that is approximated by a Fourier series as

$$\|f(A) - s(\tilde{\varepsilon}, A)\| \leqslant \tilde{\varepsilon} \leq 1;$$
$$s(\tilde{\varepsilon}, A) = \sum_{k \in F_{\tilde{\varepsilon},A}} \alpha_k(\tilde{\varepsilon}, A) \exp\left(i t_k(\tilde{\varepsilon}, A) A\right), \tag{B28}$$

where A is some $N \times N$ Hermitian matrix and for some set of Fourier parameters $\alpha_k(\tilde{\varepsilon}, A)$ and $t_k(\tilde{\varepsilon}, A)$ with tunable error parameter $\tilde{\varepsilon}$. Denote the ℓ_1-norm of the coefficients as $\alpha(\tilde{\varepsilon}, A) := \sum_k |\alpha_k(\tilde{\varepsilon}, A)|$. Suppose further that A has known Pauli decomposition $A = \sum_\ell a_\ell P_\ell$ with Pauli weight $\lambda = \sum_\ell |a_\ell|$. Then:

(a) Given unitaries U_ϕ and U_ψ to prepare $|\phi\rangle$ and $|\psi\rangle$, respectively, in depth d_ϕ, d_ψ, we give a randomized quantum algorithm that uses $\log(N) + 1$ qubits to approximate the quantity $1/q\langle\phi|f(A)|\psi\rangle$ up to additive error $\varepsilon \leqslant 1/q$ with probability at least $1 - \delta$, using

$$\mathcal{C}^{\phi}_{sample} = \mathcal{O}\left(\log\left(\frac{2}{\delta}\right)\frac{[\alpha(\varepsilon q/2, A)]^2}{\varepsilon^2 q^2}\right),$$
$$\mathcal{C}^{\phi}_{gate} = \mathcal{O}\left(\lambda^2[t_{max}(\varepsilon q/2, A)]^2 + d_\phi + d_\psi\right), \tag{B29}$$

circuit samples (each calling U_ϕ and U_ψ one time) and gate depth, respectively, where each circuit takes the form in Fig. 2(a).

(b) Given a procedure to prepare the quantum state ρ in gate depth d_ρ and perform measurements with measurement operator O, and given normalization constant q, there exists a randomized quantum algorithm that uses $\log(N) + 1$ qubits to approximate $(1/q^2)\mathrm{Tr}\left[f(A)\rho f(A)^\dagger O\right]$ up to additive error $\varepsilon \leqslant \|O\|\|f(A)\|/q^2$ with probability at least $1 - \delta$, using

$$\mathcal{C}^{O}_{sample} = \mathcal{O}\left(\log\left(\frac{2}{\delta}\right)\frac{\|O\|^2\left[\alpha\left(\frac{\varepsilon q^2}{6\|O\|\|f(A)\|}, A\right)\right]^4}{\varepsilon^2 q^4}\right),$$
$$\mathcal{C}^{O}_{gate} = \mathcal{O}\left(\lambda^2\left[t_{\max}\left(\frac{\varepsilon q^2}{6\|O\|\|f(A)\|}, A\right)\right]^2 + d_\rho\right), \tag{B30}$$

circuit samples and non-Clifford gates, respectively, where we denote $t_{max}(\varepsilon, A) = \max_{k \in F_{\varepsilon,A}} t_k(\varepsilon, A)$ and each circuit takes the form in Fig. 2(b).

Proof of Theorem 1 (Generalized sampling from Fourier approximations). We consider the two cases separately.

(i) *Preparation of* $(1/q^2)\mathrm{Tr}\left[f(A)\rho f(A)^\dagger O\right]$. We will use Proposition 6 to statistically approximate the quantity $(1/q^2)\mathrm{Tr}\left[s(\tilde{\varepsilon}, A)\rho s(\tilde{\varepsilon}, A)^\dagger O\right]$. Using Lemma 2, we will then show that, for appropriately chosen $\tilde{\varepsilon}$, this leads to an ε-close approximation of $(1/q^2)\mathrm{Tr}\left[f(A)\rho f(A)^\dagger O\right]$.

Recall that we denote $\bar{z}^{(M)}$ as the statistical approximation to $\mathrm{Tr}\left[s(\tilde{\varepsilon}, A)\rho s(\tilde{\varepsilon}, A)^\dagger O\right]$ using Algorithm 2 with M shots. We would like to find parameters such that

$$\left|\frac{1}{q^2}\bar{z}^{(M)} - \frac{1}{q^2}\mathrm{Tr}\left[f(A)\rho f(A)^\dagger O\right]\right| \leqslant \varepsilon. \tag{B31}$$

We first note that due to the triangle inequality, we can write

$$\begin{aligned}\left|\frac{1}{q^2}\bar{z}^{(M)} - \frac{1}{q^2}\mathrm{Tr}\left[f(A)\rho f(A)^\dagger O\right]\right| &\leqslant \left|\frac{1}{q^2}\bar{z}^{(M)} - \frac{1}{q^2}\mathrm{Tr}\left[s(\tilde{\varepsilon}, A)\rho s(\tilde{\varepsilon}, A)^\dagger O\right]\right| + \\ &+ \left|\frac{1}{q^2}\mathrm{Tr}\left[s(\tilde{\varepsilon}, A)\rho s(\tilde{\varepsilon}, A)^\dagger O\right] - \frac{1}{q^2}\right. \\ &\left.\times\ \mathrm{Tr}\left[f(A)\rho f(A)^\dagger O\right]\right|. \end{aligned} \tag{B32}$$

This separates the approximation error into two components, the statistical contribution of sampling from the Fourier series and the exact contribution from the quality of the Fourier-series approximation. It is sufficient to satisfy Eq. (B31) by requiring both terms on the right-hand side of Eq. (B32) to each by bounded by $1/(2)\varepsilon$.

From Lemma 2, we can see that the second term in Eq. (B32) is bounded by $1/(2)\varepsilon$ by finding Fourier parameters such that $\|f(A) - s(\tilde{\varepsilon}, A)\| \leqslant 1/(6)\varepsilon\frac{q^2}{\|O\|\|f(A)\|} \leqslant 1$. Thus, it is sufficient to pick $\tilde{\varepsilon} = \frac{1}{6}\varepsilon(q^2)/\|O\|\|f(A)\|$. From Proposition 1, the first term in Eq. (B32) is bounded by $1/(2)\varepsilon$ by choosing $M \geqslant 8e^4\log(2/\delta)\left(\|O\|^2\alpha(\varepsilon q^2/6\|O\|\|f(A)\|)^4\right)/\varepsilon^2 q^4$, $\mathcal{C}_{\mathrm{gate}} = 2\lambda^2 t^2_{\max}(\varepsilon q^2/6\|O\|\|f(A)\|) + d_\rho$.

(ii) *Preparation of* $1/q\langle\phi|f(A)|\psi\rangle$. Denote as $T'^{(M')}$ the statistical approximation to $\langle\phi|f(A)|\psi\rangle$ using Algorithm 2 with M' shots. Again, we can split up the approximation error into respective statistical and exact contributions as

$$\begin{aligned}\left|\frac{1}{q}T'^{(M')} - \frac{1}{q}\langle\phi|f(A)|\psi\rangle\right| &\leqslant \left|\frac{1}{q}T'^{(M')} - \frac{1}{q}\langle\phi|s(\tilde{\varepsilon}, A)|\psi\rangle\right| \\ &+ \left|\frac{1}{q}\langle\phi|s(\tilde{\varepsilon}, A)|\psi\rangle - \frac{1}{q}\langle\phi|f(A)|\psi\rangle\right|. \end{aligned} \tag{B33}$$

The second term in Eq. (B33) (exact contribution) is simply bounded as

$$\left|\frac{1}{q}\langle\phi|s(\tilde{\varepsilon}, A)|\psi\rangle - \frac{1}{q}\langle\phi|f(A)|\psi\rangle\right| \leqslant \frac{1}{q}\|s(\tilde{\varepsilon}, A) - f(A)\|. \tag{B34}$$

Thus, in order to bound the exact contribution by $1/(2)\varepsilon$, using Lemma 2, is sufficient to find Fourier parameters such that $\|s(\tilde{\varepsilon}, A) - f(A)\| \leqslant \varepsilon q/2 \leqslant 1$ and we can assign $\tilde{\varepsilon} = \varepsilon q/2$. The first term in Eq. (B33) (statistical contribution) is then bounded by $1/(2)\varepsilon$ by choosing $M' \geqslant 8e^4 \log(2/\delta)\,(\|O\|^2\alpha(\varepsilon q/2)^4)/\varepsilon^2 q^2$, $\mathcal{C}_{\text{gate}} = 2\lambda^2 t_{\max}^2(\varepsilon q/2) + d_\phi + d_\psi$. ■

We remark that in the above proof we have split the error into statistical and exact contributions evenly. In many cases, $\alpha(\varepsilon, A)$ can have a subpolynomial dependence on $1/\varepsilon$, such as for all three of our example algorithms. In this case, by distributing the error unevenly one can make a constant-factor saving to the sample complexity.

4. Sampling normalization constant—proof of Proposition 2

In this section, we investigate the complexity of sampling $q = \|f(A)|\phi\rangle\|$ for use as a normalization constant, where $|\phi\rangle$ is some given preparable input state. In the previous sections, we have found an estimator $\bar{z}^{(M)} := 1/M \sum_{j=1}^{M} z_j$ constructed from samples z_j which is ε-close to a desired property of a matrix function. We similarly denote our statistical approximation to q^2 as $Q^{(M_q)}$. In the following lemma, we establish how well $Q^{(M_q)}$ must approximate q^2 such that normalized quantities are additive ε-close.

Lemma 3 (Normalization precision). Suppose that we have some $\bar{z}^{(M)} \in \mathbb{R}$, pure state $|\phi\rangle$, and operators O and $f(A)$ that satisfy

$$\left|\frac{1}{q^2}\bar{z}^{(M)} - \frac{1}{q^2}\langle\phi|f(A)Of(A)^\dagger|\phi\rangle\right| \leqslant \varepsilon \leqslant 1, \tag{B35}$$

with $\|O\| \geqslant 1$. Further, denote $q = \|f(A)|\phi\rangle\|_2$. Then, if $Q^{(M_q)}$ is the statistical approximation of q^2, one can achieve

$$\left|\frac{1}{Q^{(M_q)}}\bar{z}^{(M)} - \frac{1}{q^2}\langle\phi|f(A)Of(A)^\dagger|\phi\rangle\right| \leqslant 3\varepsilon, \tag{B36}$$

if $Q^{(M_q)}$ satisfies $|Q^{(M_q)} - q^2| \leqslant \frac{1}{2}\frac{\varepsilon q^2}{\|O\|}$.

Proof. First, we note that due to the triangle inequality and submultiplicativity, we have

$$\begin{aligned}
&\left|\frac{1}{Q^{(M_q)}}\bar{z}^{(M)} - \frac{1}{q^2}\langle\phi|f(A)Of(A)^\dagger|\phi\rangle\right| \\
&\quad\leqslant \left|\frac{1}{Q^{(M_q)}}\bar{z}^{(M)} - \frac{1}{q^2}\bar{z}^{(M)}\right| \\
&\qquad + \left|\frac{1}{q^2}\bar{z}^{(M)} - \frac{1}{q^2}\langle\phi|f(A)Of(A)^\dagger|\phi\rangle\right| &\text{(B37)}\\
&\quad\leqslant \left|\frac{1}{Q^{(M_q)}} - \frac{1}{q^2}\right|\left|\bar{z}^{(M)}\right| + \varepsilon &\text{(B38)}\\
&\quad\leqslant \left|\frac{1}{Q^{(M_q)}} - \frac{1}{q^2}\right|(q^2\|O\| + q^2\varepsilon) + \varepsilon\,, &\text{(B39)}
\end{aligned}$$

assuming $\varepsilon \leqslant 1$, where in the penultimate inequality we have used the fact that $\left|\bar{z}^{(M)}\right| \leqslant \langle\phi|f(A)Of(A)^\dagger|\phi\rangle + q^2\varepsilon$, where $\langle\phi|f(A)Of(A)^\dagger|\phi\rangle \leqslant \|O\|_\infty\|f(A)|b\rangle\langle b|f(A)^\dagger\|_1 \leqslant \|O\|_\infty\|f(A)|\phi\rangle\|_2^2$. Denoting $|Q^{(M_q)} - q^2| \leqslant \nu$, we can observe

$$\begin{aligned}
\left|\frac{1}{Q^{(M_q)}} - \frac{1}{q^2}\right| &\leqslant \frac{\nu}{Q^{(M_q)}q^2} &\text{(B40)}\\
&\leqslant \frac{\nu}{q^2(q^2-\nu)} &\text{(B41)}\\
&= \frac{\nu}{q^4(1-\frac{\nu}{q^2})} &\text{(B42)}\\
&\leqslant \frac{\nu}{q^4}\left(8\left(\frac{\nu}{q^2}\right)^2 - 4\frac{\nu}{q^2} + 2\right). &\text{(B43)}
\end{aligned}$$

Then, one can check that $\nu = 1/2(\varepsilon q^2)/\|O\|$ gives $\left|1/Q^{(M_q)} - 1/q^2\right| \leqslant \varepsilon^2/q^2\|O\|^2(\varepsilon/\|O\| - 1) + \varepsilon/q^2\|O\| \leqslant \varepsilon/q^2\|O\|$ and so, returning to Eq. (B39), under this condition we have $\left|1/Q^{(M_q)}\bar{z}^{(M)} - 1/q^2\langle\phi|f(A)Of(A)^\dagger|\phi\rangle\right| \leqslant 3\varepsilon$ for $\varepsilon \leqslant 1$, $\|O\| \geqslant 1$. ■

By similar reasoning, we also have the following lemma for quantities of the form of a state overlap, rather than expectation values of general measurement observables.

Lemma 4 (Normalization precision, part 2). Suppose that we have some $\bar{z}^{(M)} \in \mathbb{R}$, pure states $|\phi\rangle$ $|\psi\rangle$, and an operator $f(A)$ that satisfy

$$\left|\frac{1}{q}\bar{z}^{(M)} - \frac{1}{q}\langle\phi|f(A)|\psi\rangle\right| \leqslant \varepsilon \leqslant 1, \tag{B44}$$

where we denote $q = \|f(A)|\phi\rangle\|_2$. Then, if $Q^{(M_q)}$ is the statistical approximation of q^2, one can achieve

$$\left|\frac{1}{\sqrt{Q^{(M_q)}}}\bar{z}^{(M)} - \frac{1}{q}\langle\phi|f(A)|\psi\rangle\right| \leqslant 3\varepsilon, \tag{B45}$$

if $Q^{(M_q)}$ satisfies $|Q^{(M_q)} - q^2| \leqslant \frac{1}{2}\varepsilon q^2$.

Proof. Similar to the proof of Lemma 3, we have

$$\begin{aligned}
&\left|\frac{1}{\sqrt{Q^{(M_q)}}}\bar{z}^{(M)} - \frac{1}{q}\langle\phi|f(A)|\psi\rangle\right| \\
&\quad\leqslant \left|\frac{1}{\sqrt{Q^{(M_q)}}} - \frac{1}{q}\right|q(1+\varepsilon) + \varepsilon. &\text{(B46)}
\end{aligned}$$

Moreover, if $|Q^{(M_q)} - q^2| \leqslant \nu$, then

$$\left|\sqrt{Q^{(M_q)}} - q\right| \leqslant \frac{\nu}{\sqrt{Q^{(M_q)}} + q} \leqslant \frac{\nu}{q}. \tag{B47}$$

Thus,

$$\left|\frac{1}{\sqrt{Q^{(M_q)}}} - \frac{1}{q}\right| \leqslant \frac{\left|\sqrt{Q^{(M_q)}} - q\right|}{\sqrt{Q^{(M_q)}} q} \tag{B48}$$

$$\leqslant \frac{\nu}{\sqrt{Q^{(M_q)}} q^2} \tag{B49}$$

$$\leqslant \frac{\nu}{q^2(q - \nu/q)} \tag{B50}$$

$$\leqslant \frac{\nu}{q^3}\left(8\left(\frac{\nu}{q^2}\right)^2 - 4\frac{\nu}{q^2} + 2\right), \tag{B51}$$

and the choice $\nu = \frac{1}{2}\varepsilon q^2$ gives

$$\left|\frac{1}{\sqrt{Q^{(M_q)}}} - \frac{1}{q}\right| \leqslant \frac{\varepsilon}{q}, \tag{B52}$$

which leads to the desired result, under the assumption $\varepsilon \leqslant 1$. ■

We will also need the following lemma, which gives a tighter result than Lemma 2 for pure states.

Lemma 5 (Tightness of expectation values, part 2). Given two operators C and D, where C is Hermitian, and given a pure state $|\phi\rangle$, we have

$$\left|\langle\phi|D^\dagger OD|\phi\rangle - \langle\phi|COC|\phi\rangle\right| \leqslant 3\,\||C|\phi\rangle\|_2\,\|C - D\|_\infty \|O\|_\infty, \tag{B53}$$

where we have assumed that $\|C - D\|_\infty \leqslant \||C|\phi\rangle\|_2$.

Proof. We directly bound the difference in expectation values as

$$\left|\langle\phi|D^\dagger OD|\phi\rangle - \langle\phi|COC|\phi\rangle\right| = \Big|\langle\phi|CO(D - C)|\phi\rangle + \langle\phi|(D^\dagger - C)OC|\phi\rangle + \langle\phi|(D^\dagger - C)O(D - C)|\phi\rangle\Big| \tag{B54}$$

$$\leqslant \||\phi\rangle\langle\phi|\,C\|_1\,\|O(D - C)\|_\infty + \|C\,|\phi\rangle\langle\phi|\|_1\,\|(D^\dagger - C) \times O\|_\infty + \||\phi\rangle\langle\phi|\|_1\,\|(D^\dagger - C)O(D - C)\|_\infty \tag{B55}$$

$$\leqslant \||\phi\rangle\langle\phi|\|_2\,\||\phi\rangle\langle\phi|\,C\|_2\,\|O\|_\infty\|D - C\|_\infty + \|C\,|\phi\rangle\langle\phi|\|_2\,\||\phi\rangle\langle\phi|\|_2\,\|D^\dagger - C\|_\infty\|O\|_\infty + \|D^\dagger - C\|_\infty\|O\|_\infty\|D - C\|_\infty \tag{B56}$$

$$= 2\||C|\phi\rangle\|_2\|C - D\|_\infty\|O\|_\infty + \|C - D\|_\infty^2\|O\|_\infty \tag{B57}$$

$$\leqslant 3\||C|\phi\rangle\|_2\|C - D\|_\infty\|O\|_\infty, \tag{B58}$$

where in the first equality we have added and subtracted $\langle\phi|COD|\phi\rangle + \langle\phi|D^\dagger OC|\phi\rangle + \langle\phi|COC|\phi\rangle$, in the first inequality we have used the triangle inequality and Hölder's tracial-matrix inequality, in the second inequality we have used the Cauchy-Schwarz inequality and the submultiplicativity of the operator norm, in the subsequent equality we have used the definition of the vector 2-norm and the fact that C is Hermitian, and in the final inequality we have used our starting assumption that $\|C - D\|_\infty \leqslant \||C|\phi\rangle\|_2$. ■

Lemmas 3 and 4 demonstrate that if one has a normalization factor that also needs to be statistically approximated, then one can simply modify the error parameter in Theorem 2 by a factor of 3 and use a separate algorithm to approximate the normalization factor. In the case of the state-normalization factor $q = \|f(A)|\psi\rangle\|$, we now demonstrate the complexities required for a randomized quantum algorithm that approximates it to sufficient error as specified by Lemmas 3, 4 and 5 (see Proposition 2 of the main text). Following previous sections, we denote our matrix function of interest as $f(A)$ given matrix A, and denote its Fourier-series approximation as $s(\varepsilon, A)$.

Proposition 7 (Sampling normalization constant—detailed version). If $q = \|f(A)|\psi\rangle\|$ where $|\psi\rangle$ is a preparable input state in depth d_ψ with unitary U_ψ, then we give a randomized algorithm to approximate q for the two algorithms in Theorem 2 that has success probability at least

$(1-\delta)$ and complexity

$$\mathcal{C}_{sample} = \mathcal{O}\left(\log\left(\frac{2}{\delta}\right)\frac{c^2\left[\alpha\left(\frac{\varepsilon q}{12c}, A\right)\right]^4}{\varepsilon^2 q^4}\right),$$
$$\mathcal{C}_{gate} = \mathcal{O}\left(\lambda^2\left[t_{max}\left(\frac{\varepsilon q}{12c}, A\right)\right]^2 + d_\psi\right), \tag{B59}$$

where $c = 1$ for the algorithm that prepares $1/q\langle\psi|Uf(A)V|\psi\rangle$ and $c = \|O\|$ for the algorithm that prepares $(1/q^2)\mathrm{Tr}\left[f(A)|\psi\rangle\langle\psi|f(A)^\dagger O\right]$.

Proof of Proposition 7. Lemmas 3 and 4 specify that, in order to statistically approximate the normalized expectation value to additive error ε, one requires a statistical approximation to q^2 using M_q shots, which we denote as $Q^{(M_q)}$, that satisfies $|Q^{(M_q)} - q^2| \leqslant 1/(6)\varepsilon q^2$ for problem (a) and satisfies $|Q^{(M_q)} - q^2| \leqslant 1/6(\varepsilon q^2)/\|O\|$ for problem (b). From hereon, we thus deal with problem (b) and note that the resources required for problem (a) can be accounted for by setting $\|O\| \to 1$.

We propose to construct $Q^{(M_q)}$ by sampling $\langle\phi|s(\varepsilon, A)^\dagger s(\varepsilon, A)|\phi\rangle$ via Algorithm 1 and the Hadamard test circuit in Fig. 2(a). We require

$$\left|Q^{(M_q)} - \langle\phi|f(A)^2|\phi\rangle\right| \leqslant \left|Q^{(M_q)} - \langle\phi|s(\varepsilon, A)^\dagger s(\varepsilon, A)|\phi\rangle\right| + \left|\langle\phi|s(\varepsilon, A)^\dagger s(\varepsilon, A)|\phi\rangle - \langle\phi|f(A)^2|\phi\rangle\right| \tag{B60}$$

to be smaller than $1/6(\varepsilon q^2)/\|O\|$. The second term on the right-hand side is bounded by Lemma 5 as

$$\left|\langle\phi|s(\varepsilon, A)^\dagger s(\varepsilon, A)|\phi\rangle - \langle\phi|f(A)^2|\phi\rangle\right| \leq 3q\left\|s(\varepsilon, A) - f(A)\right\|. \tag{B61}$$

Thus, evenly distributing error across statistical and exact contributions, we require a Fourier-series approximation that is tight up to $\|s(\varepsilon, A) - f(A)\| \leqslant 1/36(\varepsilon q)/\|O\|$, which sets the second term in Eq. (B60) to be smaller than $1/12(\varepsilon q^2)/\|O\|$. Using Proposition 1, we can bound the first term in Eq. (B60) by the value $1/12(\varepsilon q^2)/\|O\|$ by using

$$M_q \geqslant 2e^4 \log\left(\frac{2}{\delta}\right)\frac{\alpha\left(\frac{1}{36}\frac{\varepsilon q}{\|O\|}\right)^2}{\left(\frac{1}{12}\frac{\varepsilon q^2}{\|O\|}\right)^2} = 288e^4 \log\left(\frac{2}{\delta}\right)\frac{\|O\|^2\alpha\left(\frac{1}{12}\frac{\varepsilon q}{\|O\|}\right)^2}{\varepsilon^2 q^4}$$

shots and gate depth $\mathcal{C}_{\text{gate}} = 2\lambda^2 t_{\max}^2(1/12(\varepsilon q)/\|O\|)$. ■

5. Statistical encoding of classical vectors

In this section, we show how to statistically "encode" a classical vector $\vec{b} = (b_1, \ldots, b_N)$ to recover its properties when determining state overlaps and expectation values using our randomized algorithms. We show that with $\mathcal{O}(s)$ classical preprocessing steps, one can start from classical access to a vector $\vec{b}$ and invoke it as part of a larger randomized algorithm with a cost of increased circuit samples depending on $\|\vec{b}\|_1$ and minimal quantum gate overhead. This can then replace the usual (quantum) state oracle in linear algebra algorithms, such as in our statistical algorithm for linear systems as described in Proposition 4 in the main text. We note that, traditionally, quantum algorithms that act on classical data assume that an encoding is given via a normalized input vector of the form $|\vec{b}\rangle := 1/\|\vec{b}\|_2 \sum_i b_i|i\rangle$. However, our compilation scheme allows for arbitrary normalization; thus we can recover the action of the true vector $\vec{b}$ and take its magnitude into account.

Definition 1 (Classical sparse-access model). We say that a classical vector $\vec{b} = (b_1, \ldots, b_N)$ is stored with sparse access if the following set of tuples is stored and accessible in classical memory:

$$\mathcal{B} = \{(b_i, i) \mid b_i \neq 0\}. \tag{B62}$$

We suppose that we wish to randomly compile properties of a matrix G decomposed into a linear combination of implementable unitaries $G = \sum_j g_j V_j$ with weight of coefficients $g := \sum_j |g_j|$. Using our randomized sampling scheme with the circuits in Figs. 2(a) and 2(b), we can prepare quantities of the form $\langle\psi|G|\vec{b}\rangle$ and $\langle\vec{b}|G^\dagger OG|\vec{b}\rangle$, with respective sample complexities $\mathcal{O}(g^2/\varepsilon^2)$ and $\mathcal{O}(g^4\|O\|^2/\varepsilon^2)$ and a gate depth scaling with the largest gate depth in $\{V_j\}_j$. This presumes that there is a procedure to prepare the states $|\psi\rangle$, $|\vec{b}\rangle$. If $|\vec{b}\rangle$ is not provided via an oracle, the following proposition shows how to statistically recover the encoding and how the sample complexities change.

Proposition 8 (Random compiling classical input vector—detailed version). Suppose that a vector $\vec{b} = (b_1, \ldots, b_N)$ with sparsity s is stored classically with sparse access as defined in Definition 1. Suppose further that we have access to some measurement observable O and a unitary U_ψ to prepare state $|\psi\rangle$. Then, by providing classical preprocessing in $\mathcal{O}(s)$ time, there exists a randomized quantum algorithm that returns an approximation of (a) $\langle\psi|G|\vec{b}\rangle$, and (b) $\langle\vec{b}|G^\dagger OG|\vec{b}\rangle$ to additive error ε and constant arbitrary probability, using

$$\mathcal{C}^{\psi}_{sample} = \mathcal{O}\left(\frac{g^2}{\varepsilon^2}\frac{\|\vec{b}\|_1^2}{\|\vec{b}\|_2^2}\right), \quad \mathcal{C}^{O}_{sample} = \mathcal{O}\left(\frac{g^4\|O\|^2}{\varepsilon^2}\frac{\|\vec{b}\|_1^4}{\|\vec{b}\|_2^4}\right), \tag{B63}$$

for scenarios (a) and (b), respectively, where each circuit uses one application of U_ϕ. The additional gate overhead

for randomly compiling the input vector is one or two controlled multiqubit NOT gates.

Proof of Proposition 8. We first give an overview of the main steps of our idea, before giving more precise analysis of quantum and classical resources required. We express $|\vec{b}\rangle = \sum_{i=1}^{N} b_i|i\rangle/\|\vec{b}\|_2$ as a weighted probabilistic sum over states

$$|\vec{b}\rangle = \frac{\|\vec{b}\|_1}{\|\vec{b}\|_2}\sum_{i=1}^{N} p_i^{(\vec{b})}\cdot \mathrm{sgn}(b_i)|i\rangle = \frac{\|\vec{b}\|_1}{\|\vec{b}\|_2}\sum_{i=1}^{N} p_i\cdot \mathrm{sgn}(b_i)X_i|0\rangle, \tag{B64}$$

where we denote the probabilities $p_i^{(\vec{b})} = |b_i|/\|\vec{b}\|_1$, we denote $\mathrm{sgn}(b_i) = b_i/|b_i|$, and X_i is the $\lceil \log N\rceil$-qubit Pauli operator in $\{\mathbb{1}, X\}^{\otimes \lceil \log N\rceil}$ that corresponds to the binary representation of i. If the description in Eq. (B64) is classically accessible, then $|\vec{b}\rangle$ can be statistically encoded by sampling from this distribution as follows.

(a) *Recovering* $\langle\psi|G|\vec{b}\rangle$. We can express $\langle\psi|G|\vec{b}\rangle$ as the probabilistic sum

$$\langle\psi|G|\vec{b}\rangle = \frac{\|\vec{b}\|_1}{\|\vec{b}\|_2}\sum_{i=1}^{N} p_i^{(\vec{b})}\cdot \mathrm{sgn}(b_i)\langle 0|U_\psi^\dagger G X_i|0\rangle = \frac{\|\vec{b}\|_1}{\|\vec{b}\|_2} g\sum_{i=1}^{N} p_i^{(\vec{b})}p_j^{(G)}\cdot \mathrm{sgn}(b_i g_j)\langle 0|U_\psi^\dagger V_j X_i|0\rangle, \tag{B65}$$

where in the first equality we have used the fact that $|\psi\rangle = U_\psi|0\rangle$ and Eq. (B64), and in the second equality we have decomposed G into its constituent unitaries as $G = \sum_j g_j V_j$ with weight of coefficients $g := \sum_j |g_j|$. The quantity $\langle\psi|G|\vec{b}\rangle$ can then be statistically recovered as follows. (1) Sample indices i' and j' from the probability distribution $\{p_i^{(\vec{b})}p_i^{(G)}\}_{i,j}$. (2) Run two Hadamard test circuits [see Fig. 2(a)] to obtain one measurement sample each of $\mathrm{Re}(\langle 0|U_\psi^\dagger V_{j'}X_{i'}|0\rangle)$ and $\mathrm{Im}(\langle 0|U_\psi^\dagger V_{j'}X_{i'}|0\rangle)$, respectively. (3) Classically multiply the result by $\|\vec{b}\|_1 g\,\mathrm{sgn}(b_{i'}g_{j'})/\|\vec{b}\|_2$ and store the result. (4) Repeat the process $\mathcal{C}^{\phi}_{\mathrm{sample}}$ times and average over results. Due to Hoeffding's inequality, it is sufficient to take

$$\mathcal{C}^{\phi}_{\mathrm{sample}} = 8\log\left(\frac{2}{\delta}\right)\frac{g^2}{\varepsilon^2}\frac{\|\vec{b}\|_1}{\|\vec{b}\|_2} \tag{B66}$$

samples to attain an answer within additive error ε and probability at least $(1-\delta)$. (The additional factor of 4 comes from approximating $\mathrm{Re}(\langle 0|U_\psi^\dagger V_j X_i|0\rangle)$ and $\mathrm{Im}(\langle 0|U_\psi^\dagger V_j X_i|0\rangle)$ each to additive error $\varepsilon/\sqrt{2}$ separately.)

(b) *Recovering* $\langle\vec{b}|G^\dagger OG|\vec{b}\rangle$. Similar to the above, we can write

$$\langle\vec{b}|G^\dagger OG|\vec{b}\rangle = \frac{\|\vec{b}\|_1^2}{\|\vec{b}\|_2^2}\sum_{i,\ell} p_i^{(\vec{b})}p_\ell^{(\vec{b})}\cdot \mathrm{sgn}(b_i b_\ell)\langle 0|X_i G^\dagger OG X_\ell|0\rangle. \tag{B67}$$

$$= \frac{\|\vec{b}\|_1^2}{\|\vec{b}\|_2^2} g^2\sum_{i,j,k,\ell} p_i^{(\vec{b})}p_j^{(G)}p_k^{(G)}p_\ell^{(\vec{b})}\cdot \mathrm{sgn}(b_i g_j g_k b_\ell)\langle 0|X_i V_j^\dagger O V_k X_\ell|0\rangle. \tag{B68}$$

We consider a very similar protocol to before. (1) Sample indices i', j', k', ℓ' from the probability distribution $\{p_i^{(\vec{b})}p_j^{(G)}p_k^{(G)}p_\ell^{(\vec{b})}\}_{ijk\ell}$. (2) Run the circuit in Fig. 2 to obtain one measurement sample of $\langle 0|X_i V_j^\dagger O V_k X_\ell|0\rangle$. (3) Classically multiply the result by $\|\vec{b}\|_1^2 g^2\,\mathrm{sgn}(b_i g_j g_k b_\ell))/\|\vec{b}\|_2^2$ and store the result. (4) Repeat the process $\mathcal{C}^{O}_{sample}$ times and average over results. Due to Hoeffding's inequality, it is sufficient to take

$$\mathcal{C}^{O}_{\mathrm{sample}} = 2\log\left(\frac{2}{\delta}\right)\frac{g^4\|O\|^2}{\varepsilon^2}\frac{\|\vec{b}\|_1^4}{\|\vec{b}\|_2^4} \tag{B69}$$

shots to attain an answer within additive error ε and probability at least $(1-\delta)$.

We now consider the classical resources required to determine $\|\vec{b}\|_1$, $\|\vec{b}\|_2$, $\{p_i\}$, and $\{\mathrm{sgn}(b_i)\}$ given $\mathcal{B}$. As one can obtain the set of tuples $\overline{\mathcal{B}} = \{(|b_i|, i)\,|\,b_i \neq 0\}$ in $\mathcal{O}(s)$ steps from $\mathcal{B}$, it then follows that all four sets of quantities can be obtained in $\mathcal{O}(s)$ additional steps by first obtaining $\overline{\mathcal{B}}$ and then combining quantities in $\mathcal{B}$ and $\overline{\mathcal{B}}$. ■

Remark (Arbitrary normalization). We can additionally consider arbitrarily normalized states with normalization constant m as $1/m|\vec{b}\rangle$ by changing the sample complexity by factor $\|\vec{b}\|_1 \to 1/m\|\vec{b}\|_1$.

6. Sampling from output vector

In this section, we demonstrate a modification of Theorem 2 that allows sampling from the vector corresponding to sampling $G|\psi\rangle$ in the computational basis for some state $|\psi\rangle$ and operator G of interest, which is a generalization of our algorithmic framework. As in the previous section, we suppose that there is a known decomposition $G = \sum_i g_i V_i$ into implementable unitaries with $g_i \in \mathbb{C}$. For further simplicity, we denote $\sum_i g_i V_i = g\sum_i p_i U_i$, where $g := \sum_j |g_j|, p_i := |g_i|/g$ and $U_i := (g_i/|g_i|)V_i$ is a unitary that absorbs the phase of g_i. Our discussion will remain general to any randomized algorithm that samples from

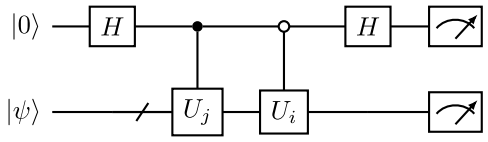

FIG. 3. The generalized Hadamard test. We use this circuit along with a modified measurement in Algorithm 2 to sample from the solution vector. Rather than measuring the observable $Z \otimes O$, we measure all qubits in the computational basis.

such a collection of unitaries. In what follows, we show that an unbiased estimator can be constructed for the vector the entries of which take the values $|\langle \vec{z}_n| G |\psi\rangle|^2$ for each $\vec{z}_n \in \{0,1\}^n$. This mimics the more common approach in quantum algorithms where one prepares the quantum state $G|\psi\rangle / \|G|\psi\rangle\|$ and collects measurement samples in the computational basis. As with Theorem 2, our scheme moves part of the (coherent) quantum complexity into sample complexity, as now the operator G does not need to be materialized quantumly.

In order to produce our desired estimator, we will simply implement the gates in the circuit corresponding to the generalized Hadamard test of Fig. 2(b) and measure all qubits in the computational basis, with minor classical postprocessing. Consider the circuit in Fig. 3, which explicitly shows the gates in the circuit in Fig. 2(b) before computational-basis measurement.

The state before measurement is $1/2\big(|0\rangle (U_i + U_j)|\psi\rangle + |1\rangle (U_i - U_j)|\psi\rangle\big)$. Measurement in the computational basis yields

$$(0, \vec{z}_n) \quad \text{with probability} \quad \frac{1}{4}\left|\langle \vec{z}_n| (U_i + U_j)|\psi\rangle\right|^2, \tag{B70}$$

$$(1, \vec{z}_n) \quad \text{with probability} \quad \frac{1}{4}\left|\langle \vec{z}_n| (U_i - U_j)|\psi\rangle\right|^2, \tag{B71}$$

for all $\vec{z}_n \in \{0,1\}^n$.

Our procedure is as follows. As in Algorithm 2, we sample two unitaries (U_i, U_j) independently at a time with probability $p_i p_j$. We run the circuit in Fig. 3, concluding with computational-basis measurement on all qubits, and assign vector element $g^2(-1)^z |\vec{z}_n\rangle$ upon receiving string $(z, \vec{z}_n)$. One can check that this gives an unbiased estimator for the vector

$$\sum_{ij} p_i p_j \sum_{\vec{z}_n \in (0,1)^n} \sum_{z \in \{0,1\}} \text{Prob}\big((z, \vec{z}_n)\big|U_i, U_j\big) \cdot g^2 (-1)^z |\vec{z}_n\rangle = \tag{B72}$$

$$= \sum_{ij} p_i p_j \sum_{\vec{z}_n \in (0,1)^n} \left(\frac{1}{4}\left|\langle \vec{z}_n| (U_i + U_j)|\psi\rangle\right|^2 \cdot (g^2 |\vec{z}_n\rangle) + \frac{1}{4}\left|\langle \vec{z}_n| (U_i - U_j)|\psi\rangle\right|^2 \cdot (-1 g^2 |\vec{z}_n\rangle)\right) \tag{B73}$$

$$= \sum_{\vec{z}_n \in (0,1)^n} \sum_{ij} g^2 p_i p_j \left(\frac{1}{2}\langle\psi| U_i^\dagger |\vec{z}_n\rangle \langle \vec{z}_n| U_j |\psi\rangle \cdot |\vec{z}_n\rangle + \frac{1}{2}\langle\psi| U_j^\dagger |\vec{z}_n\rangle \langle \vec{z}_n| U_i |\psi\rangle \cdot |\vec{z}_n\rangle\right) \tag{B74}$$

$$= \sum_{\vec{z}_n \in (0,1)^n} \left|\langle \vec{z}_n| G |\psi\rangle\right|^2 \cdot |\vec{z}_n\rangle, \tag{B75}$$

which indeed is the vector with entries $\sum_{\vec{z}_n \in (0,1)^n} \left|\langle \vec{z}_n| G |\psi\rangle\right|^2$ as desired, which, from hereon, we denote as $\vec{G}$.

How many shots do we need to take in order to obtain good convergence? We appeal to vector Bernstein inequalities. Reference [104] essentially gives a vector Bernstein inequality (adapted from Refs. [105,106]) for the sample mean $\vec{Y}_M = \sum_{i=1}^M \vec{X}_i$ of M i.d.d. random variables $\vec{X}_i \in \mathbb{R}^N$ of

$$\text{Prob}\left(\left\|\vec{Y}_M - \mathbb{E}[\vec{Y}_M]\right\| \geqslant \varepsilon\right) \leqslant \exp\left(-\frac{M\varepsilon^2}{8\sigma^2}\right), \tag{B76}$$

where σ^2 is an upper bound on the second moment satisfying $\sigma^2 \geqslant \mathbb{E}\big[\|\vec{X}_i - \mathbb{E}[\vec{X}_i]\|_2^2\big]$ for all $i \in [N]$, and $\mathbb{E}[\vec{Y}_M] = \vec{G}$. Due to the linearity of the trace, we have $\mathbb{E}\big[\|\vec{X}_i - \mathbb{E}[\vec{X}_i]\|_2^2\big] \leqslant \mathbb{E}\big[\|\vec{X}_i\|_2^2\big]$ and thus we can set $\sigma^2 = g^4$. This implies that we obtain an estimator $\vec{Y}_{\mathcal{C}_{\text{sample}}}$ satisfying

$$\left\|\vec{Y}_{\mathcal{C}_{\text{sample}}} - \vec{G}\right\| \leqslant \varepsilon, \tag{B77}$$

with probability $(1 - \delta)$, using a number of shots satisfying

$$\mathcal{C}_{\text{sample}} = \mathcal{O}\left(\frac{g^4}{\varepsilon^2} \log\left(\frac{1}{\delta}\right)\right), \tag{B78}$$

which is the same sample complexity as specified in Theorem 2 to recover a general observable. Thus, here we have shown that instead of measuring an observable, one can use the same number of shots (asymptotically) to recover an approximation of the classical vector corresponding to output probabilities in the computational basis.

7. Classical algorithm for matrix polynomials

In this section, we show that, given a bounded degree polynomial series $h(A) = \sum_{k\in F_d} \alpha_k A^k$ in matrix A where A has known Pauli decomposition with bounded Pauli weight, one can then approximate quantities of the form $\langle t|h(A)|s\rangle$ efficiently classically by means of a randomized algorithm. This is a more detailed restatement of Proposition 3 in the main text.

Proposition 9 (Sampling from polynomial series—detailed version). Suppose that we have a polynomial series of degree d

$$h(A) = \sum_{k\in F_d} \alpha_k A^k, \tag{B79}$$

in some $N \times N$ Hermitian matrix A, where $F_d \subseteq [d]$, and denote the ℓ_1-norm of the coefficients as $\alpha := \sum_{k\in F_d} |\alpha_k|$. Suppose further that A has known Pauli decomposition $A = \sum_\ell a_\ell P_\ell$ with Pauli weight $\lambda = \sum_\ell |a_\ell|$. Then, given pure states $|t\rangle$ and $|s\rangle$ that are implementable by initializing in the $|0\rangle$ state and performing $\mathcal{O}(m)$ Clifford gates, one can classically approximate $\langle t|h(A)|s\rangle$ up to additive error ε with probability at least $1-\delta$, using

$$\mathcal{C}_{sample} = \mathcal{O}\left(\log\left(\frac{2}{\delta}\right)\frac{(\alpha\lambda^d)^2}{\varepsilon^2}\right) \tag{B80}$$

calls to independent classical subroutines, each requiring $\mathcal{O}(\log^2(N))$ bits and at most $\mathcal{O}(d\log^2(N) + m\log(N))$ time.

Proof of Proposition 9. We first note that we can write

$$A = \lambda \sum_\ell p_\ell \cdot \mathrm{sgn}(a_\ell) P_\ell, \tag{B81}$$

where $\lambda = \sum_\ell |a_\ell|$ is the Pauli weight and we denote $p_\ell = |a_\ell|/\lambda$ and $\mathrm{sgn}(a_\ell) = a_\ell/|a_\ell|$. We can then write

$$\begin{aligned} h(A) &= \sum_{k\in F_d} \alpha_k \lambda^k \sum_{\ell_1,\ldots,\ell_k} p_{\ell_1}\cdots p_{\ell_k}\,\mathrm{sgn}(a_{\ell_1})\cdots\mathrm{sgn}(a_{\ell_k}) \\ &\quad\times P_{\ell_1}\cdots P_{\ell_k} \qquad\qquad (B82)\\ &= R_{h(A)} \sum_{k\in F_d}\sum_{\ell_1,\ldots,\ell_k} q_k p_{\ell_1}\ldots p_{\ell_k}\,\phi(\alpha_k) \\ &\quad\times \mathrm{sgn}(a_{\ell_1})\cdots\mathrm{sgn}(a_{\ell_k}) P_{\ell_1}\cdots P_{\ell_k}, \qquad (B83)\end{aligned}$$

where we denote $R_{h(A)} = \sum_{k\in F_d} |\alpha_k\lambda^k| \leqslant \alpha\lambda^d$, $q_k = |\alpha_k|\lambda^k/R_{h(A)}$, and $\phi(\alpha_k) = \alpha_k/|\alpha_k|$. We note that $p_{\ell_1},\ldots,p_{\ell_k}$ and q_k are probabilities. This then allows us to statistically recover $h(A)$ with weight $R_{h(A)}$. For simplicity, we subsume all indices and write

$$h(A) = G\sum_{i\in S} \tilde{p}_i\tilde{\phi}_i\widetilde{P}_i\,, \tag{B84}$$

where $\tilde{p}_i = q_k p_{\ell_1}\ldots p_{\ell_k}$ subsumes all probabilities, $\tilde{\phi}_i = \phi(\alpha_k)\,\mathrm{sgn}(a_{\ell_1})\cdots\mathrm{sgn}(a_{\ell_k})$ subsumes all phases, and $\widetilde{P}_i = P_{\ell_1}\cdots P_{\ell_k}$ is a product of k Paulis.

Preparation of $\langle t|h(A)|s\rangle$. Denote the collection of Cliffords to prepare $|t\rangle$ and $|s\rangle$ as U_t and U_s, respectively. Using Eq. (B84), we can simply write

$$\langle t|h(A)|s\rangle = R_{h(A)}\sum_{i\in S}\tilde{p}_i\tilde{\phi}_i\langle 0|U_t^\dagger\widetilde{P}_i U_s|0\rangle. \tag{B85}$$

Inspecting this, we can consider the following protocol: (1) sample from the above probability distribution, obtaining index i with probability $\tilde{p}_i$; (2) sample one measurement result corresponding to $\langle 0|U_t^\dagger\widetilde{P}_i U_s|0\rangle$ by running the circuit in Fig. 2(a); (3) multiply the result by $R_{h(A)}\tilde{\phi}_i$; (4) repeat M times and take the mean over the results.

We note that in step (2), one requires sampling one measurement result from a circuit that contains at most $2d$ controlled n-qubit Pauli operations plus $\mathcal{O}(m)$ other Clifford operations. The former part can simply be decomposed into at most $2dn$ controlled single-qubit Pauli operations. In order to simulate the circuit overall, one requires $\mathcal{O}(n^2)$ bits and $\mathcal{O}(dn^2 + mn)$ time [107,108]. Due to Hoeffding's inequality, in order to recover $\langle t|h(A)|s\rangle$ to additive error ε with probability at least $(1-\delta)$, it is sufficient to use M shots for any

$$M \geqslant 2\log\left(\frac{2}{\delta}\right)\frac{R_{h(A)}^2}{\varepsilon^2}. \tag{B86}$$

Using the upper bound $R_{h(A)} \leqslant \alpha\lambda^d$ we observe that it is sufficient to perform

$$\mathcal{C}_{\mathrm{sample}} = 2\log\left(\frac{2}{\delta}\right)\frac{(\alpha\lambda^d)^2}{\varepsilon^2} \tag{B87}$$

shots.

Note that quantities of the form $\mathrm{Tr}\left[h(A)\rho\,h(A)^\dagger O\right]$ (where ρ and O are implementable by Cliffords) can be obtained by taking the modulus squared of quantities of the form $\langle t|h(A)|s\rangle$. Alternatively, they can also be prepared by simulating the circuit in Fig. 2(b). ■

The above proposition can be trivially extended to consider polynomials of multiple matrices. This implies that subroutines such as matrix multiplication can be efficiently sampled from, given the matrices have bounded Pauli weights.

APPENDIX C: APPLICATION-SPECIFIC RESULTS

1. Linear systems

In this appendix, we detail our results for linear systems. We first summarize a result from Ref. [47]. In this work, the authors find an approximation of A^{-1} as a linear combination of unitaries $\sum_i \alpha_i \exp(-iAz_j)$. More precisely, they find a Fourier representation of the inverse function as

$$\frac{1}{x} = \frac{i}{\sqrt{2\pi}} \int_0^\infty \mathrm{d}y \int_{-\infty}^\infty \mathrm{d}z\, z e^{-z^2/2} e^{-ixyz}. \tag{C1}$$

Further, they show that when truncating the integration range as

$$g(x) := \frac{i}{\sqrt{2\pi}} \int_0^{y_J} \mathrm{d}y \int_{-z_K}^{z_K} \mathrm{d}z\, z e^{-z^2/2} e^{-ixyz}, \tag{C2}$$

we have

$$\left| g(x) - \frac{1}{x} \right| \leqslant \frac{1}{|x|} e^{-(xy_J)^2/2} + \frac{2}{|x|} e^{-z_K^2/2} \leqslant \varepsilon, \tag{C3}$$

on the domain $D_b := [-1, -1/b] \cup [1/b, 1]$ for some $y_J(\varepsilon, b) = \Theta(b\sqrt{\log(b/\varepsilon)})$ and $z_K(\varepsilon, b) = \Theta(\sqrt{\log(b/\varepsilon)})$ (Ref. [47, Lemma 12]). An important point is that this choice is totally independent of the upper limit of the domain and is wholly dependent on b, i.e., the same statement holds over the domain $[-a, -1/b] \cup [1/b, a]$ with arbitrary $a \geqslant 1/b$. This is due to the fact that the upper bound on the approximation error in Eq. (C3) is a decreasing function in $|x|$. The integral in Eq. (C2) can be discretized as

$$h(x) := \frac{i}{\sqrt{2\pi}} \sum_{j=0}^{J-1} \Delta_y \sum_{k=-K}^{K} \Delta_z z_k e^{-z_k^2/2} e^{-ixy_j z_k}, \tag{C4}$$

where the integration range has been discretized into J and $2K+1$ steps of size Δ_y and Δ_z, respectively. Moreover, taking step sizes $\Delta_y = \Theta(\varepsilon/\sqrt{\log(b/\varepsilon)})$ and $\Delta_z = \Theta((b\sqrt{\log(b/\varepsilon)})^{-1})$ guarantees that $h(x)$ is ε-close to $1/x$ on the domain D_b. Note, however, as we will only sample from this distribution, for our purposes the resolution of this discretization can be taken to be arbitrarily small. By inspecting Eq. (C4), one can observe that this corresponds to a Fourier series with maximum time parameter

$$t_{\max} = \Theta\left(y_J(\varepsilon, b) z_K(\varepsilon, b)\right) = \Theta\left(b \log(b/\varepsilon)\right) \tag{C5}$$

and coefficients with weight

$$\frac{1}{\sqrt{2\pi}} \sum_{j=0}^{J-1} \Delta_y \sum_{k=-K}^{K} \Delta_z |z_k| e^{-z_k^2/2} = \Theta\left(y_J(\varepsilon, b)\right) = \Theta\left(b\sqrt{\log(b/\varepsilon)}\right). \tag{C6}$$

By mapping the domain D_b to the spectrum of some matrix A, this allows us to establish the following lemma.

Lemma 6 (Fourier-series approximation of inverse operator, adapted from Ref. [47]). Given some matrix A with finite $\|A^{-1}\|$, we have

$$\left\| A^{-1} - \sum_{i \in S_{\varepsilon,A}} \alpha_i(\varepsilon, A) \exp(-iAz_j(\varepsilon, A)) \right\| \leqslant \varepsilon, \tag{C7}$$

where $S_{\varepsilon,A}$ is some index set, with maximum time parameter $t_{\max} := \max_{i \in S_{\varepsilon,A}}(z_j)$ satisfying

$$t_{\max}(\varepsilon, A) = \Theta\left(\|A^{-1}\| \log\left(\frac{\|A^{-1}\|}{\varepsilon} \right) \right), \tag{C8}$$

and Fourier coefficients $\alpha_i(\varepsilon, A)$ with ℓ_1-norm satisfying

$$\alpha(\varepsilon, A) := \sum_i |\alpha_i(\varepsilon, A)| = \Theta\left(\|A^{-1}\| \sqrt{\log\left(\frac{\|A^{-1}\|}{\varepsilon} \right)} \right). \tag{C9}$$

Our result for linear systems (Corollary 1 in the main text) then follows directly from Theorem 2 and Proposition 7 as follows.

Corollary 4 (Linear systems—detailed version). Consider a Hermitian matrix A with known Pauli decomposition $A = \sum_\ell a_\ell P_\ell$; $\lambda := \sum_\ell |a_\ell|$. Denote q as a freely chosen normalization parameter. Finally, suppose that we have the ability to prepare state $|\vec{b}\rangle$ in $\mathcal{O}(d_{\vec{b}})$ depth:

(i) Given the ability to implement $U_\psi|0\rangle = |\psi\rangle$ in gate depth d_ψ, there exists a randomized quantum algorithm that returns $1/q\langle\psi|A^{-1}|\vec{b}\rangle$ up to additive error ε with probability at least $1-\delta$, utilizing $\mathcal{O}\left(\log(2/\delta) \frac{\|A^{-1}\|^2}{\varepsilon^2 q^2} \log\left(\|A^{-1}\|^2/\varepsilon q^2\right)\right)$ circuit runs each with gate depth $\mathcal{O}\left(\|A^{-1}\|^2 \lambda^2 \log^2\left(\|A^{-1}\|^2/\varepsilon q^2\right) + d_\psi + d_{\vec{b}}\right)$.

(ii) Given the ability to measure observable O; $\|O\| \leqslant 1$, there exists a randomized quantum algorithm that returns $1/q^2\langle\vec{b}|A^{-1}OA^{-1}|\vec{b}\rangle$ up to additive error ε with probability at least $1-\delta$, utilizing $\mathcal{O}\left(\log(2/\delta) \frac{\|A^{-1}\|^4}{\varepsilon^2 q^4} \log^2\left(\|A^{-1}\|^2/\varepsilon q^2\right)\right)$ circuit runs

each of gate depth $\mathcal{O}\left(\|A^{-1}\|^2\lambda^2\log^2\left(\|A^{-1}\|^2/\varepsilon q^2\right)+d_{\vec{b}}\right)$.

(iii) In the case $q=\|A^{-1}|\vec{b}\rangle\|$, the value of which is not given, there exists an auxiliary algorithm that approximates the value of q for the above algorithms with the sample complexity $\mathcal{O}\left(\log(2/\delta)\frac{\|A^{-1}\|^4}{\varepsilon^2 q^4}\log^2\left(\|A^{-1}\|/\varepsilon q\right)\right)$ and gate depth $\mathcal{O}\left(\|A^{-1}\|^2\lambda^2\log^2\left(\|A^{-1}\|/\varepsilon q\right)+d_{\vec{b}}\right)$.

Proof of Corollary 1 (Linear systems). In order to quantify the complexity of preparing $1/q\langle\psi|Uf(A)V|\psi\rangle$, we now require $\alpha(\varepsilon q,A)$ and $t_{\max}(\varepsilon q,A)$, as specified by Eq. (B29) of Theorem 1. Again inspecting Eqs. (C8) and (C9) of Lemma 6, these quantities scale as

$$t_{\max}(\varepsilon q,A)=\Theta\left(\|A^{-1}\|\log\left(\frac{\|A^{-1}\|}{\varepsilon q}\right)\right), \tag{C10}$$

and

$$\alpha(\varepsilon q,A)=\Theta\left(\|A^{-1}\|\sqrt{\log\left(\frac{\|A^{-1}\|}{\varepsilon q}\right)}\right). \tag{C11}$$

Now, substituting the above into Eq. (B29) of Theorem 1, we obtain the desired result.

Similarly, in order to quantify the complexity of preparing $(1/q^2)\mathrm{Tr}\left[f(A)\rho f(A)^\dagger O\right]$, we need to evaluate the quantities $\alpha\left(\varepsilon q^2/\|f(A)\|,A\right)$ and $t_{\max}\left(\varepsilon q^2/\|f(A)\|,A\right)$, as specified by Eq. (B30) of Theorem 1. Inspecting Eqs. (C8) and (C9) of Lemma 6, we see that these quantities satisfy

$$t_{\max}\left(\frac{\varepsilon q^2}{\|f(A)\|},A\right)=\Theta\left(\|A^{-1}\|\log\left(\frac{\|A^{-1}\|^2}{\varepsilon q^2}\right)\right), \tag{C12}$$

and

$$\alpha\left(\frac{\varepsilon q^2}{\|f(A)\|},A\right)=\Theta\left(\|A^{-1}\|\sqrt{\log\left(\frac{\|A^{-1}\|^2}{\varepsilon q^2}\right)}\right). \tag{C13}$$

Substituting these two expressions into Eq. (B30) of Theorem 1, we obtain the desired result.

Finally, we can use Proposition 2 to characterize the complexity of approximating $\|A^{-1}|\vec{b}\rangle\|$. Equation (B59) of Proposition 2 is written in terms of $\alpha(\varepsilon q,A)$ and $t_{\max}(\varepsilon q,A)$, the scalings of which we have already quoted above. Substituting this into Eq. (B59), we obtain the result in Corollary 1. ■

Remark (Non-Hermitian matrices). Any non-Hermitian matrix B can be embedded in a larger Hermitian matrix A with the aid of a single qubit as

$$A=\begin{bmatrix}0 & B\\ B^\dagger & 0\end{bmatrix}. \tag{C14}$$

Then, one can verify that $A^{-1}\begin{pmatrix}\vec{b}\\0\end{pmatrix}=B^{-1}\vec{b}$, $\|A\|=\|B\|$ and $\|A^{-1}\|=\|B^{-1}\|$. In Pauli representation, given respective Hermitian and anti-Hermitian components $H(B)$ and $iH_2(B)$ of B such that $B=H_1(B)+iH_2(B)$, this embedding can be explicitly written as

$$\begin{aligned}A&=X\otimes H_1(B)-Y\otimes H_2(B)\\&=\frac{1}{2}X\otimes(B+B^\dagger)-\frac{1}{2}Y\otimes(B-B^\dagger).\end{aligned} \tag{C15}$$

From this, it is clear that the Pauli weight of A is bounded as $\lambda\leqslant 2\lambda_B$.

Using the above remark, we see that the asymptotic complexities for the linear-systems problem as stated in A can be simply be translated to B via the substitution $\lambda\to\lambda_B$ and $\|A^{-1}\|\to\|B^{-1}\|$.

2. Ground-state sampling

In this section, we study the task of sampling properties of the ground state of a given Hamiltonian. In order to establish our result, the relevant function that we consider is the Gaussian function $e^{-(1/2)\tau^2x^2}$. We refer to the following lemmas, which show how this can obtain approximations of ground-state observables.

Lemma 7 (Ground-state projection—adapted from Keen et al. [68]). Suppose that we have Hamiltonian $H=\sum_l E_l|E_l\rangle\langle E_l|$ with all eigenvalues $E_l\geqslant 0$. Assume that the spectral gap is lower bounded by $\Delta\leqslant E_1-E_0$. Additionally, we suppose that we have an initial trial state $|\psi_0\rangle$ with overlap with the ground state $\gamma:=|\langle\psi_0|E_0\rangle|$. Then, the state $|\psi\rangle=|\widetilde{\psi}\rangle/\||\widetilde{\psi}\rangle\|$, where $|\widetilde{\psi}\rangle=e^{-\frac{1}{2}\tau^2H^2}|\psi_0\rangle$, satisfies

$$1-|\langle\psi|E_0\rangle|\leqslant\frac{1}{2}\varepsilon^2, \tag{C16}$$

for any $\tau\geqslant\tau_\varepsilon$ where τ_ε satisfies

$$\tau_\varepsilon=\frac{1}{\Delta}\sqrt{2\log\frac{1}{\varepsilon\gamma}}. \tag{C17}$$

With the following lemma, we can see how this affects the closeness of expectation values.

Lemma 8 (Tightness of expectation values, part 3). For some operator O and pure states $|\psi\rangle$ and E_0, we have

$$\left|\mathrm{Tr}\left[|\psi\rangle\langle\psi|\,O\right]-\mathrm{Tr}\left[|E_0\rangle\langle E_0|\,O\right]\right| \leqslant 2\sqrt{2}\|O\|\sqrt{1-|\langle\psi|E_0\rangle|}. \tag{C18}$$

Proof. We have

$$\begin{aligned}&\left|\mathrm{Tr}\left[|\psi\rangle\langle\psi|\,O\right]-\mathrm{Tr}\left[|E_0\rangle\langle E_0|\,O\right]\right| \\ &\leqslant \|O\|_\infty \||\psi\rangle\langle\psi|E_0\rangle\langle E_0|\|_1 &\text{(C19)}\\ &\leqslant 2\|O\|_\infty\sqrt{1-|\langle\psi|E_0\rangle|^2} &\text{(C20)}\\ &= 2\|O\|_\infty\sqrt{(1-|\langle\psi|E_0\rangle|)\,(1+|\langle\psi|E_0\rangle|)} &\text{(C21)}\\ &\leqslant 2\sqrt{2}\|O\|_\infty\sqrt{1-|\langle\psi|E_0\rangle|}, &\text{(C22)}\end{aligned}$$

where in the first line we use Hölder's tracial-matrix inequality, in the second line we use the relation between the trace distance and the fidelity for pure states, in the third line we complete the square, and in the final line we use the fact that $|\langle\psi|E_0\rangle|\leqslant 1$. ■

Lemma 9 (Ground-state observable projection). Under the conditions specified in Lemma 7 with $\tau\geqslant\tau_\varepsilon$, the normalized state $|\psi\rangle = e^{-(1/2)\tau^2H^2}|\psi_0\rangle/\|e^{-(1/2)\tau^2H^2}|\psi_0\rangle\|$ satisfies

$$\mathrm{Tr}\left[|\psi\rangle\langle\psi|\,O\right]-\mathrm{Tr}\left[|E_0\rangle\langle E_0|\,O\right]\leqslant 2\|O\|\varepsilon, \tag{C23}$$

for any measurement operator O.

Proof. This follows as a direct implication of Lemmas 7 and 8. ■

We now introduce the Hubbard-Stratonovich transformation [109,110]. This gives us a way to decompose the operator $e^{-(1/2)\tau^2H^2}$ into a linear combination of implementable unitaries. It states that, for Hermitian H, we have

$$e^{-\frac{1}{2}\tau^2H^2}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}dz e^{-\frac{1}{2}z^2}e^{-iz\tau H}. \tag{C24}$$

The following lemma, adapted from Ref. [68], shows that the integral in the Hubbard-Stratonovich transformation can be discretized and truncated to give an approximate (discrete) Fourier series for the operator $e^{-(1/2)\tau^2H^2}$.

Lemma 10 (Approximate Hubbard-Stratonovich transformation—adapted from Appendix A2 of Ref. [68]). Truncating and discretizing the integral, and assuming $\|H\|\leqslant 1$, we have a Fourier-series approximation

$$\left\|\frac{1}{\sqrt{2\pi}}\sum_{k=-N_z}^{N_z}\Delta_z e^{-\frac{1}{2}z_k^2}e^{-iz_k\tau H}-e^{-\frac{1}{2}\tau^2H^2}\right\|\leqslant\varepsilon, \tag{C25}$$

for choice of $\Delta_z=\mathcal{O}\left(\tau^{-1}\right)$ and $\Delta_zN_z=\sqrt{2\log(2/\varepsilon)}$, where we have denoted $z_k=k\Delta_z$. This has maximum time-evolution parameter

$$t_{\max}(\varepsilon):=\max_k(z_k\tau)=\Delta_zN_z\tau=\tau\sqrt{2\log\left(\frac{2}{\varepsilon}\right)}. \tag{C26}$$

Further, the coefficients $\{\alpha_k(\varepsilon)\}_k=\{1/(\sqrt{2\pi})\Delta_ze^{-(1/2)z_k^2}\}_k$ have weight

$$\alpha(\varepsilon):=\sum_k|\alpha_k(\varepsilon)|\leqslant 1+\frac{1}{\sqrt{2\pi}}\Delta_z. \tag{C27}$$

Proof. The bound on $\alpha(\varepsilon)$ can be seen by noting that

$$\sum_k|\alpha_k(\varepsilon)|=\frac{1}{\sqrt{2\pi}}\sum_{k=-N_z}^{N_z}\Delta_ze^{-\frac{1}{2}z_k^2}\leqslant\frac{1}{\sqrt{2\pi}}\sum_{k=-\infty}^{\infty}\Delta_ze^{-\frac{1}{2}z_k^2} \tag{C28}$$

is simply a discretized Gaussian integral, where the discretization error can be bounded by the step size Δ_z multiplied by the maximum value of the function $1/\sqrt{2\pi}$. For the rest of the proof of the claim on $t_{\max}(\varepsilon)$, see Ref. [68, Appendix A2]. ■

From hereon, we will make the soft imposition that we choose $\Delta_z\leqslant 1$. This gives an error-independent bound for Lemma 10 of $\alpha(\varepsilon)\leqslant 1+1/\sqrt{2\pi}<1.4$.

Remark (Hamiltonians with nonpositive spectra). For Hamiltonians with nonpositive spectra and where we have upper bound on magnitude of ground-state energy $\lambda_0\geqslant|E_0|$, we can shift the spectrum $H'=H+E_0\mathbb{1}$ without changing the weight of the coefficients or the Hamiltonian-simulation problem. This is because in the Hubbard-Stratonovich transformation, the Hamiltonian appears in the term $e^{-iz_k\tau H'}=e^{-iz_k\tau H}e^{-iz_k\tau E_0\mathbb{1}}=e^{-iz_k\tau E_0}e^{-iz_k\tau H}$ and the phase factor $e^{-iz_k\tau E_0}$ can be absorbed into the coefficients α_k without changing the weight α. Further, the eigenstates of H' are clearly eigenstates of H, with shifted eigenenergies.

We can now present our result for ground-state property estimation (Corollary 2 in the main text).

Corollary 5 (Ground-state property estimation—detailed version). Consider a Hamiltonian $H=\sum_lE_l|E_l\rangle\langle E_l|$

with all eigenvalues $E_l \geqslant 0$ and known Pauli decomposition $H = \sum_\ell a_\ell P_\ell$ with $\lambda := \sum_\ell |a_\ell|$. Assume that the spectral gap is lower bounded by $\Delta \leqslant E_1 - E_0$. Additionally, we suppose that we have an initial trial state $|\psi_0\rangle$ with overlap with the ground state $\gamma := |\langle\psi_0|E_0\rangle|$. Finally, we assume that $E_0 \leqslant \Delta/\sqrt{2\log(\|O\|)/\varepsilon\gamma}$. Then, given the ability to measure observable O, there exists a random algorithm that returns $\langle E_0|O|E_0\rangle$ up to additive error ε with probability at least $(1-\delta)^2$, which consists of:

(i) a core routine that requires $\mathcal{O}\left(\log(2/\delta)\,\|O\|^2/\varepsilon^2\gamma^4\right)$ circuit runs of the form in Fig. 2(a) each of non-Clifford depth at most $\mathcal{O}\left(\frac{\lambda^2}{\Delta^2}\log^2\left(\|O\|/\varepsilon\gamma^2\right)\right)$

(ii) a subroutine that approximates the normalization constant, using asmyptotically equivalent resources to the core routine

Proof of Corollary 2. We first presume that the state-normalization constant $q = \|e^{-(1/2)\tau^2H^2}|\psi_0\rangle\|$ is given to us exactly and use Lemma 9 to show how to obtain an $\varepsilon/6$-additive statistical approximation to $1/q^2\langle\psi_0|e^{-(1/2)\tau^2H^2}Oe^{-(1/2)\tau^2H^2}|\psi_0\rangle$. We then use Proposition 7 to consider the overhead of approximating q, which will relax the approximation error to $\varepsilon/2$. This gives the desired ε-approximation to $\langle E_0|O|E_0\rangle$ for an appropriate choice of τ as specified by Lemma 7.

In order to quantify the complexity of approximating $1/q^2\langle\psi_0|e^{-(1/2)\tau^2H^2}Oe^{-(1/2)\tau^2H^2}|\psi_0\rangle$, we need to evaluate the quantities $\alpha\left(\varepsilon q^2/\|O\|\|e^{-(1/2)\tau^2H^2}\|, H\right)$ and $t_{\max}\left(\varepsilon q^2/\|O\|\|e^{-(1/2)\tau^2H^2}\|, H\right)$, as specified by Eq. (B30) of Theorem 1 and upper bound $1/q$. We recall that from the above discussion, by choosing that the step size Δ_z in the Fourier series in Lemma 10 to be less than 1, we fix $\alpha\left(\varepsilon q^2/\|O\|\|e^{-(1/2)\tau^2H^2}\|, H\right) = \mathcal{O}(1)$. Moreover, we have $1/q \leqslant \frac{1}{\gamma}e^{(1/2)\tau^2{E_0}^2} \leqslant 2/\gamma$ for $E_0\tau \leqslant 1$ as $e^{-\frac{1}{2}\tau^2{E_0}^2} \geqslant 1 - (1/2)\tau^2{E_0}^2 \geqslant 1/2$. Therefore, the sample complexity can be evaluated simply as

$$\mathcal{C}^O_{\text{sample}} = \mathcal{O}\left(\log\left(\frac{2}{\delta}\right)\frac{\|O\|^2}{\varepsilon^2q^4}\right) = \mathcal{O}\left(\log\left(\frac{2}{\delta}\right)\frac{\|O\|^2}{\varepsilon^2\gamma^4}\right). \tag{C29}$$

For the gate complexity, using Eq. (C26) of Lemma (10), we have

$$t_{\max}\left(\frac{\varepsilon q^2}{\|O\|\|e^{-\frac{1}{2}\tau^2H^2}\|}, H\right) = \tau\sqrt{\log\left(\frac{\|O\|\|e^{-\frac{1}{2}\tau^2H^2}\|}{\varepsilon q^2}\right)} \tag{C30}$$

$$\leqslant \tau\sqrt{\log\left(\frac{\|O\|}{\varepsilon\gamma^2}\right)}, \tag{C31}$$

where in the second line we have again used the fact that $1/q \leqslant \frac{1}{\gamma}e^{(1/2)\tau^2{E_0}^2}$. Suppose that we would like the exact approximation error to be $\varepsilon/2$, i.e., $\left|(1/q^2)\,\langle\psi_0|e^{-(1/2)\,\tau^2H^2}Oe^{-(1/2)\tau^2H^2}|\psi_0\rangle - \langle E_0|O|E_0\rangle\right| \leqslant \varepsilon/2$. Lemma 9 specifies that, in order to satisfy this, it is sufficient to have $\tau = \tau_{\varepsilon/2\|O\|} = 1/\Delta\sqrt{2\log(2\|O\|/\varepsilon\gamma)}$. Thus, overall, we have

$$t_{\max}\left(\frac{\varepsilon q^2}{\|O\|\|e^{-\frac{1}{2}\tau^2H^2}\|}, H\right) \leqslant \frac{1}{\Delta}\sqrt{2\log\left(\frac{2\|O\|}{\varepsilon\gamma}\right)}\sqrt{\log\left(\frac{\|O\|}{\varepsilon\gamma^2}\right)} = \mathcal{O}\left(\frac{1}{\Delta}\log\left(\frac{\|O\|}{\varepsilon\gamma^2}\right)\right), \tag{C32}$$

where we have used the fact that $\gamma \leqslant 1$. Substituting this into Eq. (B30) of Theorem 2, we obtain the stated result for non-Clifford gate complexity.

Finally, we consider the complexity of approximating the state norm $q = \|e^{-(1/2)\tau^2_{\varepsilon/\|O\|}H^2}|\psi_0\rangle\|$ by following Proposition 7. Equation (B59) of Proposition 7 expresses complexities in terms of $\alpha(\varepsilon q, H)$ and $t_{\max}(\varepsilon q, H)$. From the above discussion, we have that $\alpha(\varepsilon q, H) = \mathcal{O}(1)$. Using Eq. (C26) and Corollary 2, we can assign

$$t_{\max}(\varepsilon q, H) = \tau\sqrt{2\log\left(\frac{2}{\varepsilon q}\right)} \tag{C33}$$

$$= \frac{1}{\Delta}\sqrt{2\log\left(\frac{\|O\|}{\varepsilon\gamma}\right)}\sqrt{2\log\left(\frac{2}{\varepsilon q}\right)} \tag{C34}$$

$$= \mathcal{O}\left(\frac{1}{\Delta}\sqrt{\log\left(\frac{\|O\|}{\varepsilon\gamma}\right)\log\left(\frac{4}{\varepsilon\gamma}\right)}\right), \tag{C35}$$

where in the final line we have used the fact that $1/q \leqslant 2/\gamma$. This leads to the stated complexities for the normalization subroutine of

$$\mathcal{C}^{norm}_{\text{gate}} = \mathcal{O}\left(\log\left(\frac{2}{\delta}\right)\frac{1}{\gamma^4\varepsilon^2}\right),$$
$$\mathcal{C}^{norm}_{\text{sample}} = \mathcal{O}\left(\frac{\lambda^2}{\Delta^2}\log\left(\frac{\|O\|}{\varepsilon\gamma}\right)\log\left(\frac{4}{\varepsilon\gamma}\right)\right). \tag{C36}$$

■

a. Power method

In this section, we will explore the feasibility of using a randomized scheme based on the power method to find dominant eigenvalues. Quantum algorithms for estimating eigenvalues via the power method have previously been studied [111,112], though not in a randomized setting starting from Pauli access. Given an observable of interest, the aim will be to approximate $\langle E_0|O|E_0\rangle$ with

$\langle\psi_0|H^k O H^k|\psi_0\rangle / \|H^k|\psi_0\rangle\|^2$, where $|E_0\rangle$ is the ground state of Hamiltonian H, $|\psi_0\rangle$ is a given trial state with overlap $\gamma := |\langle E_0|\psi_0\rangle|$, and $\|H^k|\psi_0\rangle\|^2$. With this, we have the following proposition.

Supplementary Proposition 1 (Power method). For Hamiltonians H with negative spectra and first excited energy $E_1 < 0$, we have

$$\left| \frac{\langle\psi_0|H^k O H^k|\psi_0\rangle}{\|H^k|\psi_0\rangle\|^2} - \langle E_0|O|E_0\rangle \right| \leqslant \varepsilon, \tag{C37}$$

for a given error parameter $\varepsilon \leqslant (4\|O\|\sqrt{1-\gamma^2})/\gamma$ if k satisfies

$$k = \Omega\left(\frac{\log\left(\frac{\gamma\varepsilon}{\|O\|\sqrt{1-\gamma^2}}\right)}{\log\left(1-\frac{\Delta}{|E_0|}\right)} \right) = \Omega\left(\frac{|E_0|}{\Delta} \log\left(\frac{\|O\|\sqrt{1-\gamma^2}}{\gamma\varepsilon} \right) \right). \tag{C38}$$

Proof. Consider the decomposition of the trial state $|\psi_0\rangle$ in the energy eigenbasis of H as

$$|\psi_0\rangle = c_0|E_0\rangle + \sum_{j>0} c_j|E_j\rangle, \tag{C39}$$

where $\{c_k\}_k$ are coefficients. Using this, we can express $H^k|\psi_0\rangle$ as

$$H^k|\psi_0\rangle = c_0 E_0^k|E_0\rangle + \sum_{j>0} c_j E_j^k|E_j\rangle \tag{C40}$$

and thus the overlap of the normalized power-method approximation with the true ground state

$$1 - \left| \frac{\langle E_0|H^k|\psi_0\rangle}{\|H^k|\psi_0\rangle\|} \right| = 1 - \frac{|c_0 E_0^k|}{\sqrt{c_0^2 E^{2k} + \sum_{j>0} c_j^2 E_j^{2k}}} \tag{C41}$$

$$= 1 - \left[1 + \sum_{j>0} \frac{c_j^2}{c_0^2} \left(\frac{E_j}{E_0}\right)^{2k} \right]^{-1/2} \tag{C42}$$

$$\leqslant 1 - \left[1 + \frac{1-\gamma^2}{\gamma^2} \left(\frac{E_1}{E_0}\right)^{2k} \right]^{-1/2} \tag{C43}$$

$$\leqslant 1 - \left[1 - \frac{1}{2}\frac{1-\gamma^2}{\gamma^2} \left(\frac{E_1}{E_0}\right)^{2k} \right] \tag{C44}$$

$$= \frac{1-\gamma^2}{2\gamma^2} \left(\frac{E_1}{E_0}\right)^{2k}, \tag{C45}$$

where in the first inequality we have used the fact that $\gamma = E_0^2$ and $\sum_{j>0} c_j^2 = 1-\gamma^2$ and the second is due to the fact that $(1+x)^{1/2} \geqslant 1 + x/2$ for $x \geqslant -1$. Thus, in order to constrain the state overlap $1 - \left|\langle E_0|H^k|\psi_0\rangle / \|H^k|\psi_0\rangle\|\right| = \varepsilon'$, each of the following conditions are sufficient:

$$1 - \frac{|c_0 E_0^k|}{\sqrt{c_0^2 E^{2k} + \sum_{j>0} c_j^2 E_j^{2k}}} \leqslant \varepsilon', \tag{C46}$$

$$\Rightarrow \quad \left(1 - \frac{\Delta}{|E_0|}\right)^{2k} \leqslant \frac{2\gamma^2\varepsilon'}{1-\gamma^2}, \tag{C47}$$

$$\Rightarrow \quad k \geqslant \frac{1}{2} \frac{\log\left(\frac{2\gamma^2\varepsilon'}{1-\gamma^2}\right)}{\log\left(1-\frac{\Delta}{|E_0|}\right)}. \tag{C48}$$

We recall that Lemma 8 relates the overlap of pure-state expectation values to the overlap of states, which specifies that the state overlap ε' must be at most $\varepsilon^2/8\|O\|^2$ in order to constrain the expectation value within additive error ε. Using this, we obtain the first equality of Eq. (C38), where starting our assumption on ε ensures that the bound on k is positive and thus meaningful. The second equality can be established by noting that

$$1 - \frac{|E_0|}{\Delta} \leqslant \frac{1}{\log\left(1-\frac{\Delta}{|E_0|}\right)} \leqslant -\frac{|E_0|}{\Delta}, \tag{C49}$$

for $E_1 < 0$. ■

We now suppose that we are given a recipe to prepare the trial state by a series of Clifford gates or we have access to the amplitudes (for which Proposition 8 can be used). The above lemma along with Proposition 9 implies that, in this setting, there is a classical algorithm that solves the ground-state property-estimation problem with an exponential number of samples in Δ^{-1} and polynomial in all other parameters.

3. Gibbs-state property estimation

In this section, we demonstrate how the Fourier decomposition of the exponential function found in Ref. [86] leads to a randomized quantum algorithm to sample properties of the Gibbs state in our scheme. In Ref. [86], the authors assume that one has access to the purification of the Gibbs state of some intermediate Hamiltonian H_0 and aims to construct an approximation of the Gibbs state of some other Hamiltonian H. More precisely, we start with the state

$$|\Psi_0\rangle = \frac{1}{\sqrt{\mathcal{Z}_0}} \sum_i e^{-\beta E_{0,i}/2} |E_{0,i}\rangle |E_{0,i}^*\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B, \tag{C50}$$

where $\mathcal{Z}_0 := \mathrm{Tr}[e^{-\beta H_0}]$ is the partition function for H_0, and $\{|E_{0,i}\rangle\}_i$ are the eigenstates of H_0 with corresponding

eigenvalues $\{E_{0,i}\}_i$ ($\{|E^*_{0,i}\rangle\}_i$ are the eigenstates of H_0^*, the complex conjugate of H_0 in the computational basis). This state satisfies

$$\mathrm{Tr}_A[|\Psi_0\rangle\langle\Psi_0|] = \mathrm{Tr}_B[|\Psi_0\rangle\langle\Psi_0|] = \frac{e^{-\beta H_0}}{\mathcal{Z}_0}, \tag{C51}$$

i.e., it is the purification of the Gibbs state of H_0. The goal of Ref. [86] is then to prepare the Gibbs state of H,

$$\mathrm{Tr}_A[|\Psi\rangle\langle\Psi|] = \mathrm{Tr}_B[|\Psi\rangle\langle\Psi|] = \frac{e^{-\beta H}}{\mathcal{Z}} =: \gamma_\beta, \tag{C52}$$

where $\mathcal{Z} := \mathrm{Tr}[e^{-\beta H}]$. As our algorithmic framework only allows the estimation of observable and state overlaps, our randomized algorithm will approximate the expectation value $\mathrm{Tr}[\gamma_\beta O]$ for a given measurement operator O.

It is observed that the so-called "work operator" $W := H \otimes \mathbb{1} + \mathbb{1} \otimes H_0^*$ enables the transformation

$$\sqrt{\frac{\mathcal{Z}_0}{\mathcal{Z}}} e^{-\beta W/2} |\Psi_0\rangle = |\Psi\rangle\,, \tag{C53}$$

where $\sqrt{\mathcal{Z}_0/\mathcal{Z}} = 1/\|e^{-\beta W/2}|\Psi_0\rangle\|$ is the normalization factor. Thus, the relevant function of interest for us here is $e^{-\beta W/2}$. Reference [86] presents a Fourier decomposition of $e^{-\beta W/2}$ as follows.

Lemma 11 (Exponential operator—adapted from Lemmas 3.1, 3.2, and 3.4 of Ref. [86]). There exists an LCU Fourier decomposition $X = \sum_{j=0}^{2J} \alpha_j e^{i\tau_j W}$ that satisfies

$$\left\| \frac{\mathrm{Tr}_A[X|\Psi_0\rangle\langle\Psi_0|X^\dagger]}{\|X|\Psi_0\rangle\|^2} - \gamma_\beta \right\|_1 \leqslant 2\varepsilon, \tag{C54}$$

where if $[H_0, H] = 0$, the parameters in the decomposition satisfy

$$\alpha := \sum_{j=0}^{2J} |\alpha_j| \leqslant 2e^{\max\{4,\sqrt{\ln 6/\varepsilon}\}} e^{\beta\|V\|/2}\,, \tag{C55}$$

$$\|X|\Psi_0\rangle\| \geqslant \frac{1}{2}\|e^{-\beta W/2}|\Psi_0\rangle\| = \frac{1}{2}\sqrt{\frac{\mathcal{Z}}{\mathcal{Z}_0}}\,, \tag{C56}$$

$$\tau_{\max} := \max_j |\tau_j| = \frac{\pi\beta}{z}\left(\lceil \frac{1}{3} z^{3/2} \rceil - 1\right), \tag{C57}$$

where $z \leqslant \beta(\|W\| + \|V\|) + 2(\max\{4, \sqrt{\ln 6/\varepsilon}\})^2$ and we denote $V := H - H_0$.

We remark that due to Hölder's tracial-matrix inequality, Eq. (C54) implies that the expectation value with respect to any observable O is close up to additive error $2\|O\|\varepsilon$. Lemma 11 already specifies the conditions required to approximate observables with respect to the Gibbs state. Thus, we will not need to make use of Theorem 1 here and we can directly use Proposition 1. Equation (C54) can also be satisfied for noncommuting Hamiltonians and such a setting can also be transported to our framework. For simplicity, we only detail the commuting case here.

Corollary 6 (Gibbs-state property estimation—detailed version). Suppose access to the quantum state $|\Psi_0\rangle$ defined in Eq. (C50) and the ability to measure the observable O. Further, suppose that the Pauli decompositions of H_0 and H are known and $[H_0, H] = 0$. Then, we give a randomized quantum algorithm to approximate $\mathrm{Tr}[\gamma_\beta O]$ to additive error ε and success probability at least $(1-\delta)^2$, utilizing:

(i) a core routine using $\mathcal{O}(\log(1/\delta)(e^{\sqrt{\ln\|O\|/\varepsilon}})/(\varepsilon^2)\mathcal{Z}_0^2/(\mathcal{Z}^2)e^{2\beta\|V\|})$ circuit runs each of non-Clifford depth at most $\mathcal{O}\big(\lambda_W^2\beta^3\big(\|W\| + \|V\| + \log\big(\|O\|/\varepsilon\big)\big)\big)$
(ii) a subroutine to give the appropriate normalization, using asymptotically equivalent resources to the core routine

Proof. Being explicit, we first note that Lemma 11 straightforwardly implies that there exists an LCU Fourier decomposition $X = \sum_{j=0}^{2J} \alpha_j e^{i\tau_j W}$ for which the state

$$\rho = \frac{\mathrm{Tr}_A[X|\Psi_0\rangle\langle\Psi_0|X^\dagger]}{\|X|\Psi_0\rangle\|^2} \tag{C58}$$

satisfies

$$\big|\mathrm{Tr}[\rho O] - \mathrm{Tr}[\gamma_\beta O]\big| \leqslant \varepsilon/2\,, \tag{C59}$$

where the parameters in the decomposition satisfy

$$\alpha := \sum_{j=0}^{2J} |\alpha_j| \leqslant 2e^{\max\{4,\sqrt{\ln 24\|O\|/\varepsilon}\}} e^{\beta\|V\|/2}, \tag{C60}$$

$$\|X|\Psi_0\rangle\| \geqslant \frac{1}{2}\|e^{-\beta W/2}|\Psi_0\rangle\| = \frac{1}{2}\sqrt{\frac{\mathcal{Z}}{\mathcal{Z}_0}}, \tag{C61}$$

$$\tau_{\max} := \max_j |\tau_j| = \frac{\pi\beta}{z}\left(\lceil \frac{1}{3} z^{3/2} \rceil - 1\right), \tag{C62}$$

where $z \leqslant \beta(\|W\| + \|V\|) + 2(\max\{4, \sqrt{\ln 24\|O\|/\varepsilon}\})^2$. Proposition 6 and Lemma 3 then give the resources required to statistically approximate $\mathrm{Tr}[\rho O]$ to additive error $\varepsilon/2$. By the triangle inequality, this implies a statistical approximation of the exact answer $\mathrm{Tr}[\gamma_\beta O]$ to additive error ε. We detail this below.

We start by presuming that $\|X\,|\Psi_0\rangle\,\|$ is known exactly. Then, with this exact quantity, Proposition 6 gives a randomized algorithm to statistically approximate $\mathrm{Tr}[\rho_1 O]$ to additive error $\varepsilon/6$ using

$$\begin{aligned}&\mathcal{C}_{\text{sample}}\\&= 2e^4 \log\left(\frac{2}{\delta}\right)\frac{36\|O\|^2\alpha^4}{\|X\,|\Psi_0\rangle\,\|^4\varepsilon^2}\\&\leqslant 18432e^4\log\left(\frac{2}{\delta}\right)\frac{\|O\|^2}{\varepsilon^2}\frac{\mathcal{Z}_0^2}{\mathcal{Z}^2}e^{4\max\{4,\sqrt{\ln 24\|O\|/\varepsilon}\}}e^{2\beta\|V\|}\end{aligned} \tag{C63}$$

circuits, each of non-Clifford depth at most

$$\mathcal{C}_{\text{gate}} = 2\lambda_W^2\tau_{\max}^2 \leqslant 2\lambda_W^2\frac{\pi^2\beta^2}{z^2}\left(\lceil\frac{1}{3}z^{3/2}\rceil - 1\right)^2, \tag{C64}$$

and thus our stated result follows.

We now check the complexity of approximating $\|X\,|\Psi_0\rangle\,\|^2$. We recall that Lemma 3 specifies that in order to statistically approximate $\mathrm{Tr}[\rho O]$ to additive error $\varepsilon/2$, one requires the above-stated conditions (an approximation of $\mathrm{Tr}[\rho O]$ to additive error $\varepsilon/6$) and an approximation to $\|X\,|\Psi_0\rangle\,\|^2$ with additive error $1/12(\varepsilon\|X\,|\Psi_0\rangle\,\|^2)/\|O\|$. Thus (once again using Proposition 6), we see that it is sufficient to use

$$\begin{aligned}\mathcal{C}_{\text{sample}}^{norm} &= 2e^4\log\left(\frac{2}{\delta}\right)\frac{144\|O\|^2\alpha^4}{\|X\,|\Psi_0\rangle\,\|^4\varepsilon^2}\\&\leqslant 73728e^4\log\left(\frac{2}{\delta}\right)\frac{\|O\|^2}{\varepsilon^2}\frac{\mathcal{Z}_0^2}{\mathcal{Z}^2}\\&\quad\times e^{4\max\{4,\sqrt{\ln 24\|O\|/\varepsilon}\}}e^{2\beta\|V\|}\end{aligned} \tag{C65}$$

circuits, each of non-Clifford depth at most

$$\mathcal{C}_{\text{gate}}^{norm} = 2\lambda_W^2\tau_{\max}^2 \leqslant 2\lambda_W^2\frac{\pi^2\beta^2}{z^2}\left(\lceil\frac{1}{3}z^{3/2}\rceil - 1\right)^2. \tag{C66}$$

We see that this equivalent to the complexity of the core routine, up to a constant factor. ■

We remark that, as with all our algorithms, the constant factor appearing in the sample complexity can be refined slightly by dividing the statistical and exact contributions to error unevenly. For instance, by choosing the statistical contribution to be 90% of the total error (rather than 50%), the sample complexity for the normalization constant becomes

$$\begin{aligned}\mathcal{C}_{\text{sample}}^{norm} &\leqslant 22756e^4\log\left(\frac{2}{\delta}\right)\frac{\|O\|^2}{\varepsilon^2}\frac{\mathcal{Z}_0^2}{\mathcal{Z}^2}\\&\quad\times e^{4\max\{4,\sqrt{\ln 120\|O\|/\varepsilon}\}}e^{2\beta\|V\|}.\end{aligned}$$

4. Evaluating Green's functions

In this section, we present the detailed statement of our result for evaluating Green's functions as defined in Eqs. (30) and (31) (Proposition 5 in the main text).

Proposition 10 (Green's-function estimation—detailed version). Consider a Hamiltonian $H = \sum_l E_l\,|E_l\rangle\,\langle E_l|$ with all eigenvalues $E_l \geqslant 0$ and known Pauli decomposition $H = \sum_\ell a_\ell P_\ell$; $\lambda := \sum_\ell |a_\ell|$. Assume that the spectral gap is lower bounded by $\Delta \leqslant E_1 - E_0$. Additionally, we suppose that we can freely prepare an initial trial state $|\psi_0\rangle$ with overlap with the ground state $\gamma := |\langle\psi_0|E_0\rangle|$. Given parameters ω, η, and the ground-state energy E_0, there exists a random compiler that returns Eqs. (30) and (31) up to additive error ε with probability at least $(1-\delta)^2$, utilizing

$$\mathcal{O}\left(\log\left(\frac{2}{\delta}\right)\frac{\|(\Gamma^{(\pm)})^{-1}\|^2}{\gamma^4\varepsilon^2}\log\left(\frac{\|(\Gamma^{(\pm)})^{-1}\|}{\varepsilon}\right)\right)$$

circuit runs, respectively, each of non-Clifford depth at most

$$\begin{aligned}&\mathcal{O}\left(\frac{\lambda_H^2}{\Delta^2}\log^2\left(\frac{2\|(\Gamma^{(\pm)})^{-1}\|}{\varepsilon\gamma^2}\right)\right.\\&\left.+\,(|\hbar\omega \pm E_0| + \eta + \lambda_H)^2\|\Gamma^{(\pm)-1}\|^2\log^2\left(\frac{\|\Gamma^{(\pm)-1}\|}{\varepsilon}\right)\right),\end{aligned}$$

with a normalization subroutine that has smaller complexities than the main algorithm.

Proof of Proposition 5. We recall that, from Secs. III and IV, we have decompositions of the matrix inverse function and the Gaussian function in terms of Pauli gates and Pauli rotations, given that we have prior knowledge of the Pauli decomposition of the matrix.

Namely, by combining Lemmas 2 and 10, we have, for some Hamiltonian,

$$\left\|e^{-\frac{1}{2}\tau^2H^2} - f_{GS}(\varepsilon, H)\right\| \leqslant \varepsilon, \tag{C67}$$

where $f_{GS}(\varepsilon,H) = \sum_{i\in S_{\varepsilon,H}} h_i(\varepsilon,H)\,V_i(\varepsilon,H)$ is a linear combination of gates with weight $R_{GS} := \sum_{i\in S_{\varepsilon,H}}|h_i(\varepsilon,H)|$. The weight of this linear combination satisfies $R_{GS} = \Theta(1)$ and each V_i has non-Clifford gate depth at most $\Theta\left(\lambda_H^2\tau^2\log(2/\varepsilon)\right)$, where $\lambda_H := \sum_\ell |a_\ell|$ is the Pauli weight of H. Moreover, we recount that Corollary 9 states that applying the operator $e^{-(1/2)\tau^2H^2}$ (after normalization) with $\tau = 1/\Delta\sqrt{2\log(1/\varepsilon\gamma)}$ approximately projects to the ground state with $2\|O\|\varepsilon$ additive error when considering an expectation value of O.

Likewise, by combining Lemma 1 and Eq. (C7) in Lemma 6, given some operator $\Gamma = \sum_l a_l P_l$, we have

$$\left\| \Gamma^{-1} - f_{inv}(\varepsilon, \Gamma) \right\| \leqslant \varepsilon, \tag{C68}$$

where $f_{inv}(\varepsilon,\Gamma) = \sum_{i\in S_{\varepsilon,\Gamma}} \gamma_i(\varepsilon,\Gamma)\, U_i(\varepsilon,\Gamma)$ is a linear combination of gates with weight $R_{inv} := \sum_{i\in S_{\varepsilon,\Gamma}} |\gamma_i(\varepsilon,\Gamma)|$ satisfying

$$R_{inv} = \Theta\left(\|\Gamma^{-1}\| \sqrt{\log\left(\frac{\|\Gamma^{-1}\|}{\varepsilon} \right)} \right), \tag{C69}$$

and each U_i has non-Clifford gate depth at most $\Theta\left(\lambda_\Gamma^2 \|\Gamma^{-1}\|^2 \log^2\left(\|\Gamma^{-1}\|/\varepsilon\right)\right)$, where λ_Γ is the Pauli weight of Γ.

We now specifically consider the Green's-function problem. We wish to find the inverse of the (non-Hermitian) operator $\Gamma^{(\pm)} = (\hbar\omega \pm E_0)\mathbb{1} \mp H + i\eta\mathbb{1}$. Following the remark in Sec. III, this can be embedded in a Hermitian operator by dilating the space and considering $Y^{(\pm)} = X \otimes ((\hbar\omega \pm E_0)\mathbb{1} \mp H) + Y \otimes \eta\mathbb{1}$. These operators have Pauli weight $\lambda_{Y^{(\pm)}} \leqslant |\hbar\omega \pm E_0| + \lambda_H + \eta$ and satisfy $\|Y^{(\pm)}\| = \|\Gamma^{(\pm)-1}\|$. We can then write the quantity that we wish to prepare as

$$G^{(\pm)} = \langle 1, E_0 | \hat{a}_i \left(Y^{(\pm)}\right)^{-1} \hat{a}_j^\dagger | 0, E_0 \rangle . \tag{C70}$$

In order to sample from the ground state, we will need to approximate the normalization constant $q^2 := \|e^{-(1/2)\tau^2 H^2}|\psi_0\rangle\|^2$, where $|\psi_0\rangle$ is the trial ground state. We denote the statistical approximation of this quantity with M_q shots as $Q^{(M_q)}$. We denote the statistical approximation of $\langle 1, \psi_0 | e^{-(1/2)\tau^2 H^2} \hat{a}_i (Y^{(\pm)})^{-1} \hat{a}_j^\dagger e^{-(1/2)\tau^2 H^2} | 0, \psi_0 \rangle$ as E^M. We can then express the full approximation error as

$$\left| \frac{1}{Q^{(M_q)}} E^M - \langle 1, E_0 | \hat{a}_i \left(Y^{(\pm)}\right)^{-1} \hat{a}_j^\dagger | 0, E_0 \rangle \right| \leqslant \tag{C71}$$

$$\leqslant \left| \langle 1, E_0 | \hat{a}_i \left(Y^{(\pm)}\right)^{-1} \hat{a}_j^\dagger | 0, E_0 \rangle - \frac{1}{q^2} \langle 1, \psi_0 | e^{-\frac{1}{2}\tau^2 H^2} \hat{a}_i (Y^{(\pm)})^{-1} \hat{a}_j^\dagger e^{-\frac{1}{2}\tau^2 H^2} | 0, \psi_0 \rangle \right| \tag{C72}$$

$$+ \frac{1}{q^2} \left| \langle 1, \psi_0 | e^{-\frac{1}{2}\tau^2 H^2} \hat{a}_i (Y^{(\pm)})^{-1} \hat{a}_j^\dagger e^{-\frac{1}{2}\tau^2 H^2} | 0, \psi_0 \rangle - \langle 1, \psi_0 | f_{GS}(\tilde{\varepsilon}_1, H)^\dagger \hat{a}_i (Y^{(\pm)})^{-1} \hat{a}_j^\dagger f_{GS}(\tilde{\varepsilon}_1, H) | 0, \psi_0 \rangle \right| \tag{C73}$$

$$+ \frac{1}{q^2} \left| \langle 1, \psi_0 | f_{GS}(\tilde{\varepsilon}_1, H)^\dagger \hat{a}_i (Y^{(\pm)})^{-1} \hat{a}_j^\dagger f_{GS}(\tilde{\varepsilon}_1, H) | 0, \psi_0 \rangle - \langle 1, \psi_0 | f_{GS}(\tilde{\varepsilon}_1, H)^\dagger \hat{a}_i f_{inv}(\tilde{\varepsilon}_2, Y^{(\pm)}) \hat{a}_j^\dagger f_{GS}(\tilde{\varepsilon}_1, H) | 0, \psi_0 \rangle \right| \tag{C74}$$

$$+ \frac{1}{q^2} \left| \langle 1, \psi_0 | f_{GS}(\tilde{\varepsilon}_1, H)^\dagger \hat{a}_i f_{inv}(\tilde{\varepsilon}_2, Y^{(\pm)}) \hat{a}_j^\dagger f_{GS}(\tilde{\varepsilon}_1, H) | 0, \psi_0 \rangle - \bar{z}^{(M)} \right| \tag{C75}$$

$$+ \left| \frac{1}{q^2} - \frac{1}{Q^{(M_q)}} \right| \cdot \left| \bar{z}^{(M)} \right| , \tag{C76}$$

where in the above we have used a chain of triangle inequalities. We now specify conditions so that the right-hand side of the above is $\mathcal{O}(\varepsilon)$. Corollary 9 bounds (C72) by 2ε with the choice

$$\tau = \frac{1}{\Delta} \sqrt{2 \log \frac{\|\hat{a}_i Y^{(\pm)-1} \hat{a}_j^\dagger\|}{\varepsilon\gamma}} \leqslant \frac{1}{\Delta} \sqrt{2 \log \frac{\|Y^{(\pm)-1}\|}{\varepsilon\gamma}} = \frac{1}{\Delta} \sqrt{2 \log \frac{\|\Gamma^{(\pm)-1}\|}{\varepsilon\gamma}}.$$

The term in Eq. (C73) is bounded by 24ε by Eq. (C67) and Lemma 2 under the condition that each V_i appearing in $f_{GS}(\tilde{\varepsilon}_1, H)$ has non-Clifford gate depth at most $2\lambda_H^2 \tau^2 \log\left(2\|\Gamma^{(\pm)-1}\|/\varepsilon q^2\right)$, which ensures that $\left\| e^{-(1/2)\tau^2 H^2} - f_{GS}(\tilde{\varepsilon}_1, H) \right\| \leqslant \varepsilon q^2 / (\|\Gamma^{(\pm)-1}\|)$. The term in Eq. (C74) can be bounded with Hölder's inequality as

$$\text{(C74)} \leqslant \left\| Y^{(\pm)} - f_{inv}(\tilde{\varepsilon}_2, Y^{(\pm)}) \right\|_\infty \cdot \frac{1}{q^2} \left\| \hat{a}_j^\dagger f_{GS}(\tilde{\varepsilon}_1, H) \left|1, \psi_0\right\rangle\!\left\langle 0, \psi_0\right| f_{GS}(\tilde{\varepsilon}_1, H)^\dagger \hat{a}_i \right\|_1 \tag{C77}$$

$$= \left\| Y^{(\pm)} - f_{inv}(\tilde{\varepsilon}_2, Y^{(\pm)}) \right\|_\infty \cdot \frac{1}{q^2} \left\| f_{GS}(\tilde{\varepsilon}_1, H) \left|1, \psi_0\right\rangle\!\left\langle 0, \psi_0\right| f_{GS}(\tilde{\varepsilon}_1, H)^\dagger \right\|_1 \tag{C78}$$

$$\leqslant \left\| Y^{(\pm)} - f_{inv}(\tilde{\varepsilon}_2, Y^{(\pm)}) \right\|_\infty \cdot \frac{1}{q^2} \left\| f_{GS}(\tilde{\varepsilon}, H) \left|\psi_0\right\rangle\!\left\langle \psi_0\right| \right\|_2^2 \tag{C79}$$

$$\leqslant \left\| Y^{(\pm)} - f_{inv}(\tilde{\varepsilon}_2, Y^{(\pm)}) \right\|_\infty \cdot \frac{1}{q^2} \left(\left\| e^{-\frac{1}{2}\tau^2 H^2} \left|\psi_0\right\rangle\!\left\langle \psi_0\right| \right\|_2 + \left\| \left(e^{-\frac{1}{2}\tau^2 H^2} - f_{GS}(\tilde{\varepsilon}_1, H) \right) \left|\psi_0\right\rangle\!\left\langle \psi_0\right| \right\|_2 \right)^2 \tag{C80}$$

$$\leqslant \left\| Y^{(\pm)} - f_{inv}(\tilde{\varepsilon}_2, Y^{(\pm)}) \right\|_\infty \cdot \frac{1}{q^2} \left(\left\| e^{-\frac{1}{2}\tau^2 H^2} \left|\psi_0\right\rangle \right\|_2 + \left\| \left(e^{-\frac{1}{2}\tau^2 H^2} - f_{GS}(\tilde{\varepsilon}_1, H) \right) \right\|_\infty \right)^2 \tag{C81}$$

$$\leqslant \left\| Y^{(\pm)} - f_{inv}(\tilde{\varepsilon}_2, Y^{(\pm)}) \right\|_\infty \left(1 + 2\frac{\varepsilon q}{\|\Gamma^{(\pm)-1}\|} + \frac{\varepsilon^2 q^2}{\|\Gamma^{(\pm)-1}\|^2} \right), \tag{C82}$$

where in the equality we have used the unitary invariance of the Schatten norm, noting that $\hat{a}_j^\dagger$ and $\hat{a}_i$ can be expressed as probabilistic combinations of two unitaries. In the second inequality, we have again used Hölder's inequality and the third inequality is due to the triangle inequality. Thus, assuming $\varepsilon \leqslant (\|\Gamma^{(\pm)-1}\|)/q$ (of which we expect the right-hand-side to be much larger than 1), the term in Eq. (C74) can be bounded by 4ε by setting $\tilde{\varepsilon}_2 = \varepsilon$, which sets $R_{inv} = \Theta\left(\|\Gamma^{(\pm)-1}\|\sqrt{\log\left(\|\Gamma^{(\pm)-1}\|/\varepsilon\right)}\right)$ and the non-Clifford gate depth of each U_i as being at most $\Theta\left(\lambda_{Y^{(\pm)}}^2 \|\Gamma^{(\pm)-1}\|^2 \log^2\left(\|\Gamma^{(\pm)-1}\|/\varepsilon\right)\right)$. The term in Eq. (C75) can be bounded by ε with probability $(1-\delta)$ by asking for number of shots

$$M \geqslant 2\log\left(\frac{2}{\delta}\right)\frac{R_{inv}^2 R_{GS}^4}{\varepsilon q^4} = \mathcal{O}\left(\log\left(\frac{2}{\delta}\right)\frac{\|\Gamma^{(\pm)-1}\|^2}{\varepsilon^2\gamma^4}\log\left(\frac{\|\Gamma^{(\pm)-1}\|}{\varepsilon}\right)\right), \tag{C83}$$

where we have used the fact that $1/q \leqslant 2/\gamma$. Finally, the term in Eq. (C76) can be bounded by ε with probability at least $(1-\delta)$ by using number of shots

$$M_q \geqslant 2\log\left(\frac{2}{\delta}\right)\frac{\|Y^{(\pm)-1}\|^2 R_{GS}^4}{\varepsilon q^4} = \mathcal{O}\left(\log\left(\frac{2}{\delta}\right)\frac{\|\Gamma^{(\pm)-1}\|^2}{\varepsilon^2\gamma^4}\right), \tag{C84}$$

for the normalization subroutine.

Finally, we note the total gate complexity consists of the sum of the gate complexities for U_i and V_i, respectively. From the above discussion, we see that the non-Clifford gate complexity of V_i is $\mathcal{O}\left(\frac{\lambda_H^2}{\Delta^2}\log\left(2\|\Gamma^{(\pm)-1}\|/\varepsilon\gamma^2\right)\right)$ and the non-Clifford gate complexity of U_i is $\mathcal{O}\left((|\hbar\omega \pm E_0| + \eta + \lambda_H)^2\|\Gamma^{(\pm)-1}\|^2 \log^2\left(\|\Gamma^{(\pm)-1}\|/\varepsilon\right)\right)$. The non-Clifford gate complexity for the normalization subroutine as given by Proposition 2 is $\mathcal{O}\left(\frac{\lambda_H^2}{\Delta^2}\log\left(\|\Gamma^{(\pm)-1}\|/\varepsilon\gamma\right)\log\left(1/\varepsilon\gamma\right)\right)$. ■

We now briefly remark on the scenario if we would like to evaluate the Green's function but the exact ground-state energy is not given. In this case, an ε-additive approximation to E_0 would yield an amplification of the additive error of the evaluation of the Green's functions by an additive term $\mathcal{O}(\varepsilon\|\Gamma^{(\pm)-1}\|\|(\Gamma^{(\pm)} \pm \varepsilon\mathbb{1})^{-1}\|)$. This can be seen by noting that

$$\left\| \left(\Gamma^{(\pm)} \pm \varepsilon\mathbb{1}\right)^{-1} - \Gamma^{(\pm)-1} \right\| = \left\| \left(\mathbb{1} - \Gamma^{(\pm)-1}\left(\Gamma^{(\pm)} \pm \varepsilon\mathbb{1}\right)\right)\left(\Gamma^{(\pm)} \pm \varepsilon\mathbb{1}\right)^{-1} \right\| \tag{C85}$$

$$= \varepsilon \left\| \Gamma^{(\pm)-1}\left(\Gamma^{(\pm)} \pm \varepsilon\mathbb{1}\right)^{-1} \right\|. \tag{C86}$$

5. Additional information for complexity comparison

In this section, we provide some additional information on state-of-the-art quantum algorithms in the literature, to aid exposition of Tables I, II, III, and IV.

a. Data access

In our tables, we quote two generic block-encoding strategies. First, we use the result of Ref. [17], which gives an explicit block encoding of $A/\|A\|_F$, starting from a description of $A \in \mathbb{C}^{N\times N}$ in the computational basis. Throughout the paper, we quote the minimal gate depth implementation of this construction, which is achieved using $\mathcal{O}(N^2)$ qubits and $\mathcal{O}(\log N)$ gate depth each call to the block encoding. Reference [17] gives an alternative construction using $\mathcal{O}(N)$ qubits but $\mathcal{O}(N)$ gate depth each call. A second block-encoding strategy that we quote

is to use SELECT and PREPARE oracles when A is given in the Pauli basis $A = \sum_{\ell=1}^{L} a_\ell P_\ell$ with Pauli weight $\lambda := \sum_{\ell=1}^{L} |a_\ell|$. In general, this strategy works for any unitary decomposition where one presumes that controlled versions of the unitary can be implemented in $\mathcal{O}(1)$ gate depth. This strategy generically uses $\mathcal{O}(\log L)$ qubits and $\mathcal{O}(L)$ gate depth each call to the block encoding. The block encoding has subnormalization λ.

We also conceptualize the ability to perform Hamiltonian simulation as a type of data access. In Tables III and IV, we quote two Hamiltonian-simulation strategies; one that uses no additional ancillary qubits and a second that is optimal in gate complexity. First, we quote first-order Trotter time evolution. For our purposes, we again work in the Pauli access model and the goal is to return an approximation of $\exp(iAt)$ for some real time t. First-order Trotter uses (in the worst case) $\mathcal{O}\left(L^3(\Lambda t)^2/\epsilon\right)$ gates, where we have defined $\Lambda := \max_\ell a_\ell$. As we state in the main text, one can replace Λ with a smaller quantity by exploiting commutator structure between terms [73]. The second Hamiltonian-simulation strategy that we quote is the so-called qubitization approach of Ref. [62]. Here, one again constructs a block encoding using SELECT and PREPARE and makes calls to this block encoding. The algorithm uses $\mathcal{O}(\log L)$ ancillary qubits and has query complexity $\mathcal{O}(\lambda t + \log(1/\varepsilon)/\log\log(1/\varepsilon))$, where each query uses $\mathcal{O}(L)$ gates as stated above. Thus, each call to this time evolution as an oracle requires $\mathcal{O}(\lambda L t + \log(1/\varepsilon)/\log\log(1/\varepsilon))$ gates.

Finally, in Table I for the HHL algorithm [12], we also require a Hamiltonian-simulation subroutine when starting from a Frobenius-norm block encoding. Again using Ref. [62], an approximation of $\exp(iAt)$ can be obtained with $\mathcal{O}(\mu t + \log(1/\varepsilon)/\log\log(1/\varepsilon))$ calls to the block encoding, where μ is the block-encoding subnormalization.

b. Linear systems

The HHL algorithm [12] consists of three key steps. First, conditional Hamiltonian simulation of maximum time $\mathcal{O}(\|A^{-1}\|/\varepsilon)$ is required to perform phase estimation to estimate each inverse eigenvalue λ_j^{-1} to additive error $\mathcal{O}(1/(\lambda_j)\varepsilon/(\|A^{-1}\|)) = \mathcal{O}(\varepsilon)$. Second, after rotating by angle $\arcsin(1/\lambda_j\|A^{-1}\|)$ conditioned on λ_j and undoing phase estimation, we have the state $|\lambda_j\rangle|0\rangle \to |\lambda_j\rangle\left(\frac{1}{\lambda_j\|A^{-1}\|}|0\rangle + \sqrt{1-\frac{1}{\lambda_j^2\|A^{-1}\|^2}}\right)$. Finally, the desired state can thus be obtained with $\mathcal{O}(\kappa)$ amplitude-amplification steps. Putting this all together, starting with a Frobenius-norm block encoding (which requires $\mathcal{O}(\log N)$ gates per call), approximately preparing the state $|A^{-1}b\rangle$ requires $\mathcal{O}(\|A\|_F\|A^{-1}\|\kappa\log N/\varepsilon) = \mathcal{O}(\kappa_F\kappa\log N/\varepsilon)$ gate depth to implement.

In the quantum linear-systems algorithm of Ref. [13], the dominant part is a filtering step with complexity $\mathcal{O}(1/(\Delta)\log(1/\varepsilon))$, where Δ is a lower bound on the gap of the adiabatic Hamiltonian which interpolates between Hamiltonians

$$H_0 := \begin{pmatrix} 0 & Q_b \\ Q_b & 0 \end{pmatrix}, \quad H_1 := \begin{pmatrix} 0 & \hat{A}Q_b \\ Q_b\hat{A} & 0 \end{pmatrix}, \tag{C87}$$

for some encodable matrix $\hat{A}$ and where $Q_b = \mathbb{1} - |b\rangle\langle b|$. It is shown in Ref. [57, Appendix A] that this has a gap lower bounded by $\|\hat{A}^{-1}\|$. In Ref. [13], the authors call for a block-encoding construction; thus we block encode an adiabatic Hamiltonian with $\hat{A} = A/\mu$, leading to algorithm complexity $\mu\|A^{-1}\|\log(1/\varepsilon)$, where μ is the block-encoding subnormalization.

c. Ground states and Gibbs states

In Tables III and IV, all the algorithms with which we compare use either Hamiltonian simulation or block-encoding oracles. Thus, the quoted gate complexities in our tables can be extracted by inspecting the query complexity listed in each relevant work and multiplying by the corresponding gate complexity for oracle access detailed in Sec. C 5 a.

[1] J. Preskill, Quantum computing in the NISQ era and beyond, Quantum **2**, 79 (2018).

[2] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, Wai-Keong Mok, Sukin Sim, Leong-Chuan Kwek, and Alán Aspuru-Guzik, Noisy intermediate-scale quantum algorithms, Rev. Mod. Phys. **94**, 015004 (2022).

[3] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, Variational quantum algorithms, Nat. Rev. Phys. **3**, 625 (2021).

[4] A. Montanaro, Quantum algorithms: An overview, npj Quantum Inf. **2**, 1 (2016).

[5] R. De Wolf, Quantum computing: Lecture notes, arXiv Preprint arXiv:1907.09415 (2019).

[6] B. Bauer, S. Bravyi, M. Motta, and G. K.-L. Chan, Quantum algorithms for quantum chemistry and quantum materials science, Chem. Rev. **120**, 12685 (2020).

[7] L. Lin, Lecture notes on quantum algorithms for scientific computation, arXiv Preprint arXiv:2201.08309 (2022).

[8] A. M. Dalzell, S. McArdle, M. Berta, P. Bienias, C.-F. Chen, A. Gilyén, C. T. Hann, M. J. Kastoryano, E. T. Khabiboulline, and A. Kubica *et al.*, Quantum algorithms: A survey of applications and end-to-end complexities, arXiv Preprint arXiv:2310.03011 (2023).

[9] A. Gilyén, Y. Su, G. H. Low, and N. Wiebe, in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing* (Association for Computing Machinery, New York, NY, USA, 2019), p. 193.

[10] J. M. Martyn, Z. M. Rossi, A. K. Tan, and I. L. Chuang, Grand unification of quantum algorithms, PRX Quantum **2**, 040203 (2021).

[11] By "end-to-end," we mean that all relevant resources are accounted for in an implementation that solves the required task in full.

[12] A. W. Harrow, A. Hassidim, and S. Lloyd, Quantum algorithm for linear systems of equations, Phys. Rev. Lett. **103**, 150502 (2009).

[13] P. Costa, D. An, Y. R. Sanders, Y. Su, R. Babbush, and D. W. Berry, Optimal scaling quantum linear systems solver via discrete adiabatic theorem, arXiv Preprint arXiv:2111.08152 (2021).

[14] Quantum random access memory (QRAM) is a widely studied data access structure that allows for coherent access to classical data [16,23,113].

[15] C. T. Hann, C.-L. Zou, Y. Zhang, Y. Chu, R. J. Schoelkopf, S. M. Girvin, and L. Jiang, Hardware-efficient quantum random access memory with hybrid quantum acoustic systems, Phys. Rev. Lett. **123**, 250501 (2019).

[16] V. Giovannetti, S. Lloyd, and L. Maccone, Quantum random access memory, Phys. Rev. Lett. **100**, 160501 (2008).

[17] B. D. Clader, A. M. Dalzell, N. Stamatopoulos, G. Salton, M. Berta, and W. J. Zeng, Quantum resources required to block-encode a matrix of classical data, arXiv Preprint arXiv:2206.03505 (2022).

[18] The minimum qubit-count implementation in Ref. [17] requires $\mathcal{O}(N)$ qubits and $\mathcal{O}(N)$ depth per call to the block encoding. Alternatively, a quantum read-only memory (QROM) can be implemented using $\mathcal{O}(\log N)$ qubits, but $\mathcal{O}(N^2)$ gates.

[19] A. M. Childs, D. Maslov, Y. Nam, N. J. Ross, and Y. Su, Toward the first quantum simulation with quantum speedup, Proc. Natl. Acad. Sci. **115**, 9456 (2018).

[20] R. Babbush, C. Gidney, D. W. Berry, N. Wiebe, J. McClean, A. Paler, A. Fowler, and H. Neven, Encoding electronic spectra in quantum circuits with linear T complexity, Phys. Rev. X **8**, 041015 (2018).

[21] D. W. Berry, C. Gidney, M. Motta, J. R. McClean, and R. Babbush, Qubitization of arbitrary basis quantum chemistry leveraging sparsity and low rank factorization, Quantum **3**, 208 (2019).

[22] K. Wan, Exponentially faster implementations of Select(H) for fermionic Hamiltonians, Quantum **5**, 380 (2021).

[23] O. Di Matteo, V. Gheorghiu, and M. Mosca, Fault-tolerant resource estimation of quantum random-access memories, IEEE Trans. Quantum Eng. **1**, 1 (2020).

[24] S. Aaronson, Read the fine print, Nat. Phys. **11**, 291 (2015).

[25] T. Strohmer and R. Vershynin, A randomized Kaczmarz algorithm with exponential convergence, J. Fourier Anal. Appl. **15**, 262 (2009).

[26] A. Gilyén, Z. Song, and E. Tang, An improved quantum-inspired algorithm for linear regression, Quantum **6**, 754 (2022).

[27] C. Shao and A. Montanaro, Faster quantum-inspired algorithms for solving linear systems, ACM Trans. Quantum Comput. **3**, 1 (2022).

[28] E. T. Campbell, Early fault-tolerant simulations of the Hubbard model, Quantum Sci. Technol. **7**, 015007 (2021).

[29] L. Lin and Y. Tong, Heisenberg-limited ground-state energy estimation for early fault-tolerant quantum computers, PRX Quantum **3**, 010318 (2022).

[30] K. Wan, M. Berta, and E. T. Campbell, Randomized quantum algorithm for statistical phase estimation, Phys. Rev. Lett. **129**, 030503 (2022).

[31] Y. Dong, L. Lin, and Y. Tong, Ground state preparation and energy estimation on early fault-tolerant quantum computers via quantum eigenvalue transformation of unitary matrices, arXiv Preprint arXiv:2204.05955 (2022).

[32] G. Wang, D. Stilck-França, R. Zhang, S. Zhu, and P. D. Johnson, Quantum algorithm for ground state energy estimation using circuit depth with exponentially improved dependence on precision, arXiv Preprint arXiv:2209.06811 (2022).

[33] R. Zhang, G. Wang, and P. Johnson, Computing ground state properties with early fault-tolerant quantum computers, Quantum **6**, 761 (2022).

[34] E. Campbell, Random compiler for fast Hamiltonian simulation, Phys. Rev. Lett. **123**, 070503 (2019).

[35] P. K. Faehrmann, M. Steudtner, R. Kueng, M. Kieferová, and J. Eisert, Randomizing multi-product formulas for Hamiltonian simulation, Quantum **6**, 806 (2022).

[36] G. Wang, D. S. França, G. Rendon, and P. D. Johnson, Faster ground state energy estimation on early fault-tolerant quantum computers via rejection sampling, arXiv Preprint arXiv:2304.09827 (2023).

[37] A. M. Childs, A. Ostrander, and Y. Su, Faster quantum simulation by randomization, Quantum **3**, 182 (2019).

[38] S. Apers, A. Gilyén, and S. Jeffery, A unified framework of quantum walk search, 38th International Symposium on Theoretical Aspects of Computer Science (2021).

[39] C. Bravo-Prieto, R. LaRose, M. Cerezo, Y. Subasi, L. Cincio, and P. J. Coles, Variational quantum linear solver, arXiv Preprint arXiv:1909.05820 (2019).

[40] X. Xu, J. Sun, S. Endo, Y. Li, S. C. Benjamin, and X. Yuan, Variational algorithms for linear algebra, Sci. Bull. **66**, 2181 (2021).

[41] H.-Y. Huang, K. Bharti, and P. Rebentrost, Near-term quantum algorithms for linear systems of equations, arXiv Preprint arXiv:1909.07344 (2019).

[42] A. M. Childs and N. Wiebe, Hamiltonian simulation using linear combinations of unitary operations, Quantum Inf. Comput. **12**, 901 (2012).

[43] D. W. Berry, A. M. Childs, R. Cleve, R. Kothari, and R. D. Somma, in *Proceedings of the Forty-Sixth Annual ACM Symposium on Theory of Computing* (Association for Computing Machinery, New York, NY, USA, 2014), p. 283.

[44] D. W. Berry, A. M. Childs, and R. Kothari, in *2015 IEEE 56th Annual Symposium on Foundations of Computer Science* (IEEE, Berkeley, CA, USA, 2015), p. 792.

[45] D. W. Berry, A. M. Childs, R. Cleve, R. Kothari, and R. D. Somma, Simulating Hamiltonian dynamics with a truncated Taylor series, Phys. Rev. Lett. **114**, 090502 (2015).

[46] I. Kerenidis and A. Prakash, in *8th Innovations in Theoretical Computer Science Conference (ITCS 2017)* (Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, Berkeley, CA, USA, 2017).

[47] A. M. Childs, R. Kothari, and R. D. Somma, Quantum algorithm for systems of linear equations with exponentially improved dependence on precision, SIAM J. Comput. **46**, 1920 (2017).

[48] C. Cade and A. Montanaro, in *13th Conference on the Theory of Quantum Computation, Communication and Cryptography* (Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Sydney, Australia, 2018).

[49] S. Apers, S. Sen, and D. Szabó, A (simple) classical algorithm for estimating Betti numbers, arXiv Preprint arXiv:2211.09618 (2022).

[50] D. Janzing and P. Wocjan, A simple PromiseBQP-complete matrix problem, Theory Comput. **3**, 61 (2007).

[51] A. Montanaro and C. Shao, Quantum and classical query complexities of functions of matrices, arXiv Preprint arXiv:2311.06999 (2023).

[52] In the statement of results in the text, for simplicity we write $\mathcal{C}_{\text{sample}} = \widetilde{\mathcal{O}}(g)$ if $\mathcal{C}_{\text{sample}} = \mathcal{O}(g \operatorname{polylog} g \operatorname{polylog} h)$ and $\mathcal{C}_{\text{gate}} = \Omega(h)$, and vice versa for $\mathcal{C}_{\text{gate}}$. That is, with $\widetilde{\mathcal{O}}(\cdot)$ notation, we drop any terms that contribute polylogarithmically to the total (unparallelized) run time.

[53] X.-M. Zhang, T. Li, and X. Yuan, Quantum state preparation with optimal circuit depth: Implementations and applications, Phys. Rev. Lett. **129**, 230504 (2022).

[54] A. Ambainis, Variable time amplitude amplification and quantum algorithms for linear algebra problems, STACS'12 (29th Symposium on Theoretical Aspects of Computer Science) **14**, 636 (2012).

[55] Y. Subaşı, R. D. Somma, and D. Orsucci, Quantum algorithms for systems of linear equations inspired by adiabatic quantum computing, Phys. Rev. Lett. **122**, 060504 (2019).

[56] L. Lin and Y. Tong, Optimal polynomial based quantum eigenstate filtering with application to solving quantum linear systems, Quantum **4**, 361 (2020).

[57] D. An and L. Lin, Quantum linear system solver based on time-optimal adiabatic quantum computing and quantum approximate optimization algorithm, ACM Trans. Quantum Comput. **3**, 1 (2022).

[58] Y. Subaşı, R. D. Somma, and D. Orsucci, Quantum algorithms for systems of linear equations inspired by adiabatic quantum computing, Phys. Rev. Lett. **122**, 060504 (2019).

[59] E. Knill, G. Ortiz, and R. D. Somma, Optimal quantum measurements of expectation values of observables, Phys. Rev. A **75**, 012328 (2007).

[60] D. Wang, O. Higgott, and S. Brierley, Accelerated variational quantum eigensolver, Phys. Rev. Lett. **122**, 140504 (2019).

[61] In the tables (different from the text), for each cell we detail contributions from all parameters even if the contribution of that parameter to the total unparallelized run time is exponentially worse. This is to highlight the dependence of each parameter in each cell in isolation, so that each dependency can be compared across different schemes.

[62] G. H. Low and I. L. Chuang, Hamiltonian simulation by qubitization, Quantum **3**, 163 (2019).

[63] N.-H. Chia, A. P. Gilyén, T. Li, H.-H. Lin, E. Tang, and C. Wang, Sampling-based sublinear low-rank matrix arithmetic framework for dequantizing quantum machine learning, J. ACM **69**, 1 (2022).

[64] S. Chakraborty, A. Gilyén, and S. Jeffery, The power of block-encoded matrix powers: Improved regression techniques via faster Hamiltonian simulation, 46th International Colloquium on Automata, Languages, and Programming (ICALP 2019) (2019), https://arxiv.org/abs/1804.01973.

[65] This consists of oracles to coherently access the nonzero row and column entries, as well as an additional oracle to coherently access all matrix entries.

[66] D. Camps, L. Lin, R. Van Beeumen, and C. Yang, Explicit quantum circuits for block encodings of certain sparse matrice, arXiv Preprint arXiv:2203.10236 (2022).

[67] P. Zeng, J. Sun, and X. Yuan, Universal quantum algorithmic cooling on a quantum computer, arXiv Preprint arXiv:2109.15304 (2021).

[68] T. Keen, E. Dumitrescu, and Y. Wang, Quantum algorithms for ground-state preparation and Green's function calculation, arXiv Preprint arXiv:2112.05731 (2021).

[69] L. Lin and Y. Tong, Near-optimal ground state preparation, Quantum **4**, 372 (2020).

[70] Y. Ge, J. Tura, and J. I. Cirac, Faster ground state preparation and high-precision ground energy estimation with fewer qubits, J. Math. Phys. **60**, 022202 (2019).

[71] O. Kyriienko, Quantum inverse iteration algorithm for programmable quantum simulators, npj Quantum Inf. **6**, 1 (2020).

[72] T. A. Bespalova and O. Kyriienko, Hamiltonian operator approximation for energy measurement and ground-state preparation, PRX Quantum **2**, 030318 (2021).

[73] A. M. Childs, Y. Su, M. C. Tran, N. Wiebe, and S. Zhu, Theory of Trotter error with commutator scaling, Phys. Rev. X **11**, 011020 (2021).

[74] P. Zeng, J. Sun, L. Jiang, and Q. Zhao, Simple and high-precision Hamiltonian simulation by compensating Trotter error with linear combination of unitary operations, arXiv Preprint arXiv:2212.04566 (2022).

[75] J. Lee, D. W. Berry, C. Gidney, W. J. Huggins, J. R. McClean, N. Wiebe, and R. Babbush, Even more efficient quantum computations of chemistry through tensor hypercontraction, PRX Quantum **2**, 030305 (2021).

[76] S. Bravyi, A. Chowdhury, D. Gosset, and P. Wocjan, On the complexity of quantum partition functions, arXiv Preprint arXiv:2110.15466 (2021).

[77] P. Wocjan and K. Temme, Szegedy walk unitaries for quantum maps, arXiv Preprint arXiv:2107.07365 (2021).

[78] M.-H. Yung and A. Aspuru-Guzik, A quantum-quantum metropolis algorithm, Proc. Natl. Acad. Sci. **109**, 754 (2012).

[79] K. Temme, T. J. Osborne, K. G. Vollbrecht, D. Poulin, and F. Verstraete, Quantum metropolis sampling, Nature **471**, 87 (2011).

[80] C.-F. Chen, M. J. Kastoryano, F. G. Brandão, and A. Gilyén *et al.*, Quantum thermal state preparation, arXiv Preprint arXiv:2303.18224 (2023).

[81] C.-F. Chen and F. G. Brandão, Fast thermalization from the eigenstate thermalization hypothesis, arXiv Preprint arXiv:2112.07646 (2021).
[82] A. N. Chowdhury and R. D. Somma, Quantum algorithms for Gibbs sampling and hitting-time estimation, arXiv Preprint arXiv:1603.02940 (2016).
[83] J. Van Apeldoorn, A. Gilyén, S. Gribling, and R. de Wolf, Quantum SDP-solvers: Better upper and lower bounds, Quantum **4**, 230 (2020).
[84] J. van Apeldoorn and A. Gilyén, in *46th International Colloquium on Automata, Languages, and Programming (ICALP 2019)* (Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, Patras, Greece, 2019).
[85] Y. Tong, D. An, N. Wiebe, and L. Lin, Fast inversion, preconditioned quantum linear system solvers, fast Green's-function computation and fast evaluation of matrix functions, Phys. Rev. A **104**, 032422 (2021).
[86] Z. Holmes, G. Muraleedharan, R. D. Somma, Y. Subasi, and B. Şahinoğlu, Quantum algorithms from fluctuation theorems: Thermal-state preparation, Quantum **6**, 825 (2022).
[87] G. Giuliani and G. Vignale, *Quantum Theory of the Electron Liquid* (Cambridge University Press, Cambridge, UK, 2005).
[88] A. L. Fetter and J. D. Walecka, *Quantum Theory of Many-Particle Systems* (Dover Publications, Inc, Mineola, New York, 2012).
[89] B. Bauer, D. Wecker, A. J. Millis, M. B. Hastings, and M. Troyer, Hybrid quantum-classical approach to correlated materials, Phys. Rev. X **6**, 031045 (2016).
[90] D. Wecker, M. B. Hastings, N. Wiebe, B. K. Clark, C. Nayak, and M. Troyer, Solving strongly correlated electron models on a quantum computer, Phys. Rev. A **92**, 062318 (2015).
[91] J. Kreula, S. R. Clark, and D. Jaksch, Non-linear quantum-classical scheme to simulate non-equilibrium strongly correlated fermionic many-body dynamics, Sci. Rep. **6**, 1 (2016).
[92] J. M. Kreula, L. García-Álvarez, L. Lamata, S. R. Clark, E. Solano, and D. Jaksch, Few-qubit quantum-classical simulation of strongly correlated lattice fermions, EPJ Quantum Technol. **3**, 1 (2016).
[93] I. Rungger, N. Fitzpatrick, H. Chen, C. Alderete, H. Apel, A. Cowtan, A. Patterson, D. M. Ramo, Y. Zhu, and N. H. Nguyen *et al.*, Dynamical mean field theory algorithm and experiment on quantum computers, arXiv Preprint arXiv:1910.04735 (2019).
[94] S. Endo, I. Kurata, and Y. O. Nakagawa, Calculation of the Green's function on near-term quantum computers, Phys. Rev. Res. **2**, 033281 (2020).
[95] T. Kosugi and Y.-i. Matsushita, Construction of Green's functions on a quantum computer: Quasiparticle spectra of molecules, Phys. Rev. A **101**, 012330 (2020).
[96] A. Tosta, T. d. L. Silva, G. Camilo, and L. Aolita, Randomized semi-quantum matrix processing, arXiv Preprint arXiv:2307.11824 (2023).
[97] J. W. Demmel, *Applied Numerical Linear Algebra* (SIAM, Philadelphia, USA, 1997).
[98] G. H. Golub and C. F. Van Loan, *Matrix Computations* (JHU Press, Baltimore, USA, 2013).
[99] Y. Saad, *Iterative Methods for Sparse Linear Systems* (SIAM, Philadelphia, USA, 2003).
[100] B. D. Clader, B. C. Jacobs, and C. R. Sprouse, Preconditioned quantum linear system algorithm, Phys. Rev. Lett. **110**, 250504 (2013).
[101] C. Shao and H. Xiang, Quantum circulant preconditioner for a linear system of equations, Phys. Rev. A **98**, 062321 (2018).
[102] S. Chakraborty, Implementing linear combination of unitaries on intermediate-term quantum computers, arXiv Preprint arXiv:2302.13555 (2022).
[103] A. J. Walker, New fast method for generating discrete random numbers with arbitrary frequency distributions, Electron. Lett. **8**, 127 (1974).
[104] J. M. Kohler and A. Lucchi, in *International Conference on Machine Learning* (PMLR, Sydney, Australia, 2017), p. 1895.
[105] D. Gross, Recovering low-rank matrices from few coefficients in any basis, IEEE Trans. Inf. Theory **57**, 1548 (2011).
[106] M. Ledoux and M. Talagrand, *Probability in Banach Spaces: Isoperimetry and Processes* (Springer Science & Business Media, Berlin Heidelberg, Germany, 1991), Vol. 23.
[107] S. Aaronson and D. Gottesman, Improved simulation of stabilizer circuits, Phys. Rev. A **70**, 052328 (2004).
[108] C. Gidney, STIM: A fast stabilizer circuit simulator, Quantum **5**, 497 (2021).
[109] R. L. Stratonovich, A method for the computation of quantum distribution functions, Doklady Akademii Nauk **115**, 1097 (1957).
[110] J. Hubbard, Calculation of partition functions, Phys. Rev. Lett. **3**, 77 (1959).
[111] N. A. Nghiem and T.-C. Wei, Quantum algorithm for estimating eigenvalue, arXiv:2211.06179 [quant-ph] (2023).
[112] K. Seki and S. Yunoki, Quantum power method by a superposition of time-evolved states, PRX Quantum **2**, 010333 (2021).
[113] C. T. Hann, G. Lee, S. Girvin, and L. Jiang, Resilience of quantum random access memory to generic noise, PRX Quantum **2**, 020311 (2021).

Correction: The email address in the byline footnote for the first author contained a typographical error and has been fixed.