Implementing any Linear Combination of Unitaries on Intermediate-term Quantum Computers

Shantanav Chakraborty

CQST and CSTAR, International Institute of Information Technology Hyderabad, Telangana, India

We develop three new methods to implement any Linear Combination of Unitaries (LCU), a powerful quantum algorithmic tool with diverse applications. While the standard LCU procedure requires several ancilla qubits and sophisticated multi-qubit controlled operations, our methods consume significantly fewer quantum resources. The first method (*Single-Ancilla LCU*) estimates expectation values of observables with respect to any quantum state prepared by an LCU procedure while requiring only a single ancilla qubit, and no multi-qubit controlled operations. The second approach (*Analog LCU*) is a simple, physically motivated, continuous-time analogue of LCU, tailored to hybrid qubit-qumode systems. The third method (*Ancilla-free LCU*) requires no ancilla qubit at all and is useful when we are interested in the projection of a quantum state (prepared by the LCU procedure) in some subspace of interest. We apply the first two techniques to develop new quantum algorithms for a wide range of practical problems, ranging from Hamiltonian simulation, ground state preparation and property estimation, and quantum linear systems. Remarkably, despite consuming fewer quantum resources they retain a provable quantum advantage. The third technique allows us to connect discrete and continuous-time quantum walks with their classical counterparts. It also unifies the recently developed optimal quantum spatial search algorithms in both these frameworks, and leads to the development of new ones that require fewer ancilla qubits. Overall, our results are quite generic and can be readily applied to other problems, even beyond those considered here.

Shantanav Chakraborty: shchakra@iiit.ac.in

Contents

1 Introduction

We are currently in an era of quantum computing where theoretical advancements have been accompanied by drastic improvements in experimental capabilities [1, 2, 3, 4, 5]. With rapid progress being made, it is reasonable to envision a stage in the near future, where quantum computing will transition away from the NISQ era [6, 7]. Quantum devices available immediately after the current NISQ stage, will most likely not have the capabilities of a large-scale, fully-programmable, fault-tolerant quantum computer. These devices, known as *early fault-tolerant quantum computers* [8, 9, 10, 11, 12, 13], would have a limited number of logical qubits (restricting the availability of ancilla qubits), and short depth circuits with little to no multi-qubit controlled gates. On the other hand, for particular quantum technological platforms, it might be possible to engineer certain specific interactions more precisely, and for longer time-scales than others. For instance, it might be easier to engineer hybrid qubit-qumode systems in the intermediate-term [14, 15, 16, 17, 3, 18, 19] as many of the most promising quantum technological platforms such as superconducting systems [14], ion-traps [20], and photonic systems [21], naturally have access to continuous variables. We refer to such devices, which will become available shortly after the current stage, as "intermediate-term quantum computers".

It is thus crucial to develop quantum algorithms of practical interest that are implementable on intermediate-term quantum computers. Indeed, quantum algorithms tailored to early fault-tolerant quantum computers are already being developed [8, 9, 10, 11, 12, 13]. With many quantum technological platforms vying for supremacy, it is also essential to develop physically motivated quantum algorithms that can exploit the degrees of freedom that are naturally available to such platforms.

There are only a handful of quantum algorithmic frameworks that can be applied to a diverse range of problems. However, most of these are only implementable on fully-fault tolerant quantum computers, which might be decades away. Linear Combination of Unitaries (LCU) is one such paradigm. Over the years, it has been widely applied and has been central to the development of a plethora of useful quantum algorithms ranging from Hamiltonian simulation [22, 23, 24, 25], quantum linear systems [26, 27] and differential equations [28, 29, 30], quantum walks [31, 32, 33], ground state preparation [34, 35, 36] and a large-class of optimization problems [37, 38].

The wide applicability of this procedure stems from the generic settings in which it can be applied. Given a Hermitian matrix H, and an initial state ρ_0, the LCU procedure implements any function $f(H)$ that can be well-approximated by a linear combination of unitaries, i.e. $f(H) \approx \sum_j c_j U_j$. Specifically, it prepares the quantum state

$$\rho = \frac{f(H)\rho_0 f(H)^\dagger}{\mathrm{Tr}[f(H)\rho_0 f(H)^\dagger]}, \tag{1}$$

using U_j, controlled over multiple ancilla qubits [1]. In fact, for most of the applications mentioned previously, the problem boils down to applying a specific $f(H)$, to some initial state. Despite its broad applicability, LCU has its drawbacks when it comes to being implementable in the intermediate-term. First, for many problems of interest, there is a significant overhead in terms of the number of ancilla qubits needed. Second, the procedure requires implementing a sequence of sophisticated multi-qubit controlled-unitary operations, which is challenging for intermediate-term quantum computers. Furthermore, simply preparing the quantum state ρ is often not useful. In most practical scenarios, we

[1] In order to prepare ρ, several rounds of amplitude amplification is also required

are interested in estimating some property of ρ, such as the expectation value of some observable O, i.e. $\mathrm{Tr}[O\rho]$. Extracting useful information about ρ either requires additional runs of the LCU procedure or leads to even deeper quantum circuits.

In this work, we significantly enhance the applicability of the LCU framework to intermediate-term quantum devices. We develop three approaches that seek to reduce the resource required to implement any LCU. These methods either prepare the state ρ, or help extract useful information from it. Being considerably simpler than the standard LCU procedure, they are suitable for implementation using intermediate-term devices such as early fault-tolerant quantum computers and hybrid qubit-qumode systems. We apply each of them to develop quantum algorithms of practical interest.

Firstly, we develop a randomized quantum algorithm that estimates properties of the state ρ, prepared by any LCU procedure. More precisely, for any observable O, our algorithm estimates the quantity $\mathrm{Tr}[O\rho]$, to arbitrary accuracy. This technique, which we refer to as *Single-Ancilla LCU*, requires only one ancilla qubit that acts as a control, and implements two (controlled) unitaries sampled according to the distribution of the LCU coefficients, followed by a single-shot measurement. By repeatedly running this simple short-depth quantum circuit, one obtains samples whose average converges to the expectation value we seek to estimate. Our procedure is suitable for early fault-tolerant quantum devices as it can implement any LCU (a) using only a single ancilla qubit, and (b) no multi-qubit controlled gates. In contrast, as mentioned previously, the standard LCU method requires several ancilla qubits and a series of sophisticated (multi-qubit) controlled operations. We rigorously compare the cost of implementing *Single-Ancilla LCU* with the generic LCU procedure, and show that each coherent run of our method costs less, while requiring more classical runs. Furthermore, we apply our method to develop novel quantum algorithms for Hamiltonian simulation, estimating the properties of ground states of Hamiltonians, as well as quantum linear systems.

Secondly, we develop *Analog LCU*, a physically motivated, continuous-time analogue for implementing a linear combination of unitaries. This technique requires coupling the system Hamiltonian H to a continuous-variable ancilla system (such as a one-dimensional quantum Harmonic oscillator), initialized in some easy-to-prepare continuous-variable quantum state (such as a Gaussian). The overall system is then evolved according to the resulting interaction Hamiltonian. Although this approach requires a continuous-variable ancilla register, the overall algorithm is considerably simpler than the standard LCU procedure. Moreover, this technique might be particularly useful for intermediate-term quantum computers (e.g. hybrid qubit-qumode systems) as such interactions can already be engineered on several quantum technological platforms. Examples of discrete systems coupled to continuous-variable ones include ion traps and superconducting systems [14, 15, 17, 16, 18]. We show that this approach can be used to develop novel analog quantum algorithms for ground state preparation and solving quantum linear systems.

Suppose for a specific problem, we are interested in the projection of the LCU state $f(H)\,|\psi_0\rangle$ in some subspace, and it suffices to ensure that the measurement is successful, on average. In such scenarios, we show that the ancilla registers can be dropped entirely. We call this the *Ancilla-free LCU* technique. This approach involves randomly sampling the unitaries U_j according to the distribution of the LCU coefficients $c_j/\|c\|_1$ without any ancilla registers. On average, this prepares some density matrix for which the projection in this subspace can be proven to be at least as large. This scenario arises in the context of quantum spatial search algorithms: the problem of finding an element in a marked subset of nodes of any ergodic, reversible, Markov chain. Indeed, the goal is to prove a generic quadratic speedup over classical random walks, for which, the expected number of steps

to solve this problem is known as the hitting time (HT). This problem has only recently been resolved using generic LCU [32], following a long line of works that provided speedups in particular cases [39, 40, 41]. Consequently, we use *Ancilla-free LCU* to design optimal spatial search algorithms, also placing recent results [32, 42, 33] within this framework, along with developing new ones. As compared to quantum spatial search algorithms using *Standard LCU*, our methods achieve the same generic quadratic speedup while requiring $O(\log HT)$ fewer ancilla qubits.

In addition to providing a unified framework for quantum spatial search, *Ancilla-free LCU* also allows us to establish a relationship between the different frameworks of classical random walks and quantum walks. Finally, in order to complete the picture, we also establish a connection between discrete and continuous-time quantum walks by using the frameworks of block encoding [43, 27] and quantum singular value transformation (QSVT) [44, 45, 46].

The paper is organized as follows. In the rest of this section, we provide a brief overview of the main results in Sec. 1.1, and also relate our work to prior results in Sec. 1.2. In Sec. 2, we review some basic definitions and techniques that we will be using in this article. We formally describe the three different approaches to implementing LCU in Sec. 3. The rest of the article involves applying these techniques to develop new quantum algorithms. In Sec. 4, we apply the *Single-Ancilla LCU* method to develop a novel quantum algorithm for Hamiltonian simulation. In Sec. 5, we make use of our techniques to develop new quantum algorithms for ground state preparation of Hamiltonians (Sec. 5.1) and also for ground state property estimation (Sec. 5.2). In Sec. 6, we develop novel analog quantum linear systems algorithms, tailored to hybrid qubit-qumode systems (Sec. 6.1) and also use *Single-Ancilla LCU* to estimate expectation values with respect to the solution of quantum linear systems (Sec. 6.2). In Sec. 7, we apply *Ancilla-free LCU* we establish a relationship between different frameworks of classical and quantum walks by developing optimal quantum spatial search algorithms that reduce the number of ancilla qubits needed, also placing recently developed algorithms within this framework. Finally, we conclude and discuss possible future research directions in Sec. 8.

1.1 Summary of our results

In this section, we state the main results of this article. We begin by briefly outlining each of the three variants of implementing LCU and the applications we consider. We summarize them in Fig. 1.

1.1.1 Single-Ancilla LCU: Estimating expectation values of observables

Given any initial state ρ_0, and Hamiltonian H, we develop a randomized quantum algorithm that estimates the expectation value

$$\mathrm{Tr}[O\rho] = \frac{\mathrm{Tr}[O\ f(H)\rho_0 f(H)^\dagger]}{\mathrm{Tr}[f(H)\rho_0 f(H)^\dagger]}, \tag{2}$$

to arbitrary accuracy, for any function f that can be approximated by a linear combination of unitaries, i.e. $f(H) \approx \sum_j c_j U_j$. For this task, the standard LCU procedure requires several ancilla qubits, and sophisticated (multi-qubit) controlled unitaries. In contrast, for quantum algorithms tailored to early fault-tolerant quantum devices, both the number of ancilla qubits, as well as the number of multi-qubit controlled operations should be as low as possible. Given these restrictions, the priority of such algorithms is to reduce the cost of each coherent run, even if this results in an increase in the number of classical

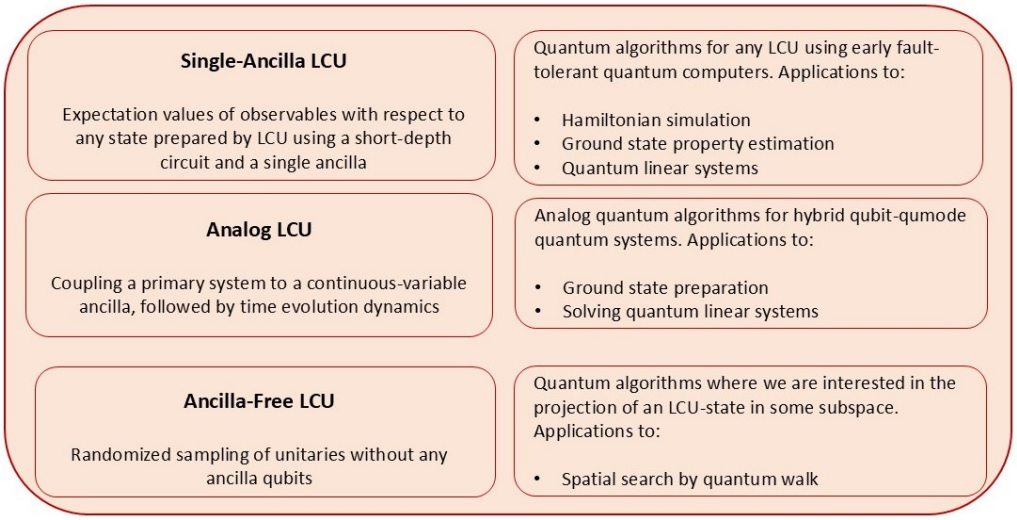

Figure 1: *Summary of the main results – The three approaches to LCU and their applications*

repetitions required (and hence, the overall cost). As quantum coherence need not be maintained across multiple runs, this leads to having multiple repetitions of a low-cost quantum circuit (in terms circuit/gate depth), which is preferred in the intermediate regime.

The *Single-Ancilla LCU* technique estimates the expectation value in Eq. (2) while satisfying the aforementioned features: (a) it uses only a single ancilla qubit, (b) requires no multi-qubit controlled operations, and (c) despite restrictions (a) and (b), the cost of each coherent run is lower than the *Standard LCU* procedure. However, it requires more classical repetitions as compared to the generic LCU technique, and hence has a higher overall cost. To this end, we develop a randomized quantum algorithm makes use of the quantum circuit of Faerhmann et al. [9] (shown in Fig. 2), wherein the authors used it to generate randomized multi-product formulas. For any $f(H)$ that can be approximated by an LCU, our method repeatedly samples from this circuit to estimate the numerator of Eq. (2). Note that the denominator is apriori unknown. We show that the knowledge of any rudimentary lower bound ℓ_* of this quantity allows us to leverage the same algorithm to estimate it [2]. Overall, our algorithm separately estimates both the numerator as well as the denominator and we rigorously calculate the accuracy with both these quantities need to be estimated so that their ratio is ε-close the expectation value we seek to estimate. Our overall procedure, and its and its correctness has been proven in detail in Sec. 3.2. Here, we state the general result informally.

Theorem 1 (Informal)**.** *Let O be some observable and ρ_0 be some initial state, prepared in cost τ_{ρ_0}. Suppose there exists a Hermitian matrix $H \in \mathbb{C}^{N\times N}$, and a function $f : [-1,1] \mapsto \mathbb{R}$ such that $f(H) \approx \sum_j c_j U_j$, where $\|c\|_1 = \sum_j |c_j|$, and each U_j is implemented with cost τ_j. Define $\langle\tau\rangle = \sum_j c_j\tau_j/\|c\|_1$. Also, suppose we know some ℓ_* such that $\ell^2 = \mathrm{Tr}[f(H)\rho_0 f(H)^\dagger] \geq \ell_*$. Then there exists a procedure that outputs μ and $\tilde{\ell}$ such that*

$$\left|\mu/\tilde{\ell} - \mathrm{Tr}[O\rho]\right| \leq \varepsilon,$$

[2] Apriori knowledge of ℓ_* is application specific. For instance, the minimum eigenvalue of $f(H)$ can be such a lower bound. Note that this information is also needed in case of *Standard LCU*

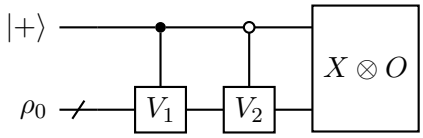

Figure 2: *Quantum circuit corresponding for the Single-Ancilla LCU procedure. For* $f(H) \approx \sum_j c_j U_j$, *repeated runs of this short-depth quantum circuit can estimate* $\mathrm{Tr}[O\, f(H)\rho_0 f(H)^\dagger]/\mathrm{Tr}[f(H)\rho_0 f(H)^\dagger]$, *to arbitrary accuracy. For this,* V_1 *and* V_2 *are sampled at random according to* $\mathcal{D} \sim \{c_j/\|c\|_1, U_j\}$. *Each run of the circuit outputs a random variable corresponding to the outcome of the measurement of the observable* $X \otimes O$. *Overall we need to repeat this circuit* T *times, such that the sample mean of the* T *observations to converge to the desired estimate.*

with a constant probability, using only one ancilla qubit and

$$T = O\left(\frac{\|O\|^2 \|c\|_1^4}{\varepsilon^2 \ell_*^2}\right)$$

repetitions of the quantum circuit in Fig. 2, where the average cost of each such run is $2\langle\tau\rangle + \tau_{\rho_0}$.

We compare the performance of *Single-Ancilla LCU* with *Standard LCU* in detail (See Sec. 3.2 and Table 1). We show that if implementing each U_j costs τ_j, and preparing the initial state ρ_0 costs τ_{ρ_0}, the average cost of each coherent run of our algorithm is $2\langle\tau\rangle + \tau_{\rho_0}$. On the other hand, *Standard LCU* requires implementing a *prepare* gate R and a multi-qubit controlled *select* unitary Q, requiring cost τ_R and τ_Q respectively. So the total cost is $O(\tau_R + \tau_Q + \tau_{\rho_0})$. Then just τ_Q is at least as high as $\langle\tau\rangle$. Thus, despite requiring a solitary ancilla qubit and no multi-qubit controlled operations, the cost of each coherent run of our algorithm is lower than *Standard LCU*, given both procedures implement U_j with the same cost. However, the number of classical repetitions (and hence, the overall cost) required for *Standard LCU* is lower. Moreover, being suitable for fully fault-tolerant quantum computers, *Standard LCU* can also leverage procedures such as quantum amplitude amplification and estimation [47] to estimate $\mathrm{Tr}[O\rho]$ coherently, which further reduces the overall cost by requiring fewer classical repetitions, while increasing the cost of each coherent run substantially. However, procedures such as quantum amplitude amplification and estimation are too involved to be implemented in the intermediate-term. Next, we discuss the applications of our method.

Applications: We apply *Single-Ancilla LCU* to several problems of practical interest such as Hamiltonian simulation, ground state property estimation and quantum linear systems. Throughout the article, for each of the applications of *Single-Ancilla LCU*, we calculate T, the number of repetitions needed of the circuit in Fig. 2 as well as an upper bound on the cost of each coherent run, given by $\tau_{\max}$. The overall cost would then be the product of these quantities that is $O(\tau_{\max} \cdot T)$.

For all our applications, we consider some Hamiltonian H, and implement a specific $f(H) \approx \sum_j c_j U_j$. However, depending on the application, how we can access H (input model) varies. This also determines what the costs, $\tau_{\max}$ and T characterize. Before moving on to the specific applications, we discuss them briefly:

- For Hamiltonian simulation, we assume H is a linear combination of strings of Pauli matrices, which can be accessed and implemented directly. We characterize the cost in terms of the gate depth per coherent run ($\tau_{\max}$) as well as the overall gate depth $O(T \cdot \tau_{\max})$.

- For both ground state property estimation as well as quantum linear systems, we assume that we can access the Hamiltonian H through the time evolution operator $U_t = \exp[-itH]$. Furthermore, given access to U_t, we can perform the time evolution controlled on a single ancilla qubit. This is referred to as the Hamiltonian evolution (HE) model as has been used in prior results specific to ground energy estimation using early fault-tolerant quantum computers [12, 11, 13, 10]. Much like these works, we calculate: (a) the maximal time of evolution of H (controlled by a single ancilla qubit) required in each coherent run, which will be denoted by $\tau_{\max}$, and (b) the total number of repetitions of the circuit T. The total evolution time is then $O(\tau_{\max} \cdot T)$. Note that $\tau_{\max}$ is different from the actual circuit depth. In fact, this relationship depends on how the time evolution is performed. If U_t can be performed exactly in time $O(t)$, then the circuit depth scales linearly with the maximal evolution time. However, if U_t is implemented by a Hamiltonian simulation algorithm, then the circuit depth depends on the particular choice of the algorithm. Recall that in the early fault-tolerant regime, we are limited by a small ancilla qubit space and the inability to perform multi-qubit controlled operations. This restricts the choice for the underlying simulation algorithm. In Appendix C, we indeed characterize the complexities of both our algorithms in terms of the circuit depth (more precisely, the gate depth) by choosing particular Hamiltonian simulation algorithms that are suited to this regime.

Having discussed the various access models we consider, we move on to the specific applications to which we apply *Single-Ancilla LCU*:

(a) **Hamiltonian simulation:** Consider any Hamiltonian H such that it is expressed as a linear combination of strings of Pauli operators P_k. That is, $H = \sum_{k=1}^{L} \lambda_k P_k$, such that $\beta = \sum_{k=1}^{L} |\lambda_k|$. If ρ_0 is some initial state prepared in cost τ_{ρ_0}, then our randomized quantum algorithm outputs a parameter μ, with probability at least $1 - \delta$, such that

$$\left|\mu - \mathrm{Tr}[O\ e^{-iHt}\rho_0 e^{iHt}]\right| \leq \varepsilon,$$

for any measurable observable O. Each coherent run of our algorithm makes use of the quantum circuit in Fig. 2 which uses only a single ancilla qubit. The gate depth is at most $\tau_{\rho_0} + 2\tau_{\max}$, where

$$\tau_{\max} = O\left(\beta^2 t^2 \frac{\log(\beta t \|O\|/\varepsilon)}{\log\log(\beta t \|O\|/\varepsilon)}\right).$$

Overall,

$$T = O\left(\frac{\|O\|^2 \ln(1/\delta)}{\varepsilon^2}\right)$$

classical repetitions of this quantum circuit is needed. For this, we decompose $f(H) = e^{-itH}$ as a linear combination of unitaries using ideas from the truncated Taylor series approach [25], as well as from [13]. We describe our method in detail in Sec. 4. The overall gate depth is given by $T \cdot \tau_{\max} = \widetilde{O}\left(\beta^2 t^2 \|O\|^2/\varepsilon^2\right)$.

Our method outperforms all first order Trotter methods [48] and their randomized variants [49, 50], requiring an exponentially shorter gate depth per coherent run (in terms of the precision $1/\varepsilon$), as well as a shorter overall gate depth. Our algorithm also uses fewer ancilla qubits than the Truncated Taylor series method [25], which additionally makes multiple uses of complicated subroutines such as oblivious amplitude amplification. The gate depth per coherent run of this algorithm is quadratically better than our method in terms of β and t, but has a linear dependence on $O(L)$, which implies that there are Hamiltonians (satisfying $\beta \ll L$) for which our method provides an advantage. The state-of-the-art Hamiltonian simulation procedure by Low and Chuang [43] is optimal in terms of the number of queries to a block encoding of H. However, the construction of the block encoding requires $O(\log L)$ ancilla qubits and multi-qubit controlled operations, and hence is not implementable in the early fault-tolerant regime. This also adds an overhead of $O(L)$ to the gate depth per coherent run. We compare our algorithm rigorously with other methods in Sec. 4. A comparison of the complexities are summarized in Table 2.

(b) **Ground state property estimation:** For any Hamiltonian H with unknown ground state $|v_0\rangle$, and any measurable observable O, we provide a randomized quantum algorithm that outputs an ϵ-additive accurate estimate of the expectation value of O with respect to $|v_0\rangle$, i.e. $\langle v_0|O|v_0\rangle$. For this, we make the following standard assumptions: (i) a lower bound on spectral gap of H is known (say Δ), (ii) an initial state $|\psi_0\rangle$ having overlap of at least η with $|v_0\rangle$, can be prepared in cost τ_{ψ_0}, (iii) the ground energy of H is known to precision $O(\Delta \cdot (\log(\varepsilon^{-1}\eta^{-1}))^{-1/2})$. Our algorithm involves expressing the function $f(H) = e^{-tH^2}$ as a LCU: we show $f(H) = \sum_j c_j e^{-ij\sqrt{2t}H}$. For a judiciously chosen value of t, $f(H)\,|\psi_0\rangle$, has the effect of preserving the component of $|\psi_0\rangle$ in the direction of $|v_0\rangle$ while exponentially suppressing all other components that are orthogonal to it, resulting in a state that is close to $|v_0\rangle$. In the Hamiltonian evolution model (H is accessed via the time evolution operator), we show that the *Single-Ancilla LCU* algorithm estimates $\langle v_0|O|v_0\rangle$ with additive accuracy ε, with probability at least $(1-\delta)^2$, using

$$T = O\left(\frac{\|O\|^2 \ln(1/\delta)}{\varepsilon^2 \eta^4}\right),$$

repetitions of the quantum circuit in Fig. 2 and only a single ancilla qubit. The maximal time evolution of H in any coherent run is at most $2\tau_{\max} + \tau_{\psi_0}$, where

$$\tau_{\max} = O\left(\frac{1}{\Delta}\log\left(\frac{\|O\|}{\varepsilon\eta}\right)\right).$$

The total evolution time is then $O(T.\tau_{\max}) = \widetilde{O}(\Delta^{-1}\eta^{-4}\|O\|^2/\varepsilon^2)$. The overall method and its correctness has been formally stated via Theorem 15. We also compare our method with other algorithms in detail (See Table 3). The *Standard LCU* procedure [34, 35] requires $O(\log(1/\Delta) + \log(\,\|O\|\,\eta^{-1}\varepsilon^{-1}))$ ancilla qubits and sophisticated multi-qubit controlled operations. Despite this, in the Hamiltonian Evolution access model, the maximal time evolution of H, for *Standard LCU* is never better than our method. However, *Standard LCU* requires fewer classical runs (by a factor of $1/\eta^2$) and hence has a lower total evolution time. It can also make use of involved subroutines such as quantum amplitude amplification and estimation to

further lower the total evolution time. However these methods substantially increase the maximal time of evolution of H. Furthermore, it also increases the number of ancilla qubits needed. Our algorithm also compares favourably to other early fault-tolerant quantum algorithms [11, 10].

(c) **Quantum linear systems:** Suppose we have a Hermitian matrix H such that its eigenvalues in $[-1, -1/\kappa] \cup [1/\kappa, 1]$, such that κ is an upper bound on the ratio between the maximum and minimum eigenvalue of H (condition number). Let us assume that the initial quantum state $|b\rangle$ can be prepared in cost τ_b. Then, we show use *Single-Ancilla LCU* to estimate $\langle x|O|x\rangle$ up to ε additive accuracy, with probability at least $(1-\delta)^2$, using

$$T = O\left(\frac{\|O\|^2 \kappa^4 \log^2\left(\frac{\|O\|\kappa}{\varepsilon}\right) \ln(1/\delta)}{\varepsilon^2}\right),$$

repetitions of the quantum circuit in Fig. 2 and only one ancilla qubit. The maximal time evolution of H is at most $2\tau_{\max} + \tau_b$, where

$$\tau_{\max} = O\left(\kappa \log\left(\frac{\|O\|\kappa}{\varepsilon}\right)\right).$$

The total evolution time is given by $\widetilde{O}(\kappa^5 \|O\|^2/\varepsilon^2)$. This approach makes use of the LCU decomposition of $f(H) = H^{-1}$ in Ref. [26]. We analyze the correctness of our method via Theorem 18. The Childs, Kothari and Somma algorithm [26], which makes use of *Standard LCU* requires $O(\kappa \log(\kappa\|O\|)/\varepsilon)$ ancilla qubits and sophisticated multi-qubit controlled operations. Despite this, the maximal time evolution of H is never better than our method. The number of classical repetitions (and hence the total evolution time), however scales better than our method. The quantum linear systems algorithm using QSVT [45] requires access to a block encoding of H. This has near optimal complexity in terms of the number of queries to the block encoding. However, constructing a block encoding of H is resource consuming and is not implementable in the intermediate-term. This is also the case with the state-of-the-art algorithm of Costa et al. [51] which has optimal query complexity on all parameters (the query depth per coherent run is $O(\log \kappa)$ better than the maximal time evolution of H in our case). In Sec. 6.2, we compare our method in detail with other algorithms for solving quantum linear systems (See Table 4).

As discussed previously, we analyzed the complexity of both the ground state property estimation algorithm as well as the quantum linear systems algorithm in the Hamiltonian Evolution input model, which is indeed the case for other early fault-tolerant quantum algorithms. In this case, $\tau_{\max}$ measures the maximal time evolution of H, while $O(\tau_{\max}.T)$ is the total evolution time. However, this is different from the actual circuit depth of the algorithm, for which one needs to specify how the time-evolution operator $U_t = \exp[-itH]$ is implemented. If this is performed by a Hamiltonian simulation algorithm, both the circuit depth per coherent run, as well as the overall circuit depth depends on the choice of the underlying simulation algorithm. In the early fault-tolerant regime, we intend to leverage algorithms that do not add any overhead in terms of the number of ancilla

qubits or multi-qubit controlled gates. This limits the Hamiltonian simulation algorithms that can be implemented in the early fault-tolerant era. For instance, state-of-the-art Hamiltonian simulation algorithms [52, 43] require access to a block encoding of H: a unitary where H is in the top left block subnormalized by some factor (See Sec. 2.3). This is resource demanding as it increases the number of ancilla qubits, as well as multi-qubit controlled operations and hence are unsuitable for early fault-tolerant quantum devices.

On the other hand, if H can be expressed as a linear combination of Pauli operators, both Trotter-based methods [48] as well as the Hamiltonian simulation algorithm by *Single-Ancilla LCU* can be incorporated into both our algorithms. This does not require any additional ancilla qubits or multi-qubit controlled gates. However, both these methods have a suboptimal dependence on t, which increases the circuit depth of both our ground state property estimation and quantum linear systems algorithms.

In Appendix C.2, we analyze the circuit depth (in terms of the gate depth per coherent run, as well as the overall gate depth) of both these algorithms while using (a) Hamiltonian simulation by *Single-Ancilla LCU* and (b) $2k$-order Trotter [48]. For this we assume that:

(i) H is a linear combination of strings of Pauli operators, i.e. $H = \sum_{j=1}^{L} \lambda_j P_j$, with $\beta = \sum_{k=1}^{L} |\lambda_k|$.

(ii) The observable O to be measured is also a linear combination of easy-to-implement unitary observables, i.e. $O = \sum_{j=1}^{L_O} h_j O_j$ such that $\|O_j\| = 1$ and $\|h\|_1 = \sum_j |h_j|$. This is motivated by the fact that such observables are ubiquitous across condensed matter physics [53] and quantum chemistry [54].

We demonstrate that under these assumptions both algorithms can still be performed using only a single ancilla qubit and no multi-qubit controlled operations (See Table A1). We comprehensively compare our methods with other quantum algorithms all of which require multiple ancilla qubits and multi-qubit controlled operations. Furthermore, we show that despite this, there are regimes where both our algorithms require a shorter gate depth per coherent run, as compared to even state-of-the-art quantum algorithms (See Table A2 for ground state property estimation and Table A3 for quantum linear systems). Next, we discuss our results in the *Analog LCU* framework.

1.1.2 Analog LCU: Coupling discrete systems with continuous variable systems

We develop a more physical model for LCU in continuous-time. Consider any Hamiltonian H, consider any $f(H)$ that can be well approximated by a truncated Fourier transform, i.e.,

$$\left\| f(H) - \int_a^b dz\ c(z) \cdot e^{-iHzt} \right\| \leq \varepsilon,$$

where $c : \mathbb{R} \mapsto \mathbb{R}\backslash\{0\}$. Then by a purely continuous-time procedure, for any initial state $|\psi_0\rangle$, we can prepare a state that $O\left(\varepsilon/\|c\|_1\right)$-close to $f(H)|\psi_0\rangle / \|f(H)|\psi_0\rangle\|$, where $\|c\|_1 = \int_a^b dz\ |c(z)|$.

This requires coupling the primary system to a continuous variable ancilla (such as a one-dimensional quantum Harmonic oscillator), prepared in a continuous variable state. We show in this work that for several applications, this state is easy to prepare (such as a Gaussian). The overall system is then evolved according to the interaction Hamiltonian $H' = H \otimes \hat{z}$ for an appropriate time T, which prepares the desired LCU state in the first register.

The technique is considerably simpler than its discrete-time counterpart. Furthermore such hybrid qubit-qumode interactions can be implemented in a number of quantum technological platforms such as trapped ions, Cavity (or Circuit QED), photonic systems, and superconducting qubits [14, 15, 17, 16, 18]. Our motivation is not only to provide an alternative approach to implementing LCU but make this paradigm more implementable for intermediate-term quantum devices.

Applications: We apply these techniques to develop analog quantum algorithms for ground state preparation and for solving quantum linear systems. As mentioned previously, our aim is to couple the system Hamiltonian with an ancillary continuous-variable system.

(a) **Ground state preparation:** Given a Hamiltonian H, we couple this system with a continuous variable ancillary system via the interaction Hamiltonian $H' = H \otimes \hat{z}$. The ancilla system is prepared in an easy-to-prepare continuous variable state, namely a Gaussian. We show that given an initial state $|\psi_0\rangle$ that has an overlap of at least η with the ground state, simply evolving the system according to H' results in a state proportional to $f(H)|\psi_0\rangle$ in the first register, where $f(H) = e^{-tH^2}$. We show that, with probability η^2, this state is ε-close to the ground state of H (provided its ground energy is known up to some precision). The overall time required is

$$T = O\left(\frac{1}{\Delta}\sqrt{\log\left(\frac{1}{\eta\varepsilon}\right)}\right),$$

where Δ is the spectral gap of H (Lemma 13). This quantum algorithm appeared in [33] and also independently in Ref. [36]. Here we place this in the context of *Analog LCU*: it provides useful intuition for (i) the quantum linear systems algorithms we develop using similar techniques and (ii) the *Single-Ancilla LCU* method for ground state property estimation.

(b) **Quantum linear systems:** We provide two quantum algorithms for this problem. For both these problems, we couple H to two ancillary continuous variable systems (Harmonic oscillators), i.e. $H' = H \otimes \hat{y} \otimes \hat{z}$. The first approach works for any Hermitian matrix H with eigenvalues in the domain $[-1, -1/\kappa] \cup [1/\kappa, 1]$, where κ is an upper bound on the condition number (ratio between the maximum and the minimum non-zero eigenvalue) of H. The first register is prepared in the initial state $|b\rangle$, the second register is prepared in the first excited state of the quantum Harmonic oscillator, while the third register is prepared in the ground state of a "particle in a ring" of unit radius [55].

This algorithm (see Sec. 6.1) can be seen as an analog variant of the quantum linear systems algorithm of Childs, Kothari and Somma [26]. In order to obtain a quantum state that is ε/κ-close to $|x\rangle = H^{-1}|b\rangle / \|H^{-1}|b\rangle\|$ in the first register, with overlap at least $1/\kappa$, we require evolving the system according to H' for a time

$$T = O\left(\kappa\sqrt{\log\left(\frac{\kappa}{\varepsilon}\right)}\right).$$

Typically in continuous variable systems, Gaussian states are easier to prepare, and engineer [56]. Thus, we also provide an analog quantum algorithm for solving quantum linear systems (for positive semidefinite Hamiltonians) in which both the ancilla

registers are now prepared in Gaussian states. Evolving this system according to H' prepares a state that is $(\varepsilon/\kappa)^{3/2}$ - close to $|x\rangle$, with overlap $\Omega(1/T)$ in time

$$T = O\left(\frac{\kappa^{3/2}}{\sqrt{\varepsilon}}\right).$$

Although the complexity is worse than the first approach, this quantum algorithm requires preparing only Gaussian states, which we expect to be easier for intermediate-term quantum computers to implement.

1.1.3 Ancilla - free LCU: Randomized sampling of unitaries

Suppose for some Hermitian matrix $H \in \mathbb{C}^{N\times N}$, we intend to implement $f(H)$ such that

$$\left\| f(H) - \sum_{j=1}^{M} c_j U_j \right\| \leq \gamma,$$

where $\gamma \in [0,1)$, $c_j \in \mathbb{R}$ and U_j is some unitary. Furthermore, we are interested in the projection of $f(H)\,|\psi_0\rangle$ in some subspace of interest and for the underlying problem, it suffices to ensure that the projective measurement is successful, on average. Then, dropping the ancilla register altogether and simply sampling V according to $\mathcal{D} \sim \{c_j/\|c\|_1, U_j\}$, results in the following mixed state, on average

$$\bar{\rho} = \mathbb{E}[V\rho_0 V^\dagger] = \frac{1}{\|c\|_1}\sum_{j=1}^{M} c_j U_j \rho_0 U_j^\dagger,$$

where $\|c\|_1 = \sum_{j=1}^{M} |c_j|$. Then, if $\|c\|_1 \leq 1$, the average projection of this density matrix $\rho = V\rho_0 V^\dagger$ on the subspace of interest is at least as large as $f(H)\rho_0 f(H)^\dagger$. That is, if Π is the projector on to this subspace,

$$\mathbb{E}[\mathrm{Tr}(\Pi\rho)] = \mathrm{Tr}[\Pi\bar{\rho}] \geq \mathrm{Tr}[\Pi f(H)\rho_0 f(H)^\dagger].$$

We call this technique *Ancilla-free LCU* as it does not require any ancilla qubits, by avoiding the need to prepare $f(H)\rho_0 f(H)^\dagger$, by *Standard LCU*. Formally, we prove the following theorem:

Theorem 2 (Randomized unitary sampling)**.** *Let $H \in \mathbb{C}^{N\times N}$ is a Hermitian matrix. Also let $\varepsilon \in (0,1)$ and suppose $f : [-1,1] \mapsto \mathbb{R}$ be some function such that*

$$\left\| f(H) - \sum_{j=1}^{M} c_j U_j \right\| \leq \frac{\varepsilon}{3\|f(H)\|},$$

for some unitaries U_j and $c_j \in \mathbb{R}\backslash\{0\}$ such that $\|c\|_1 = \sum_j |c_j| \leq 1$. Suppose V is a unitary sampled from the ensemble $\{c_j/\|c\|_1, U_j\}$, and applied to some initial state ρ_0. Then, the average density matrix, defined as

$$\bar{\rho} = \mathbb{E}\left[V\rho_0 V^\dagger\right] = \frac{1}{\|c\|_1}\sum_{j=1}^{M} c_j U_j \rho_0 U_j^\dagger,$$

satisfies,

$$\mathrm{Tr}\left[\Pi\bar{\rho}\right] \geq \mathrm{Tr}[\Pi f(H)\rho_0 f(H)^\dagger] - \varepsilon,$$

for any projector Π acting on the state space of $\bar{\rho}$.

We prove Theorem 2 in Sec. 3.4. The result of Theorem 2 can be interpreted as follows: Suppose we are interested in the projection of $f(H)\,|\psi_0\rangle$ in some subspace of interest, and in the average projection of the resulting state in this subspace. Then, instead of implementing the standard LCU procedure, we can simply sample a V as described above and apply it to the initial state $|\psi_0\rangle$. On average, this leads to the density matrix $\bar{\rho}$. Theorem 2 states that the projection of $\bar{\rho}$ in this subspace is guaranteed to be is at least as large. In other words, the average success probability of the projective measurement on $\rho = V\rho_0 V^\dagger$ would be at least as high even when *Standard LCU* is replaced with just importance sampling. This also extends to the continuous-time setting.

Thus while *Standard LCU* requires $\lceil \log M \rceil$ ancilla qubits, our method requires none. However, it is important to note that Theorem 2 does not always guarantee a high success probability: it only does so, on average. However, this is useful, for instance, if for the underlying problem, when we care about the average success probability. This makes *Ancilla-free LCU* well suited to tackle the spatial search problem, defined as follows for random walks: Consider any reversible Markov chain P such that a certain subset of its nodes (say M out of N) are marked. Then, the expected number steps required by a classical random walk to reach the marked nodes is known as the hitting time (HT). Analogously, quantum spatial search involves determining the expected number of steps after which the projection of the quantum walk on to the marked subspace is high. Indeed, demonstrating that quantum walks require $O(\sqrt{HT})$ steps on average, for any P and any M, had been a long standing open problem. Following a series of works [39, 57, 40, 41] that only partially resolved this problem (such as for specific graphs or a single marked element), a generic quadratic speedup (up to a log factor) was finally proven in Ref. [32]. This algorithm relies on *Standard LCU*, requiring $O(\log(HT))$ ancilla qubits, in addition to the walk space (edges of P). Here we show that the same generic quadratic speedup can be obtained without using any of the $O(\log(HT))$ ancilla qubits by using *Ancilla-free LCU*.

Applications: We discuss two separate quantum algorithms by discrete-time quantum walks that solve the spatial search problem quadratically faster than classical random walks, while requiring fewer ancilla qubits than the prior art. The first one relies on fast-forwarding discrete-time random walks and formalizes an observation in [42]. The second quantum algorithm relies on the fast-forwarding of continuous-time random walks, allowing us to relate both frameworks of classical random walks with discrete-time quantum walks. The running time of both these algorithms scales as the square root of the hitting time of classical random walks (up to log factors) even in the presence of multiple marked elements, which is optimal. Overall, we demonstrate that the *Ancilla-free LCU* provides a unified framework for optimal quantum spatial algorithms. It also allows us to connect discrete and continuous-time random walks with discrete and continuous-time quantum walks. We briefly describe the results we obtain:

(a) **Quantum spatial search by discrete-time quantum walks:** We use the fact that if any Hamiltonian H is encoded in the top-left block of a unitary U_H, we can obtain a block encoding [43, 27] of H^t or $e^{-t(I-H)}$ by implementing a linear combination of Chebyshev polynomials of H [26, 32]. Both these procedures require roughly $O(\sqrt{t})$ cost to be implemented. When $H = D$ (the discriminant matrix of P), D^t results in a fast-forwarding of discrete-time random walks [31], which is the key subroutine of the optimal spatial search algorithm of [32]. One the other hand,

when $H = e^{-t(I-D)}$, a fast-forwarding of continuous-time random walks can be achieved. However, quantum fast forwarding makes use of *Standard LCU*, requiring $O(\log t)$ ancilla qubits. However, for quantum spatial search, we can invoke Theorem 2 instead, and make use of *Ancilla-free LCU*. This is possible as for this problem we are interested in proving that quantum walks require $\widetilde{O}(\sqrt{HT})$ steps, on average. Thus, our methods do not require any ancilla qubits (other than the walk space).

More precisely, using the framework of interpolated Markov chains (see Sec. 7.1 for details of these terms), and for a specific initial state $|\sqrt{\pi_U}\rangle$ (related to the stationary distribution of the interpolated random walk), we can show for the first spatial search algorithm, the *Ancilla-free LCU* procedure, on average, prepares a mixed state $\bar{\rho}$ such that

$$\mathrm{Tr}[(I \otimes \Pi_M)\bar{\rho}] \geq \left\| \Pi_M D(s)^T \, |\sqrt{\pi_U}\rangle \right\|^2 - \varepsilon,$$

where Π_M is the projector on to the marked subspace. In Ref. [32], the authors proved that, on average, the RHS of the aforementioned inequality is $\tilde{\Omega}(1)$, for $T = \widetilde{O}\left(\sqrt{HT}\right)$, for some randomly chosen value of $s \in [0, 1)$ (See Algorithm 4). Thus our algorithm achieves the same quadratic speedup as [32], while requiring $O(\log HT)$ fewer ancilla qubits. This also formalizes the observation of Ref. [42].

For our second spatial search algorithm by discrete-time quantum walk (See Algorithm 6), Theorem 2, prepares, on average, the mixed state $\bar{\rho}$ such that

$$\mathrm{Tr}[(I \otimes \Pi_M)\bar{\rho}] \geq \left\| \Pi_M e^{T(D(s)-I)} \, |\sqrt{\pi_U}\rangle \right\|^2 - \varepsilon,$$

where again, we prove that the expected value of the RHS is in $\widetilde{\Omega}(1)$, for $T = \widetilde{O}\left(\sqrt{HT}\right)$. For this we prove that the probability of finding a marked vertex is at least as large as the probability of a certain event occurring in the continuous-time interpolated Markov chain $P(s)$, which in turn can be lower bounded by the probability of the same event occurring for the corresponding discrete-time Markov chain. These reductions allow us to leverage the results of [32]. Both these algorithms require no ancilla qubits (other than the walk space), while solving this problem via quantum fast-forwarding by *Standard LCU* would require $O(\log HT)$ ancilla qubits.

It is worth mentioning that optimal spatial search algorithm by continuous-time quantum walk [33] is yet another demonstration of (a continuous-time variant of) *Ancilla-free LCU*, where a quadratic speedup for this problem could be obtained by the fast forwarding of a continuous-time random walk. This algorithm involved using randomized time-evolution through which the full LCU procedure could be bypassed. Additionally, these optimal quantum spatial search algorithms allow us to relate discrete-time random walks and continuous-time random walks with their quantum counterparts through quantum spatial search. In the Appendix (Sec. E), we also establish a relationship between discrete and continuous-time quantum walks. In a seminal work, Childs showed that any dynamics generated by a continuous-time quantum walk can be simulated by a discrete-time quantum walk [58]. We show that given access to a discrete-time quantum walk, one can also generate a continuous-time quantum walk (on the edges of the underlying Markov chain). In the other direction, whether discrete-time quantum walks can be obtained from the continuous-time quantum walk evolution operator, has been unknown. In fact, using the frameworks of block encoding and quantum singular value transformation, we show that given access to a quantum walk time evolution operator, one can obtain a

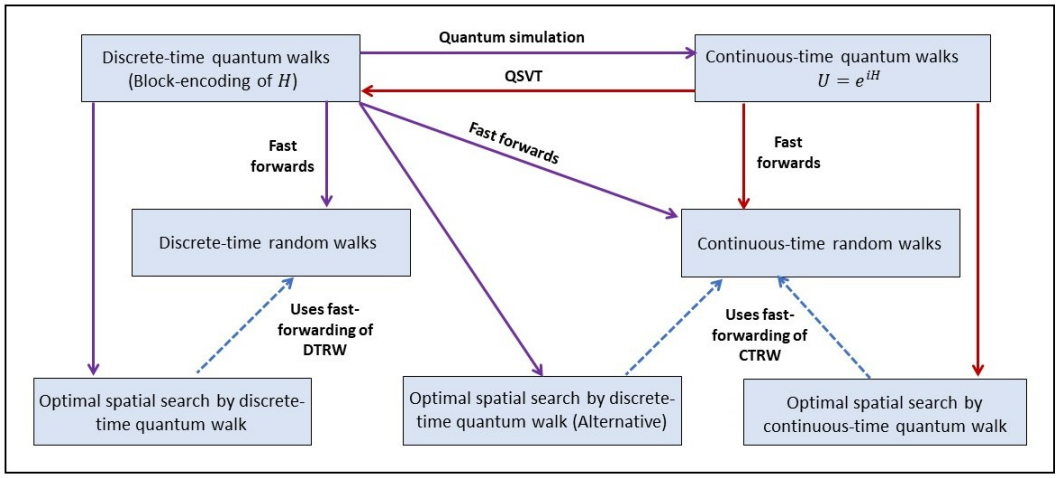

Figure 3: *Relationship between discrete and continuous-time quantum walks, and their classical counterparts:* Given the block encoding of the discriminant matrix D of any reversible Markov chain P, we can generate a continuous-time quantum walk on both the vertices and edges of P, following Ref. [58]. Conversely, given access to the continuous-time quantum time-evolution operator $U = e^{-iH/2}$, we can implement a discrete-time quantum walk on the edges of H. For this, we make use of the framework of Quantum Singular Value Transformation (QSVT). Apers and Sarlette demonstrated that discrete-time random walks can be fast forwarded by discrete-time quantum walks [31] which is the cornerstone of the recently developed optimal quantum spatial search algorithm [32]. In this work, we show that for quantum spatial search, the fast-forwarding can be done without needing any ancilla qubits (other than the quantum walk space) through *Ancilla-free LCU*. Additionally, we also demonstrate that discrete-time quantum walks can fast forward continuous-time random walks. This fact, can also be leveraged to develop new optimal quantum spatial search algorithms, which do not require any ancilla qubits (other than the quantum walk space). Finally continuous-time random walks can also be fast-forwarded by continuous-time quantum walks which is central to the optimal quantum spatial search algorithm of [33]. Thus overall, through this work we connect all different frameworks random and quantum walks.

discrete-time quantum walk. We also discuss the subtleties of this approach with regards to quantum fast forwarding and also suggest possible improvements. Overall, this helps us complete the picture (See Fig. 3) by relating both frameworks of quantum and random walks.

1.2 Prior work

In this section, we briefly sketch relevant prior work and relate them to the results we obtain. The linear combination of unitaries technique was first introduced by Childs and Wiebe [22] to develop quantum Hamiltonian simulation algorithms based on multi-product formulas. Since then, LCU has been extensively used to develop improved quantum algorithms for Hamiltonian simulation [23, 25, 24]. Subsequently, it has been used to develop a wide variety of quantum algorithms for linear algebra, such as for solving quantum linear systems, and linear regression [26, 27], preparing ground states of Hamiltonians [34] and solving optimization problems [37, 38]. Many of these quantum algorithms have been unified by the more recent framework of quantum singular value transformation (QSVT) [45], which implements polynomial transformations to the singular values of a matrix. Given access to a unitary that block encodes [27] the matrix, QSVT provides near-optimal query complexities for these problems (in terms of the number of queries made to the block encoding) and requires fewer ancilla qubits than LCU. However, constructing the block

encoding itself can be resource consuming and may require a large number of ancilla qubits and multi-qubit controlled gates which increases the overall circuit depth. Thus, it is not likely to be implemented on early fault-tolerant quantum computers.

The main contribution of this article is to demonstrate that the framework of implementing LCU can be modified so that this framework, which is also applicable to a wide variety of problems, is implementable on intermediate-term quantum computers. As discussed in the previous sections, we introduce three main variants of LCU.

The first technique, *Single-Ancilla LCU*, implements any LCU using only a single ancilla qubit and no multi-qubit controlled operations. Despite this, the cost per coherent run is lower than the generic LCU procedure. Consequently, it is useful for early fault-tolerant quantum computers. A few quantum algorithms, tailored to early fault-tolerant quantum computers have recently been developed. However, these algorithms tackle a single problem, namely, the estimation of the ground state energy of a given Hamiltonian [59, 12, 11, 13, 10]. Our method makes use of the quantum circuit of Faerhmann et al. [9], wherein it was used for multi-product Hamiltonian simulation. In a way, it is a non-trivial generalization of their technique to implement any LCU. This opens up the possibility of developing several novel quantum algorithms using early fault-tolerant quantum devices. In this work, we apply our method for Hamiltonian simulation, ground state property estimation, and quantum linear systems.

For each of these problems, several quantum algorithms have been developed over the years. Hamiltonian simulation has been widely considered as one of the potential applications of the first useful quantum computer. Algorithms more suited to near-term applicability include Trotter-based approaches [48] and their randomized variants such as qDRIFT [60, 61] and others [49, 50]. The standard LCU procedure has also been quite useful for developing near-optimal quantum simulation algorithms [24, 25, 28]. For our Hamiltonian simulation algorithm, we make use of the LCU decomposition of the time evolution operator [13], along with the Truncated Taylor series method of Berry et al. [25]. State-of-the-art quantum simulation algorithms use the framework of quantum signal processing [62], and QSVT [45], but require access to a block encoding of the Hamiltonian, which, as discussed previously, is resource consuming. We extensively compare our Hamiltonian simulation procedure with other algorithms (Sec. 4).

The first quantum algorithm for ground state preparation involved using Hamiltonian simulation along with quantum phase estimation [63]. Subsequently, Refs. [34, 35] took advantage of the fact that functions of Hamiltonians can be expressed as a linear combination of unitaries to develop fast quantum algorithms for ground state preparation and estimation. A QSVT-based quantum algorithm has also been developed recently [64], which requires an optimal number of queries to the block encoding of the Hamiltonian. The problem of ground state preparation/property estimation is also considered to be one of the first problems to be solved using near/intermediate-term quantum computers. In Sec. 5.2 and Sec. C.2, we substantively compare our procedure with other early fault-tolerant quantum algorithms for ground energy estimation, as well as with state-of-the-art quantum algorithms, suitable for fully fault-tolerant quantum computers.

Ever since the seminal algorithm by Harrow, Hassidim and Lloyd [65], quantum linear systems has been analyzed extensively. In particular, the LCU-based approach of Childs, Kothari, and Somma [26] provided a linear dependence on the condition number of the underlying sparse matrix and an exponentially improved dependence on the error. This algorithm was improved to also work for non-sparse matrices [66] and in the more general block encoding framework [67]. Recently, QSVT-based approaches for this problem have also been developed [45, 46, 67]. Another direction of research has been to develop

quantum algorithms for this problem in the adiabatic quantum computing framework [68, 69, 51]. This approach provides an optimal dependence on all parameters but requires access to a block encoding of the matrix to be inverted. The possibility of applying quantum linear systems algorithm on the near-term quantum devices has been explored in [70]. However, the techniques discussed are either variational and hence heuristic, or requires substantially higher resources as compared to our method. We rigorously compare our approach with these techniques in Sec. 6.2 and the Appendix (Sec. C.2).

Our second technique is a continuous-time variant of LCU, which is more physical. The key idea is to couple discrete systems with continuous-variable systems. Such interactions have been explored in the context of quantum phase estimation, where the system Hamiltonian is coupled with a one-dimensional free particle, acting as the pointer variable - which is the so-called von Neumann measurement model [71, 72]. In Ref. [33], the continuous-time quantum walk Hamiltonian H was coupled to a one-dimensional quantum Harmonic oscillator to implement e^{-tH^2}. This was a key ingredient of their spatial search algorithm by continuous-time quantum walk. In this work, we formalize this technique and show that it is more widely applicable and in fact, can serve as a continuous-time variant to any LCU-based quantum algorithm. We develop an analog variant of the quantum linear systems algorithm of Childs et al. [26] and a new quantum algorithm for this problem (using only Gaussian states) that is more suited for intermediate-term implementation.

Finally, we apply the *Ancilla-free LCU* technique to develop optimal quantum spatial search algorithms. LCU has been central to the development of several quantum walk algorithms. The quantum fast-forwarding scheme by Apers and Sarlette [31] quadratically fast-forwards the dynamics of a discrete-time random walk by implementing a linear combination of discrete-time quantum walk steps. Recently, Ambainis et al. [32] proved that for the spatial search problem, fast-forwarding T random walk steps on an interpolated Markov chain, prepares a quantum state that has, on average, a $\tilde{\Omega}(1)$ overlap with the marked space for $T = \widetilde{O}(\sqrt{HT})$, where HT is the classical hitting time of the random walk. Overall, this algorithm requires $O(\log n + \log HT)$ ancilla qubits, for any reversible Markov chain of n nodes. Their LCU-based quantum spatial search algorithm for discrete-time quantum walks completely solves the spatial search problem quadratically faster than classical random walks, for any number of marked nodes. This closed a long line of work which made partial progress towards solving this problem (such as for only particular reversible Markov chains or when only a single node was marked) [39, 40, 41]. Subsequently, Apers et al. [42] provided a unified framework that connected the different variants of discrete-time quantum walk search. Therein, the authors observed that the LCU procedure of [32] could be replaced by randomly sampling quantum walk steps. In the continuous-time quantum walk framework, whether the spatial search problem offered a generic quadratic speedup was also open for a long time, with a series of works that partially resolved this problem [57, 73, 74, 75]. It was only recently solved in [33]. Their analog quantum algorithm indeed managed to bypass the LCU procedure by evolving the system under the quantum walk Hamiltonian for a random time. In this work, we demonstrate that by using *Ancilla-free LCU* instead of using the full LCU procedure (thereby requiring fewer ancilla qubits), one can develop two spatial search algorithms by discrete-time quantum walks: one that relies on fast-forwarding discrete-time random walks and the other, relying on fast-forwarding continuous-time random walks. Similar to [32], these quantum algorithms provide a generic quadratic speedup over their classical counterparts, while requiring $O(\log HT)$ fewer ancilla qubits.

In the following section, we briefly define some of the key preliminary concepts that we will borrow to develop our results.

2 Preliminaries

In this section, we introduce some of the techniques that we use in this article as well as discuss the key algorithmic primitives required to develop our results. We begin by stating the complexity theoretic notations we shall be using throughout the article.

2.1 Notation

Complexity theoretic notations: Throughout the article, we shall be using the standard complexity theoretic notations. The *Big-O* notation, $g(n) = O(f(n))$ or $g(n) \in O(f(n))$, implies that g is upper bounded by f. That is, there exists a a constant $k > 0$ such that $g(n) \leq k \cdot f(n)$. The *Big-Omega* notation, $g(n) = \Omega(f(n))$, implies $g(n) \geq kf(n)$ (g is lower bounded by f). The *Theta* notation is used when g is both bounded from above and below by f, i.e. $g(n) = \Theta(f(n))$, if there exists non-negative constants k_1, and k_2 such that $k_1 f(n) \leq g(n) \leq k_2 f(n)$. The *Little-o* notation, $g(n) = o(f(n))$, when g is dominated by f asymptotically, i.e. $\lim_{n\to\infty} g(n)/f(n) = 0$.

For each of these notations, it is standard to use *tilde* ($\sim$) to hide polylogarithmic factors. For instance, $\widetilde{O}(f(n)) = O(f(n)\text{polylog}(f(n)))$. This applies to the other notations as well.

Trace, Expectation and Probability: The trace of an operator A will be denoted by $\text{Tr}[A]$, while the expectation value of the operator will be denote by $\mathbb{E}[A]$. The probability of an event X occurring will be denoted by $\Pr[X]$.

Schatten norms: The Schatten p-norm of the operator X is defined as

$$\|X\|_p = \left(\sum_j \sigma_j^p(X)\right)^{1/p},$$

where $\sigma_j(X)$ is the j^{th} singular value of X. So if $\sigma_{\max}(X)$ denotes the maximum singular value of X, we have

$$\lim_{p\to\infty} \|X\|_p = \sigma_{\max} \cdot \lim_{p\to\infty} \left(\sum_j \frac{\sigma_j^p(X)}{\sigma_{\max}^p(X)}\right)^{1/p} = \sigma_{\max},$$

which is the spectral norm of the operator X, which we denote as simply $\|X\|$.

LCU coefficients: For LCU, we implement some operator $V = \sum_j c_j U_j$, where each U_j is a unitary. Note that c_j can, in general, be any non-zero real or complex number (positive or negative). When c_j is complex or negative, we can absorb the sign as well as the imaginary phase into the description of the unitary itself. Consequently, without loss of generality, it suffices to consider that $c_j \in \mathbb{R}^+ \backslash \{0\}$. This is what we shall be considering throughout the article. In the next section, we describe the formalism of LCU.

2.2 Linear Combination of Unitaries

We will begin by stating the general framework of Linear Combination of Unitaries (LCU). Throughout the article, we shall refer to this as the *Standard LCU* procedure. Suppose for $H \in \mathbb{C}^{N\times N}$, we wish to apply some function of the Hamiltonian to an initial state $|\psi_0\rangle$. More precisely, if H has spectral decomposition $H = \sum_{j=1}^N \lambda_j |v_j\rangle \langle v_j|$, define $f(H) =$

$\sum_{j=1}^{N} f(\lambda_j) |v_j\rangle \langle v_j|$. Now suppose $f(H)$ can be well approximated by linear combinations of unitaries. That is, for some $\gamma \in (0, 1)$ suppose,

$$\left\| f(H) - \sum_{j=1}^{M} c_j U_j \right\| \leq \gamma,$$

where each U_j is a unitary matrix that we have access to, i.e $f(H)$. Without loss of generality let us define the parameters $c_j \in \mathbb{R}^+/\{0\}$. Even though $f(H)$ is not necessarily unitary, the LCU technique allows us to implement $f(H) |\psi_0\rangle$.

For this, the procedure requires $m = \lceil \log_2 M \rceil$ ancilla qubits. First, a *prepare* unitary R is applied to this ancilla register such that

$$R |\bar{0}\rangle = \sum_{j=1}^{M} \sqrt{\frac{c_j}{\|c\|_1}} |j\rangle ,$$

where $c = (c_1, \ldots, c_M)^T$. Suppose the cost of implementing this unitary is τ_R.

Furthermore, a *select* unitary Q, defined in the following way,

$$Q = \sum_j |j\rangle \langle j| \otimes U_j,$$

is also used. Note that Q is a sophisticated operation, controlled on each of the m ancilla qubits. Suppose the cost of implementing Q is τ_Q.

Then,

$$|\psi_t\rangle = (R^\dagger \otimes I) Q (R \otimes I) |\bar{0}\rangle |\psi_0\rangle = \frac{1}{\|c\|_1} |\bar{0}\rangle \sum_{j=1}^{M} c_j U_j |\psi_0\rangle + |\Phi\rangle^\perp \tag{3}$$

where $(|\bar{0}\rangle \langle \bar{0}| \otimes I) |\Phi\rangle^\perp = 0$. Note that, controlled on $|\bar{0}\rangle$, the state in the second register is $\gamma/\|c\|_1$-close to

$$|\psi\rangle = \frac{f(H) |\psi_0\rangle}{\| f(H) |\psi_0\rangle \|},$$

with probability $p = \| f(H) |\psi_0\rangle \|^2 / \|c\|_1^2$. The total cost of using the LCU procedure is then

$$\Gamma_{\max} = 2\tau_R + \tau_Q + \tau_{\psi_0}. \tag{4}$$

Let us now explore the applicability of this procedure. For instance, $f(H)$ could be well approximated by a Fourier series, in which case, $U_j = e^{-ijH}$. Since, for several applications, it indeed boils down to applying some such $f(x)$ to an initial state, LCU provides a versatile framework that has wide applicability. Consequently, several near-optimal quantum algorithms have been designed in this framework ranging from quantum algorithms for linear systems [26], ground state preparation [34, 35], sampling from thermal states [37, 76] to Hamiltonian simulation [22, 23, 24, 25] and many others.

However, the twin requirements of several ancilla qubits, as well as sophisticated multi-qubit controlled operations, imply that this standard technique to implement any LCU is useful for only full-scale fault-tolerant quantum computers. Additionally, in most applications, simply preparing the quantum state $|\psi\rangle$ is not useful, as access to the entries of the underlying quantum state are required. However, even state-of-the-art techniques in quantum state tomography result in an exponential overhead. Thus, in most cases, we are interested in predicting certain properties of the state $|\psi\rangle$, such as estimating the expectation values of observables of interest, i.e. $\langle \psi | O | \psi \rangle$.

2.3 Block encoding and Quantum Singular Value Transformation

For some of the algorithms in this article, we will consider the framework of *block encoding*, wherein it is assumed that the input matrix H (up to some sub-normalization) is stored in the left block of some unitary. The advantage of the block encoding framework, which was introduced in a series of works [43, 27, 44], is that it can be applied to a wide variety of input models.

Definition 3 (Block Encoding [27])**.** *Suppose that H is an s-qubit operator, $\alpha, \varepsilon \in \mathbb{R}^+$ and $a \in \mathbb{N}$, then we say that the $(s+a)$-qubit unitary U_H is an (α, a, ε)-block encoding of H, if*

$$\left\| H - \alpha(\langle 0|^{\otimes a} \otimes I) U_H (|0\rangle^{\otimes a} \otimes I) \right\| \leq \varepsilon. \tag{5}$$

Let $|\psi\rangle$ be an s-qubit quantum state. Then applying U_H to $|\psi\rangle |0\rangle^{\otimes a}$ outputs a quantum state that is $\frac{\varepsilon}{\alpha}$-close to

$$|0\rangle^{\otimes a} \frac{H}{\alpha} |\psi\rangle + |\Phi^{\perp}\rangle ,$$

where $\left(|0\rangle^{\otimes a} \langle 0|^{\otimes a} \otimes I_s\right) |\Phi^{\perp}\rangle = 0$. Equivalently, suppose $\tilde{H} := \alpha \left(\langle 0|^{\otimes a} \otimes I_s\right) U_H \left(|0\rangle^{\otimes a} \otimes I_s\right)$ denotes the actual matrix that is block-encoded into U_H, then $\left\| H - \tilde{H} \right\| \leq \varepsilon$.

Quantum Singular Value Transformation (QSVT) applies a polynomial transformation to the singular values of a block-encoded matrix [45]. Formally, let $P \in \mathbb{R}[x]$ be a polynomial with real coefficients of degree $n \geq 2$, such that P is either even or odd (has parity-$n \mod 2$), and for all $x \in [-1, 1]$, $|P(x)| \leq 1$. Then, QSVT allows us to implement any polynomial $P(x)$ that satisfies the aforementioned requirements. Next, we formally introduce QSVT formally via the following theorem.

Theorem 4 (Quantum Singular Value Transformation [44])**.** *Suppose $A \in \mathbb{R}^{N \times d}$ is a matrix with singular value decomposition $A = \sum_{j=1}^{d_{\min}} \sigma_j |v_j\rangle \langle w_j|$, where $d_{\min} = \min\{N, d\}$ and $|v_j\rangle$ ($|w_j\rangle$) is the left (right) singular vector with singular value σ_j. Furthermore, let U_A be a unitary such that $A = \widetilde{\Pi} U_A \Pi$, where Π and $\widetilde{\Pi}$ are orthogonal projectors. Then, if any real polynomial $P(x)$ of degree n is even or odd, and satisfies $|P(x)| \leq 1$, for all $x \in [-1, 1]$, there exists a vector $\Phi = (\phi_1, \phi_2, \cdots \phi_n) \in \mathbb{R}^n$ and a unitary*

$$U_\Phi = \begin{cases} e^{i\phi_1(2\widetilde{\Pi} - I)} U_A \left[\prod_{k=1}^{(n-1)/2} e^{i\phi_{2k}(2\widetilde{\Pi} - I)} U_A^\dagger e^{i\phi_{2k+1}(2\widetilde{\Pi} - I)} U_A \right], & n \text{ is odd} \\ \left[\prod_{k=1}^{n/2} e^{i\phi_{2k-1}(2\widetilde{\Pi} - I)} U_A^\dagger e^{i\phi_{2k}(2\widetilde{\Pi} - I)} U_A \right], & n \text{ is even,} \end{cases} \tag{6}$$

such that

$$P^{SV}(A) = \begin{cases} \left(\langle +| \otimes \widetilde{\Pi}\right) \left(|0\rangle \langle 0| \otimes U_\Phi + |1\rangle \langle 1| \otimes U_{-\Phi}\right) \left(|+\rangle \otimes \Pi\right), & n \text{ is odd} \\ \left(\langle +| \otimes \Pi\right) \left(|0\rangle \langle 0| \otimes U_\Phi + |1\rangle \langle 1| \otimes U_{-\Phi}\right) \left(|+\rangle \otimes \Pi\right), & n \text{ is even,} \end{cases} \tag{7}$$

where $P^{SV}(A)$ is the polynomial transformation of the matrix A defined as

$$P^{SV}(A) := \begin{cases} \sum_j P(\sigma_j) |v_j\rangle \langle w_j|, & P \text{ is odd} \\ \sum_j P(\sigma_j) |w_j\rangle \langle w_j|, & P \text{ is even.} \end{cases} \tag{8}$$

Theorem 4 tells us that for any real, bounded n-degree polynomial P with definite parity, we can implement $P^{SV}(A)$ using one ancilla qubit, $\Theta(n)$ applications of U_A, $U_A^\dagger$

and controlled reflections $I-2\Pi$ and $I-2\widetilde{\Pi}$. Furthermore, if in some well-defined interval, some function $f(x)$ is well approximated by an n-degree polynomial $P(x)$, then Theorem 4 also allows us to implement a transformation that approximates $f(A)$, where

$$f(A) := \begin{cases} \sum_j f(\sigma_j) \left|v_j\right\rangle \left\langle w_j\right|, & P \text{ is odd} \\ \sum_j f(\sigma_j) \left|w_j\right\rangle \left\langle w_j\right|, & P \text{ is even.} \end{cases} \tag{9}$$

The following theorem from Ref. [45] deals with the robustness of the QSVT procedure, i.e. how errors propagate in QSVT. In particular, for two matrices A and $\tilde{A}$, it shows how close their polynomial transformations ($P^{SV}(A)$ and $P^{SV}(\tilde{A})$, respectively) are, as a function of the distance between A and $\tilde{A}$.

Lemma 5 (Robustness of Quantum Singular Value Transformation, [45])**.** *Let $P \in \mathbb{R}[x]$ be a n-degree real polynomial that satisfies the requirements of QSVT. Let $A, \tilde{A} \in \mathbb{C}^{N\times M}$ be matrices of spectral norm at most 1. Then,*

$$\left\| P^{SV}(A) - P^{SV}(\tilde{A}) \right\| \leq 4n\sqrt{\left\| A - \tilde{A} \right\|}.$$

Having discussed the preliminary concepts, in the next section, we explain the three variants of LCU we consider in this article.

3 Three different approaches for implementing LCU on intermediate-term quantum computers

In this section, we present the three different LCU techniques. Before stating these techniques, we prove a general lemma that will be invoked while proving results related to our approaches.

3.1 Robustness of expectation values of observables

In this section, we develop general results on the robustness of expectation values of observables which we shall use for both the *Single-Ancilla LCU* (Sec. 3.2) and the *Ancilla-free LCU* (Sec. 3.4) approaches.

Consider that there exist two operators P and Q such that $\|P-Q\| \leq \gamma$. In this section, we demonstrate that the expectation value of O with respect to $P\rho P^\dagger$ is not far off from the expectation value of O with respect to $Q\rho Q^\dagger$, for any density matrix ρ. More precisely, we prove

$$\left|\mathrm{Tr}[O\ P\rho P^\dagger] - \mathrm{Tr}[O\ Q\rho Q^\dagger]\right| \leq 3\|P\|\|O\|\,\gamma.$$

In order to prove this result, we need to use the tracial version Hölder's inequality which is stated below for completeness:

Lemma 6 (Tracial version of Hölder's inequality [77])**.** *Define two operators A and B and parameters $p, q \in [1,\infty]$ such that $1/p + 1/q = 1$. Then the following holds:*

$$\mathrm{Tr}[A^\dagger B] \leq \|A\|_p \|B\|_q .$$

Here $\|X\|_p$ corresponds to the Schatten p-norm of the operator X. For the special case of $p = \infty$ and $q = 1$, the statement of Lemma 6 can be rewritten as

$$\mathrm{Tr}[A^\dagger B] \leq \|A\|_\infty \|B\|_1 = \|A\| \|B\|_1 . \tag{10}$$

Now we are in a position to formally state the main result.

Theorem 7. *Suppose P and Q are operators such that $\|P - Q\| \leq \gamma$ for some $\gamma \in [0, 1]$. Furthermore, let ρ be any density matrix and O be some Hermitian operator with spectral norm $\|O\|$. Then, if $\|P\| \geq 1$, the following holds:*

$$\left|\mathrm{Tr}[OP\rho P^\dagger] - \mathrm{Tr}[OQ\rho Q^\dagger]\right| \leq 3\|O\|\|P\|\gamma.$$

Proof. For the operators P and Q, we have

$$(Q - P)\rho(P^\dagger - Q^\dagger) = Q\rho(P^\dagger - Q^\dagger) - P\rho(P^\dagger - Q^\dagger) \tag{11}$$

Now adding and subtracting $P\rho P^\dagger$ in the RHS we obtain

$$\begin{aligned}(Q - P)\rho(P^\dagger - Q^\dagger) + P\rho(P^\dagger - Q^\dagger) &= Q\rho(P^\dagger - Q^\dagger) + P\rho P^\dagger - P\rho P^\dagger && (12)\\ &= P\rho P^\dagger - Q\rho Q^\dagger - (P - Q)\rho P^\dagger && (13)\end{aligned}$$

This gives us

$$P\rho P^\dagger - Q\rho Q^\dagger = (Q - P)\rho(P^\dagger - Q^\dagger) + P\rho(P^\dagger - Q^\dagger) + (P - Q)\rho P^\dagger. \tag{14}$$

Now, using the tracial version of Holder's inequality with $p = \infty$ and $q = 1$, we have

$$\left|\mathrm{Tr}[O\ P\rho P^\dagger] - \mathrm{Tr}[O\ Q\rho Q^\dagger]\right| \leq \|O\| \cdot \left\|P\rho P^\dagger - Q\rho Q^\dagger\right\|_1 . \tag{15}$$

For the second term in the RHS of the above equation, we can use the expression in Eq. (14). That is,

$$\left\|P\rho P^\dagger - Q\rho Q^\dagger\right\|_1 = \left\|(Q - P)\rho(P^\dagger - Q^\dagger) + P\rho(P^\dagger - Q^\dagger) + (P - Q)\rho P^\dagger\right\|_1 . \tag{16}$$

At this stage we can successively apply the tracial version of Holder's inequality (Lemma 6 with $p = \infty$ and $q = 1$) to the RHS of the expression above to obtain

$$\begin{aligned}\left\|P\rho P^\dagger - Q\rho Q^\dagger\right\|_1 &\leq \|Q - P\|\left\|\rho(P^\dagger - Q^\dagger)\right\|_1 + \|P\|\left\|\rho(P^\dagger - Q^\dagger)\right\|_1 + \|P - Q\|\left\|\rho P^\dagger\right\|_1 && (17)\\ &\leq \|P - Q\|^2\|\rho\|_1 + \|P\|\|P - Q\|\|\rho\|_1 + \|P - Q\|\|P\|\|\rho\|_1 && (18)\\ &\leq \|P - Q\|^2 + 2\|P - Q\|\|P\| \qquad [\text{ Using } \|\rho\|_1 = 1\ ] && (19)\end{aligned}$$

Now, substituting this back in the RHS of Eq. (15), we obtain

$$\begin{aligned}\left|\mathrm{Tr}[OP\rho P^\dagger] - \mathrm{Tr}[OQ\rho Q^\dagger]\right| &\leq \|O\|\|P - Q\|^2 + 2\|O\|\|P\|\|P - Q\| && (20)\\ &\leq \gamma^2\|O\| + 2\|O\|\|P\|\gamma \qquad [\text{ As } \|P - Q\| \leq \gamma] && (21)\\ &\leq 3\gamma\|O\|\|P\| \qquad [\text{ As } \|P\| \geq 1] && (22)\end{aligned}$$

□

It is easy to see why Theorem 7 is useful to develop robust versions of *Single Ancilla LCU* and *Ancilla-free LCU*. Typically, $f(H)$ is often not exactly equal to a linear combination of unitaries but is γ-close to it. Formally, $\|f(H) - g(H)\| \leq \gamma$, where $g(H)$ can be exactly expressed as an LCU. Consequently, for the variants of LCU that we develop in the subsequent sections, we will be estimating $\mathrm{Tr}[O\ g(H)\rho g(H)^\dagger]$. But, by Theorem 7, we can always bound

$$\left|\mathrm{Tr}[O\ f(H)\rho f(H)^\dagger] - \mathrm{Tr}[O\ g(H)\rho g(H)^\dagger]\right| \leq 3\|O\|\|f(H)\|\,\gamma.$$

We shall be using this result in the subsequent sections.

3.2 Single-Ancilla LCU: Estimating expectation values of observables

In this section, we describe the *Single-Ancilla LCU* technique, which allows us to sample from quantum states obtained by applying LCU. Suppose we are given a Hermitian matrix $H \in \mathbb{C}^{N\times N}$ and we wish to implement $f(H)$, which can be approximated by a linear combination of unitaries. That is, for some $\gamma \in [0, 1)$,

$$\left\| f(H) - \sum_{j=1}^{M} c_j U_j \right\| \leq \gamma,$$

for unitaries U_j and $c_j \in \mathbb{R}^+\backslash\{0\}$. Let us define the ℓ_1-norm of the LCU coefficients as $\|c\|_1 = \sum_{j=1}^{M} c_j$. Furthermore, suppose we have access to the quantum circuit for U_j. Then given any initial state ρ_0, we can estimate the expectation value

$$\mathrm{Tr}[O\rho] = \frac{\mathrm{Tr}[O\ f(H)\rho_0 f(H)^\dagger]}{\mathrm{Tr}[f(H)\rho_0 f(H)^\dagger]},$$

to arbitrary accuracy, for any observable O, using the quantum circuit in Fig. 2. We estimate both the expectation value (numerator) and the norm (denominator), by making use of Algorithm 1,whose steps are stated thusly: the ancilla qubit is prepared in the state $|+\rangle$ so that the overall initial state is $\rho_1 = |+\rangle\langle+| \otimes \rho_0$. We sample two unitaries V_1 and V_2, independently from the ensemble $\mathcal{D} = \{U_j, c_j/\|c\|_1\}$ and then implement controlled and anti-controlled versions of V_1 and V_2, respectively. Finally, we make a measurement of the observable $X \otimes O$ and store the outcome. We then sample from this quantum circuit T times and calculate the mean of all the outcomes. This is the final output of Algorithm 1.

If τ_j is the cost of implementing the unitary U_j, then the cost of implementing V_1 (or V_2) depends on this quantity. Indeed, the average cost of implementing these unitaries (in each coherent run of Algorithm 1) is given by $2\langle\tau\rangle$, where

$$\langle\tau\rangle = \frac{1}{\|c\|_1}\sum_{j=1}^{M} c_j\tau_j. \tag{23}$$

Additionally, if τ_{ρ_0} is the cost of preparing the initial state ρ_0, we can we can upper bound the average cost of each coherent run of Algorithm 1 is $2\langle\tau\rangle + \tau_{\rho_0}$. In the worst case, if $\tau_{\max} = \max_j \tau_j$, we have $\langle\tau\rangle \leq \tau_{\max}$. Thus, a pessimistic upper bound on the cost per coherent run would be $2\tau_{\max} + \tau_{\rho_0}$. Next, we calculate the number of classical iterations T required to estimate the desired quantity.

In order to estimate $\mathrm{Tr}[O\rho]$, we need to run this Algorithm twice: first to estimate the expectation value, and then to estimate the norm. First, we formally show via Theorem

Algorithm 1: `Expectation-observable` $(O, \{c_j, U_j\}, \rho_0, T)$

1. Prepare the state $\rho_1 = |+\rangle\langle+| \otimes \rho_0$.
2. Obtain i.i.d. samples V_1, V_2 from the distribution $\left\{U_j, \frac{c_j}{\|c\|_1}\right\}$.
3. For $\tilde{V}_1 = |0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes V_1$ and $\tilde{V}_2 = |0\rangle\langle 0| \otimes V_2 + |1\rangle\langle 1| \otimes I$, measure $(X \otimes O)$ on the state
$$\rho' = \tilde{V}_2\tilde{V}_1\rho_1\tilde{V}_1^\dagger\tilde{V}_2^\dagger.$$
4. For the j^{th} iteration, store into μ_j, the outcome of the measurement in Step 4.
5. Repeat Steps 1 to 4, a total of T times.
6. Output
$$\mu = \frac{\|c\|_1^2}{T}\sum_{j=1}^{T}\mu_j.$$

8 that if the number of samples obtained T is large enough, the sample mean of the outcomes converges to the expectation value $\mathrm{Tr}[O\ f(H)\rho_0 f(H)^\dagger]$.

Theorem 8 (Estimating expectation values of observables)**.** *Let $\varepsilon, \delta, \gamma \in (0,1)$ be some parameters. Let O be some observable and ρ_0 be some initial state. Suppose there is a Hermitian matrix $H \in \mathbb{C}^{N\times N}$, such that $\left\|f(H) - \sum_j c_j U_j\right\| \leq \gamma$, where U_j is some unitary such that*
$$\gamma \leq \frac{\varepsilon}{6\|O\|\|f(H)\|}.$$
Furthermore, let
$$T \geq \frac{8\|O\|^2\ln(2/\delta)\|c\|_1^4}{\varepsilon^2}.$$
Then, Algorithm 1 estimates μ such that
$$\left|\mu - \mathrm{Tr}[O\ f(H)\rho_0 f(H)^\dagger]\right| \leq \varepsilon,$$
with probability at least $1-\delta$, using of one ancilla qubit and T repetitions of the quantum circuit in Fig. 2.

Proof. Let $g(H) = \sum_j c_j U_j$. First observe from Algorithm 1 that the initial state $\rho_1 = |+\rangle\langle+| \otimes \rho_0$ transforms to
$$\rho' = \tilde{V}_2\tilde{V}_1\rho_1\tilde{V}_1^\dagger\tilde{V}_2^\dagger \tag{24}$$
$$= \frac{1}{2}\left[|0\rangle\langle 0| \otimes V_2\rho_0 V_2^\dagger + |0\rangle\langle 1| \otimes V_2\rho_0 V_1^\dagger + |1\rangle\langle 0| \otimes V_1\rho_0 V_2^\dagger + |1\rangle\langle 1| \otimes V_1\rho_0 V_1^\dagger\right]. \tag{25}$$

So after measuring the observable $X \otimes O$, we have

$$\mathrm{Tr}\left[(X \otimes O)\rho'\right] = \frac{1}{2}\mathrm{Tr}\left[O\left(V_1\rho_0 V_2^\dagger + V_2\rho_0 V_1^\dagger\right)\right].$$

Note that the expected values,

$$\mathbb{E}\left[V_1\right] = \mathbb{E}\left[V_2\right] = \frac{1}{\|c\|_1}\sum_j c_j U_j.$$

So, the expected outcome of the j^{th} iteration is

$$\mathbb{E}\left[\mu_j\right] = \mathbb{E}\left[\mathrm{Tr}[(X \otimes O)\rho']\right] = \frac{1}{\|c\|_1^2}\mathrm{Tr}[O\ g(H)\rho_0 g(H)^\dagger].$$

Next, we need to estimate two things:

(a) How fast does the sample mean $\mu = \sum_j \|c\|_1^2 \mu_j/T$ converge to its expectation value? For this, we use Hoeffding's inequality.

(b) What is the accuracy of the observation with respect to $f(H)$ as a function of the distance between $f(H)$ and $g(H)$? For this, we invoke Theorem 7.

Observe that the POVM measurement yields some outcome of O in the range $[-\|O\|, \|O\|]$. So each random variable lies in the range

$$-\|O\|\|c\|_1^2 \leq \|c\|_1^2\,\mu_j \leq +\|O\|\|c\|_1^2.$$

We evaluate (a), by using Hoeffding's inequality to obtain

$$\Pr\left[\left|\mu - \mathrm{Tr}[O\ g(H)\rho_0\ g(H)^\dagger]\right| \geq \varepsilon/2\right] \leq 2\exp\left[-\frac{T\varepsilon^2}{8\|c\|_1^4\|O\|^2}\right]. \tag{26}$$

This immediately gives us that for

$$T \geq \frac{8\|O\|^2\ln(2/\delta)\|c\|_1^4}{\varepsilon^2}, \tag{27}$$

$$\left|\mu - \mathrm{Tr}[O\ g(H)\ \rho_0\ g(H)^\dagger]\right| \leq \varepsilon/2,$$

with probability at least $1-\delta$. Now, in order to evaluate (b), we first apply triangle inequality, we obtain

$$\left|\mu - \mathrm{Tr}[O\ f(H)\rho_0 f(H)^\dagger]\right| \leq \left|\mu - \mathrm{Tr}[O\ g(H)\rho_0 g(H)^\dagger]\right| + \tag{28}$$

$$\left|\mathrm{Tr}[O\ f(H)\rho_0 f(H)^\dagger] - \mathrm{Tr}[O\ g(H)\rho_0 g(H)^\dagger]\right|. \tag{29}$$

The first term in the RHS of the above inequality is upper bounded by $\varepsilon/2$. In order to bound the second term, note that $\|f(H) - g(H)\| \leq \gamma$. For any such operators that are at most γ-separated, we can use Theorem 7 to obtain:

$$\left|\mathrm{Tr}[O\ f(H)\rho_0 f(H)^\dagger] - \mathrm{Tr}[O\ g(H)\rho_0 g(H)^\dagger]\right| \leq 3\|O\|\|f(H)\|\,\gamma \leq \varepsilon/2,$$

for γ upper bounded as in the statement of Theorem 8. So, overall we have

$$\left|\mu - \mathrm{Tr}[O\ f(H)\rho_0 f(H)^\dagger]\right| \leq \varepsilon,$$

which completes the proof. □

When $f(H)$ is unitary, the operator $\rho = f(H)\rho_0 f(H)^\dagger$ is a normalized quantum state and hence, $\mathrm{Tr}[O\rho] = \mathrm{Tr}[O\ f(H)\rho_0 f(H)^\dagger]$. However, this is not the case in general. In such scenarios, as described previously,

$$\mathrm{Tr}[O\rho] = \frac{\mathrm{Tr}[O\ f(H)\rho_0 f(H)^\dagger]}{\mathrm{Tr}[f(H)\rho_0 f(H)^\dagger]}.$$

Thus far, we could only obtain the numerator of the RHS of this equation. Next, we describe the procedure to obtain an estimate of the norm $\ell^2 = \mathrm{Tr}[f(H)\rho_0 f(H)^\dagger]$.

Note that we do not have an accurate knowledge of this quantity apriori. Provided we have knowledge of some rudimentary lower bound of the norm, using Algorithm 1 and Theorem 8, we can obtain an estimate $\tilde{\ell}$ of ℓ^2 to arbitrary accuracy by simply setting $O = I$. The crucial question in this regard is, how accurate should the estimate $\tilde{\ell}$ be such that $\mu/\tilde{\ell}$ is ε-close to $\mathrm{Tr}[O\rho]$?

Suppose we have knowledge of some lower bound (say ℓ_*) of this quantity, i.e. $\mathrm{Tr}[f(H)\rho_0 f(H)^\dagger] = \ell^2 \geq \ell_*$. Then, we prove the following:

Theorem 9 (Robustness of normalization factors)**.** *Let $\varepsilon \in (0,1)$, ρ_0 be some initial state and P be an operator. Furthermore, let $\ell_* \in \mathbb{R}^+$ satisfies $\ell^2 = \mathrm{Tr}[P\rho_0 P^\dagger] \geq \ell_*$, and O be some observable with $\|O\| \geq 1$. Suppose we obtain an estimate $\tilde{\ell}$ such that*

$$\left|\tilde{\ell} - \ell^2\right| \leq \frac{\varepsilon \ell_*}{3\|O\|}, \tag{30}$$

and some parameter μ such that,

$$\left|\mu - \mathrm{Tr}[O\ P\rho_0 P^\dagger]\right| \leq \frac{\varepsilon \ell_*}{3}, \tag{31}$$

then,

$$\left|\frac{\mu}{\tilde{\ell}} - \frac{\mathrm{Tr}[O\ P\rho_0 P^\dagger]}{\ell^2}\right| \leq \varepsilon.$$

Proof. This is proven in Sec. A of the Appendix. □

This Theorem tells us the precision with which both μ and $\tilde{\ell}$ should be estimated so that $\mu/\tilde{\ell}$ is ε-close to $\mathrm{Tr}[O\rho]$. The overall procedure makes use of Algorithm 1, and is formally stated via Algorithm 2. This involves running Algorithm 1 twice: first obtain μ as stated previously, and then obtain $\tilde{\ell}$, by setting $O = I$ (the identity matrix) and following the same steps.

The correctness of the Algorithm is stated formally through the following Theorem.

Theorem 10. *Let $\varepsilon, \delta \in (0,1)$, O be some observable and ρ_0 be some initial state, prepared in cost τ_{ρ_0}. Suppose $H \in \mathbb{C}^{N\times N}$ be a Hermitian matrix such that for some function $f : [-1,1] \mapsto \mathbb{R}$ and unitaries $\{U_j\}_j$, $\left\|f(H) - \sum_j c_j U_j\right\| \leq \gamma$, where*

$$\gamma \leq \frac{\varepsilon \ell_*}{18\|O\|\|f(H)\|},$$

and $\ell^2 = \mathrm{Tr}[f(H)\rho_0 f(H)^\dagger] \geq \ell_*$. Furthermore, suppose each unitary U_j is implementable in cost τ_j such that $\langle\tau\rangle = \sum_j c_j\tau_j/\|c\|_1$. Then Algorithm 2 outputs μ and $\tilde{\ell}$ such that

$$\left|\mu/\tilde{\ell} - \mathrm{Tr}[O\rho]\right| \leq \varepsilon,$$

Algorithm 2: `Single-ancilla-LCU` $(O, \{c_j, U_j\}, \rho_0)$

Choose some $T \in O\left(\frac{\|O\|^2 \ln(1/\delta)\|c\|_1^4}{\varepsilon^2 \ell_*^2}\right)$.

1. Run `Expectation-observable` $(O, \{c_j, U_j\}, \rho_0, T)$ to obtain μ such that
$$|\mu - \mathrm{Tr}[O\ f(H)\rho_0 f(H)^\dagger]| \leq \varepsilon \ell_*/3.$$
2. Run `Expectation-observable` $(I, \{c_j, U_j\}, \rho_0, T)$ to obtain $\tilde{\ell}$ such that
$$|\tilde{\ell} - \ell^2| \leq \frac{\varepsilon \ell_*}{3\|O\|}.$$
3. Output $\mu/\tilde{\ell}$.

with probability $(1-\delta)^2$*, using*

$$T = O\left(\frac{\|O\|^2 \|c\|_1^4 \ln(1/\delta)}{\varepsilon^2 \ell_*^2}\right)$$

repetitions of the quantum circuit in Fig. 2, the average cost of each coherent run is $2\langle\tau\rangle + \tau_{\rho_0}$.

Proof. Algorithm 2 calls Algorithm 1 twice: first to estimate μ and then to estimate $\tilde{\ell}$. For the upper bound of γ defined in the statement of the theorem, we can indeed obtain a μ, using Algorithm 1 such that

$$\left|\mu - \mathrm{Tr}[O\ f(H)\rho_0 f(H)^\dagger]\right| \leq \varepsilon \ell^*/3.$$

This follows from Theorem 8, by simply replacing ε with $\varepsilon\ell^*/3$. The number of iterations of Theorem 8 scales as

$$T_1 = O\left(\frac{\|O\|^2 \|c\|_1^4 \ln(1/\delta)}{\varepsilon^2 \ell_*^2}\right),$$

with each coherent run costing at most $\tau_{\rho_0} + 2\langle\tau\rangle$.

For obtaining the estimate $\tilde{\ell}$, we set $O = I$. Furthermore, in Theorem 8, we replace ε by $\frac{\varepsilon\ell_*}{3\|O\|}$. For these parameters, Algorithm 1 outputs $\tilde{\ell}$ such that

$$\left|\tilde{\ell} - \ell^2\right| \leq \frac{\varepsilon \ell^*}{3\|O\|}.$$

The total number of iterations is

$$T_2 = O\left(\frac{\|O\|^2 \|c\|_1^4 \ln(1/\delta)}{\varepsilon^2 \ell_*^2}\right),$$

with each coherent run having an average cost $\tau_{\rho_0} + 2\langle\tau\rangle$.

Finally from Theorem 9, we have that these estimates μ and $\tilde{\ell}$ satisfy

$$\left|\frac{\mu}{\tilde{\ell}} - \mathrm{Tr}[O\rho]\right| \leq \varepsilon,$$

where the total number of iterations scales as

$$T = T_1 + T_2 = O\left(\frac{\|O\|^2 \|c\|_1^4 \ln(1/\delta)}{\varepsilon^2 \ell_*^2}\right),$$

with each coherent run requiring an average cost $\tau_{\rho_0} + 2\langle\tau\rangle$. □

3.2.1 Comparison with the standard LCU procedure

First, we compare the performance of the *Single-Ancilla LCU* technique with the standard LCU procedure. In order to fairly compare the two approaches, we will analyze the various methods by which the standard LCU technique can be used to estimate $\mathrm{Tr}[O\rho]$. As described in Sec. 2.2, standard LCU prepares the quantum state

$$|\psi_t\rangle = \frac{\|f(H)|\psi_0\rangle\|}{\|c\|_1} |\bar{0}\rangle |\psi\rangle + |\Phi\rangle^{\perp}, \tag{32}$$

where

$$|\psi\rangle = \frac{f(H)|\psi_0\rangle}{\|f(H)|\psi_0\rangle\|}. \tag{33}$$

One obvious advantage of *Single-Ancilla LCU* is that it requires only a single ancilla qubit, while the standard LCU procedure and its variants described below require atleast $\lceil \log_2 M \rceil$ ancilla qubits, and more sophisticated controlled operations.

Also from Sec. 2.2, the cost of preparing $|\psi_t\rangle$ is $O(2\tau_R + \tau_Q + \tau_{\psi_0})$. Here, τ_{ψ_0} is the cost of preparing the initial state $|\psi_0\rangle$, τ_R is the cost of implementing the *prepare* unitary R that initializes the state of the ancilla registers, and τ_Q is the cost of implementing the multi-qubit controlled *select* unitary $Q = \sum_{j=1}^{M} |j\rangle\langle j| \otimes U_j$.

Let us begin by comparing τ_Q with $\langle\tau\rangle$, the average cost of implementing unitaries V_1 and V_2 in each coherent run of the *Single-Ancilla LCU* algorithm. Clearly, $\langle\tau\rangle \leq \tau_Q$. Also, in the general setting, the average cost $\langle\tau\rangle$ cannot exceed the cost of implementing the most expensive unitary, $\tau_{\max}$. At the same time, τ_Q cannot be lower than the cost of implementing the most expensive unitary. Combining these facts, we have

$$\langle\tau\rangle \leq \tau_{\max} \leq \tau_Q. \tag{34}$$

Indeed, when there is no apriori information about the U_j's, τ_Q can be much greater than even the cost of implementing the most expensive unitary, $\tau_{\max}$. In fact, in the worst case, if every U_j costs $\tau_{\max}$, $\tau_Q = O(M\tau_{\max})$. However, in particular cases, when the U_i's are related, it is possible that both τ_Q and $\langle\tau\rangle$ scale as $O(\tau_{\max})$. In fact in Lemma 8 of Ref. [26], it was shown that when the U_j's are powers of one single unitary, i.e. $U_j = Y^j$, for some unitary Y, $\tau_Q = O(\tau_{\max})$.

Thus, despite requiring only a single ancilla qubit, the cost per coherent run of *Single-Ancilla LCU* is lower than *Standard LCU* (as it additionally requires implementing the *prepare* unitary, requiring τ_R cost), provided the cost of implementing each U_i is the same for both procedures. Additionally, our method requires no multi-qubit controlled gates, which are unlikely to be implemented in the intermediate-term.

After having prepared $|\psi_t\rangle$, there are three ways in which $\mathrm{Tr}[O\rho_t]$ may be estimated. The overall cost varies based on the choice of the particular variant. We compare each of these three approaches with *Single-Ancilla LCU* by considering the cost of each coherent run, the number of ancilla qubits needed, and the overall cost (which is the product of the cost per run and the number of classical repetitions needed). Overall, there is a trade-off

between the cost of each run and the number of classical repetitions as we discuss below.

Standard LCU with amplitude amplification and classical repetitions: One can use the generic LCU procedure to prepare the state $|\psi_t\rangle$ with probability $\left\|f(H)\,|\psi_0\rangle\right\|^2/\|c\|_1^2$, and then apply quantum amplitude amplification in order to prepare the state $|\psi\rangle$, with a high probability. This procedure requires a cost

$$\Gamma_{\max} = O\left(\frac{\|c\|_1}{\sqrt{\ell_*}}\left(2\tau_R + \tau_Q + \tau_{\psi_0}\right)\right).$$

This is already higher than the cost of each run by *Single-Ancilla LCU*.

It is then possible to estimate $\mathrm{Tr}[O\rho] = \langle\psi|O|\psi\rangle$, by repeatedly measuring the observable O, which requires several classical repetitions of the procedure to prepare $|\psi\rangle$, costing $\Gamma_{\max}$.

For this, we first prove a general result. Broadly, consider any state preparation procedure V that prepares a quantum state $|x\rangle$ to (roughly) $\varepsilon/\|O\|$-accuracy. Then, for any observable O, we can obtain an ε-accurate estimate of $\langle x|O|x\rangle$ using $O(\,\|O\|^2\ln(1/\delta)/\varepsilon^2)$ runs of V. Formally, we have the following lemma:

Lemma 11. *Suppose, there exists a quantum procedure V which, starting from the state $|\psi_0\rangle$, prepares a quantum state $|\tilde{x}\rangle$, such that*

$$\left\||\tilde{x}\rangle - |x\rangle\right\| \leq \frac{\varepsilon}{2\sqrt{2}\|O\|}.$$

Then, to in order to output μ such that

$$|\mu - \langle x|O|x\rangle| \leq \varepsilon,$$

with probability at least $1-\delta$,

$$T = O\left(\frac{\|O\|^2\ln(1/\delta)}{\varepsilon^2}\right)$$

repetitions of the procedure V is required.

Proof. A single run of V, prepares the state $\rho_{\tilde{x}} = |\tilde{x}\rangle\langle\tilde{x}|$. Then, measuring the observable O, outputs an estimate that lies between $[-\|O\|,\|O\|]$. Then, in order to ensure that the estimate is $\varepsilon/2$-close to $\mathrm{Tr}[O\rho_{\tilde{x}}]$, with probability $1-\delta$, we require

$$T = O\left(\frac{\|O\|^2\ln(1/\delta)}{\varepsilon^2}\right),$$

repetitions of the procedure V, which follows from Hoeffding's inequality. Also,

$$\begin{aligned}
\left|\mathrm{Tr}[O\ \ |x\rangle\langle x|] - \mathrm{Tr}[O\ \ |\tilde{x}\rangle\langle\tilde{x}|]\right| &\leq \|O\|_\infty \left\||x\rangle\langle x| - |\tilde{x}\rangle\langle\tilde{x}|\right\|_1 && (35)\\
&\leq 2\|O\|\sqrt{1-|\langle x|\tilde{x}\rangle|} && (36)\\
&\leq \varepsilon/2. && (37)
\end{aligned}$$

Thus, after T repetitions, we are able to output μ such that

$$\begin{aligned}
|\mu - \langle x|O|x\rangle| &\leq |\mu - \mathrm{Tr}[O\ \ |\tilde{x}\rangle\langle\tilde{x}|] + \left|\mathrm{Tr}[O\ \ |x\rangle\langle x|] - \mathrm{Tr}[O\ \ |\tilde{x}\rangle\langle\tilde{x}|]\right| && (38)\\
&\leq \varepsilon/2 + \varepsilon/2 = \varepsilon. && (39)
\end{aligned}$$

□

We now compare our method to the standard LCU approach to estimate the expectation value of O. For any H, suppose

$$\left\| f(H) - \sum_{j=1}^{M} c_j U_j \right\| \leq \frac{\varepsilon}{2\sqrt{2}\|O\|},$$

such that $\|c\|_1 = \sum_j c_j$. Then the standard LCU procedure prepares a quantum state $|\tilde{\psi}\rangle$ such that

$$\left\| |\psi\rangle - |\tilde{\psi}\rangle \right\| \leq \Theta\left(\frac{\varepsilon}{2\sqrt{2}\|O\|}\right).$$

Furthermore, from Lemma 11,

$$T = O\left(\frac{\|O\|^2 \ln(1/\delta)}{\varepsilon^2}\right)$$

repetitions of the standard LCU procedure to estimate $\langle\psi|O|\psi\rangle$ with probability at least $1-\delta$, where each such run costs $\Gamma_{\max}$. Thus the overall cost of $O(\Gamma_{\max}||O||^2/\varepsilon^2)$ has a better dependence on $\|c\|_1$ and ℓ_*, as compared to *Single-Ancilla LCU*.

However, in addition to requiring more ancilla qubits and sophisticated controlled operations, the cost of each run of this procedure is significantly larger than our algorithm. Each run of our algorithm classically samples only two U_j's and implements controlled (over a single ancilla qubit) versions of these sampled unitaries. Thus, the average cost of each run for our procedure $O(\langle\tau\rangle + \tau_{\psi_0}) < \Gamma_{\max}$. In order to estimate $\langle\psi|O|\psi\rangle$ however, our procedure requires more classical repetitions, and so the overall cost of the *Single-Ancilla LCU* procedure is higher.

Coherent estimation of $\mathrm{Tr}[\mathbf{O}\rho]$ using Standard LCU: It is possible to use quantum amplitude estimation to coherently estimate $\langle\psi|O|\psi\rangle$ to an additive accuracy ε. This procedure then does not require more than $O(1)$ classical repetitions, and hence the overall cost is lower than the *Single Ancilla LCU procedure*. However, this also increases the cost of each run, as well as number of ancilla qubits substantially.

First, we need to access (any general, non-unitary) O via a block encoding. For instance, in many cases O is itself a linear combination of unitaries. Constructing this block encoding requires additional ancilla qubits and controlled operations. Note that modern versions of quantum amplitude estimation procedure require only one additional ancilla qubit [78, 79].

Let U_O be an $(\alpha_O, a_O, 0)$-block encoding of O, requiring cost τ_O. Then, if T_ψ is cost of preparing the state $|\psi\rangle$, then the cost of estimating $\mathrm{Tr}[O\rho]$ to ε-additive accuracy is given by

$$\Gamma_{\max} = O\left(\frac{\alpha_O}{\varepsilon}\left(T_\psi + T_O\right)\right) = O\left(\frac{\alpha_O}{\varepsilon}\left(\frac{\|c\|_1}{\sqrt{\ell_*}}\left(\tau_Q + 2\tau_R + \tau_{\psi_0}\right) + T_O\right)\right),$$

where we have replaced T_ψ by the cost of preparing $|\psi\rangle$ using standard LCU with quantum amplitude amplification. As compared to the *Single-Ancilla LCU* method, we find that the overall complexity of coherently estimating $\mathrm{Tr}[O\rho]$ via standard LCU is lower. In particular, the overall dependence on precision is now $O(1/\varepsilon)$, instead of $O(1/\varepsilon^2)$. However the cost of of each coherent run is substantially increased. Moreover, the total number of ancilla qubits needed is now $O(a_O + \log(M))$.

Method	No. of. Ancilla qubits	Cost of each coherent run	Classical repetitions
Standard LCU (with QAA + classical repetitions)	$O(\log M)$	$O\left(\frac{\|c\|_1}{\sqrt{\ell_*}}\left(2\tau_R+\tau_Q+\tau_{\psi_0}\right)\right)$	$O\left(\frac{\|O\|^2}{\varepsilon^2}\right)$
Standard LCU (with QAE)	$O(a_O+\log M)$	$O\left(\frac{\alpha_O}{\varepsilon}\left(\frac{\|c\|_1}{\sqrt{\ell_*}}\left(\tau_Q+2\tau_R+\tau_{\psi_0}\right)+T_O\right)\right)$	$O(1)$
Standard LCU (without QAA or QAE)	$O(\log M)$	$O\left(2\tau_R+\tau_Q+\tau_{\psi_0}\right)$	$O\left(\frac{\|O\|^2\|c\|_1^2}{\varepsilon^2\ell_*}\right)$
Single-Ancilla LCU (this work)	1	$2\langle\tau\rangle+\tau_{\psi_0}$	$O\left(\frac{\|O\|^2\|c\|_1^4}{\varepsilon^2\ell_*^2}\right)$

Table 1: For any $H\in\mathbb{C}^{N\times N}$, this table compares the *Single-Ancilla LCU* method with standard LCU for estimating $\mathrm{Tr}[O\rho]$ to additive accuracy ε, for any measurable observable O. Here, ρ is defined in Eq. (1) such that $f(H)\approx\sum_i c_iU_i$ and $\|c\|_1=\sum_i c_i$. Here τ_{ψ_0} is the cost the unitary that prepares the initial state, τ_R and τ_Q denote the cost of the *prepare* unitary R, and the *select* unitary Q, respectively. Also, $\langle\tau\rangle=\sum_j c_j\tau_j/\|c\|_1$, where τ_j is the cost of implementing U_j. As described in Sec. 3.4, the standard LCU method can be used to estimate $\mathrm{Tr}[O\rho]$ in three ways: (i) with quantum amplitude amplification (QAA) followed by classical repetitions, (ii) with quantum amplitude estimation (QAE) to coherently estimate the desired expectation value, and (iii) incoherently with just classical repetitions. In particular, for method (ii), we assume that there exists an $(\alpha_O,a_O,0)$-block encoding of O, implementable in a cost τ_O. As compared to the *Single-Ancilla LCU* procedure, each of these variants require higher ancilla qubits and multi-qubit controlled operations. Furthermore, as $\langle\tau\rangle\leq\tau_{\max}\leq\tau_Q$, the cost of each coherent run of the *Single-Ancilla LCU* is lower than *Standard LCU*, provided the cost of implementing the unitaries U_j is the same for both. This is because: (a) Standard LCU and its variants require implementing the *prepare* unitary, requiring cost τ_R, and (b) also, $\tau_{\max}$, which is upper bounded by the maximum cost of implementing the most expensive U_j, cannot exceed τ_Q, the cost of implementing the *select* unitary. However, the overall complexity (product of the cost of each coherent run and the total number of classical repetitions) of *Standard LCU* scales better than *Single-Ancilla LCU*.

Standard LCU without amplitude amplification or estimation: Without amplitude amplification, the standard LCU procedure, prepares the desired quantum state in the second register with probability $p=\left\|f(H)\left|\psi_0\right\rangle\right\|^2/\|c\|_1^2$. Then this would require cost,

$$\Gamma_{\max}=O\left(2\tau_R+\tau_Q+\tau_{\psi_0}\right),$$

which is less than the previous two variants, but still higher than *Single Ancilla LCU* (as $\tau_Q\geq\tau_{\max}\geq\langle\tau\rangle$). Then, the following strategy can be used to estimate the desired expectation value: Measure the first register. Whenever the first register is in $|\bar{0}\rangle$, apply the observable O to the state of the second register. The first register is in state $|\bar{0}\rangle$ with probability p, and so, $O(\ln(1/\delta)/p)$ classical repetitions are needed to obtain the same. Overall, these many repetitions of the *Standard LCU* procedure are needed per measurement of O. Also, one needs to measure O, a total of $O(\ ||O||^2\ln(1/\delta)/\varepsilon^2)$ to estimate $\mathrm{Tr}[O\rho]$ with ε-additive accuracy. Thus, combining these, we obtain that overall

$$T=O\left(\frac{\|O\|^2\|c\|_1^2\ln^2(1/\delta)}{\varepsilon^2\ell_*}\right),$$

repetitions are needed to output the desired expectation value with ε-additive accuracy and at least $(1-\delta)^2$ success probability. This is quadratically lower than the total number

of iterations required by the *Single Ancilla LCU* technique, in terms of $\|c\|_1$ and ℓ_*. Thus, even in the worst case, this method has a cost per coherent run that is higher than our method, while requiring quadratically lower overall cost. Recall that unlike this procedure, our method required (i) only a single ancilla qubit, (ii) no multi-qubit controlled operations.

Overall, *Single-Ancilla LCU* provides a generic framework to implement any LCU, using only a single ancilla qubit, and no multi-qubit controlled operations. This makes the method particularly suitable for early fault-tolerant quantum computers. The cost of each coherent run of *Single-Ancilla LCU* is lower than *Standard LCU*, provided the unitaries U_j's are implemented at the same cost for both methods. However, the overall cost of estimating expectation values is lower for standard LCU approaches, as the *Single-Ancilla LCU* technique requires more classical repetitions. In Table 1, we compare the cost (number of ancilla qubits, cost of each coherent run and the number of classical repetitions) of estimating expectation values of observables using the standard LCU approach, with *Single-Ancilla LCU*. The wide applicability of LCU in the development of various quantum algorithms, also implies that our method can be employed to develop novel algorithms for these problems. We apply this technique for several problems of interest such as Hamiltonian simulation (Sec. 4), ground state property estimation (Sec. 5.2), and extracting useful information from the solution of quantum linear systems (Sec. 6.2). Next, we discuss the *Analog LCU* method.

3.3 Analog LCU: coupling a discrete primary system to a continuous-variable ancilla

Suppose we have some Hermitian matrix $H \in \mathbb{C}^{N\times N}$ of unit spectral norm and we wish to implement $f(H)$ for some function $f : [-1, 1] \mapsto \mathbb{R}$ which satisfies:

$$\left| f(x) - \int_a^b dz\ c(z) \cdot e^{-itxz} \right| \leq \varepsilon,$$

where $c : \mathbb{R} \mapsto \mathbb{R}^+ \backslash \{0\}$.

Now suppose H is coupled to a continuous variable system such that the resulting interaction Hamiltonian is $H' = H \otimes \hat{z}$. Suppose the first register is prepared in some initial state $|\psi_0\rangle$ and the ancilla system is prepared in the continuous-variable quantum state

$$|\bar{0}\rangle_c = \int_a^b dz\ \sqrt{\frac{c(z)}{\|c\|_1}}\, |z\rangle\,,$$

where $\|c\|_1 = \int_a^b dz\ |c(z)|$. For instance, $\hat{z}$ can represent a degree of freedom (position or momentum) of a one-dimensional quantum Harmonic oscillator, and the state $|\bar{0}\rangle_c$, could be its ground state (a Gaussian), a free resource state for continuous variable systems. For several of our applications, we shall see that this is indeed the case.

Now we shall simply evolve the system according to the interaction Hamiltonian H' to obtain

$$|\eta_t\rangle = e^{-iH't}\, |\psi_0\rangle\, |\bar{0}\rangle_c = \int_a^b dz\ \sqrt{\frac{c(z)}{\|c\|_1}} e^{-iHtz}\, |\psi_0\rangle\, |z\rangle \tag{40}$$

$$= \frac{1}{\|c\|_1} \int_a^b dz\ c(z) e^{-iHtz}\, |\psi_0\rangle\, |\bar{0}\rangle_c + |\Phi\rangle^{\perp}\,, \tag{41}$$

where $|\Phi\rangle^{\perp}$ is a quantum state (not normalized) such that $(I \otimes |\bar{0}\rangle_c \langle\bar{0}|_c)|\Phi\rangle^{\perp} = 0$. We arrive at the Eq. (41) from Eq. (40) by observing that

$$(I \otimes \langle\bar{0}|_c) \int_a^b dz \, \sqrt{\frac{c(z)}{\|c\|_1}} e^{-iHtz} |\psi_0\rangle |z\rangle = \frac{1}{\|c\|_1} \int_a^b dz \, dz' \, \delta_{z,z'} \sqrt{c(z)c(z')^*} e^{-itHz} |\psi_0\rangle .$$

Thus, we have prepared a quantum state that is $O(\varepsilon/\|c\|_1)$-close to

$$|\psi\rangle = \frac{f(H)}{\|c\|_1} |\psi_0\rangle |\bar{0}\rangle_c + |\Phi\rangle^{\perp} . \tag{42}$$

Now post-selecting on having $|\bar{0}\rangle_c$ in the second register we obtain a state that is ε-close to $f(H)|\psi_0\rangle / \|f(H)|\psi_0\rangle\|$ in the first register with probability $\|f(H)|\psi_0\rangle\|^2 / \|c\|_1^2$. We will use this procedure to develop an analog quantum algorithm for preparing ground states of Hamiltonians in Sec. 5.

This continuous-time algorithm can be naturally generalized to the scenario where we want to implement $f(H)$ for some function $f : [-1, 1] \mapsto \mathbb{R}$ such that

$$\left| f(x) - \int_{a_1}^{b_1} dz_1 \, c(z_1) \int_{a_2}^{b_2} dz_2 \, c(z_2) \cdots \int_{a_k}^{b_k} dz_k \, c(z_k) e^{-itxz_1 z_2 \cdots z_k} \right| \leq \varepsilon.$$

This can be implemented by coupling the Hamiltonian H with k different ancillary continuous-variable systems such that the effective interaction Hamiltonian is $\tilde{H} = H \otimes \hat{z}_1 \otimes \cdots \otimes \hat{z}_k$. The j-th ancilla system is prepared in the quantum state

$$|\bar{0}\rangle_{c_j} = \int_{a_j}^{b_j} dz_j \, \sqrt{\frac{c(z_j)}{\|c_j\|_1}} |z_j\rangle .$$

Then by evolving the initial state $|\psi_0\rangle |\bar{0}\rangle_{c_1} \cdots |\bar{0}\rangle_{c_k}$ according to $\tilde{H}$ for time t results in a quantum state that is $O\left(\frac{\varepsilon}{\Pi_{j=1}^k \|c_j\|_1}\right)$-close to

$$|\eta_t\rangle = \frac{f(H)}{\Pi_{j=1}^k \|c_j\|_1} |\psi_0\rangle |\bar{0}\rangle_{c_1} \cdots |\bar{0}\rangle_{c_k} + |\Phi\rangle^{\perp} . \tag{43}$$

In Sec. 6, our analog quantum linear systems algorithm requires coupling the system Hamiltonian to two ancillary systems, which is captured by this generalization of analog LCU. Interestingly, for the applications we consider, the ancillary states are the ground or the first excited state of a one-dimensional quantum Harmonic oscillator or the ground state of a "particle in a ring".

3.4 Ancilla-free LCU: Randomized unitary sampling

As in the previous sections, suppose we are given a Hermitian matrix H and we wish to implement $f(H)$, which can be approximated by a linear combination of unitaries. That is, for some $\gamma \in [0, 1)$,

$$\left\| f(H) - \sum_{j=1}^{M} c_j U_j \right\| \leq \gamma,$$

for unitaries U_j, $c_j \in \mathbb{R}^+ \backslash \{0\}$, and $\|c\|_1 = \sum_{j=1}^M c_j$. Here, we assume that we are interested in the projection of $f(H)|\psi_0\rangle$ in some subspace of interest, and in the resulting average

success probability. Then, we formally prove that instead of implementing *Standard LCU*, it suffices to simply sample U_j according to the distribution of the LCU coefficients. The resulting projection on the subspace of interest is at least as high, on average. This is the key idea behind the *Ancilla-free LCU* technique.

Suppose we obtain a unitary V sampled from the ensemble $\mathcal{D} \sim \{c_j/\|c\|_1, U_j\}$, and apply it to some initial state ρ_0, such that $\rho = V\rho_0 V^\dagger$. Then, on average, this leads to the following mixed state

$$\bar{\rho} = \mathbb{E}[\rho] = \mathbb{E}[V\rho_0 V^\dagger] = \frac{1}{\|c\|_1}\sum_{j=1}^{M} c_j U_j \rho_0 U_j^\dagger. \tag{44}$$

If each U_j costs τ_j, then the average cost of preparing $\bar{\rho}$ is given by $\langle\tau\rangle = \sum_j c_j\tau_j/\|c\|_1$. As with *Single-Ancilla LCU*, this is upper bounded by $\tau_{\max} = \max_j \tau_j$, the cost of implementing the most expensive U_j. Now, for some projector Π onto the subspace of interest, the average projection of the resulting density matrix ρ, is given by $\mathbb{E}[\mathrm{Tr}(\Pi\rho)] = \mathrm{Tr}[\Pi\bar{\rho}]$. Then, it is possible to prove that the projection of $\bar{\rho}$ in this subspace is at least as large as the projection of $f(H)\rho_0 f(H)^\dagger$, if $\|c\|_1 \leq 1$. We formally prove this via the following theorem:

Theorem 2 (Randomized unitary sampling)**.** *Let $H \in \mathbb{C}^{N\times N}$ is a Hermitian matrix. Also let $\varepsilon \in (0,1)$ and suppose $f : [-1,1] \mapsto \mathbb{R}$ be some function such that*

$$\left\| f(H) - \sum_{j=1}^{M} c_j U_j \right\| \leq \frac{\varepsilon}{3\|f(H)\|},$$

for some unitaries U_j and $c_j \in \mathbb{R}\backslash\{0\}$ such that $\|c\|_1 = \sum_j |c_j| \leq 1$. Suppose V is a unitary sampled from the ensemble $\{c_j/\|c\|_1, U_j\}$, and applied to some initial state ρ_0. Then, the average density matrix, defined as

$$\bar{\rho} = \mathbb{E}\left[V\rho_0 V^\dagger\right] = \frac{1}{\|c\|_1}\sum_{j=1}^{M} c_j U_j \rho_0 U_j^\dagger,$$

satisfies,

$$\mathrm{Tr}\left[\Pi\bar{\rho}\right] \geq \mathrm{Tr}[\Pi f(H)\rho_0 f(H)^\dagger] - \varepsilon,$$

for any projector Π acting on the state space of $\bar{\rho}$.

Proof. Let $g(H) = \sum_{j=1}^{M} c_j U_j$. Then, given any initial state $\rho_0 = |\psi_0\rangle\langle\psi_0|$, if R is the *prepare* unitary while $Q = \sum_j |j\rangle\langle j| \otimes U_j$ is the *select* unitary, then the standard LCU procedure first prepares the state $|\psi_t'\rangle$ where

$$\begin{aligned} |\psi_t'\rangle &= Q(R\otimes I)|\bar{0}\rangle|\psi_0\rangle \\ &= \sum_{j=1}^{M}\sqrt{\frac{c_j}{\|c\|_1}}|j\rangle U_j|\psi_0\rangle. \end{aligned}$$

Then, the standard LCU procedure implements $(R^\dagger \otimes I)$ to prepare:

$$|\psi_t\rangle = |\bar{0}\rangle\frac{g(H)}{\|c\|_1}|\psi_0\rangle + |\Phi^\perp\rangle.$$

Note that in the last step, the unitary $R^\dagger$ acts only on the first register, whereas we are interested in the measurement outcomes of the second register. That is, consider the projection $I \otimes \Pi$. Indeed, as Π acts only on the second register, we can ignore the last step of *Standard LCU*, (i.e. the action of $R^\dagger \otimes I$) and by the equivalence of partial trace, we obtain the following equality:

$$\mathrm{Tr}\left[(I \otimes \Pi) |\psi_t\rangle \langle\psi_t|\right] = \mathrm{Tr}\left[(I \otimes \Pi) |\psi_t'\rangle \langle\psi_t'|\right]. \tag{45}$$

Next, by randomly sampling U_j according to $\{c_j/\|c\|_1\}$, we prepare, on average, the density matrix $\bar{\rho}$. It is easy to verify that $|\psi_t'\rangle$ is a purification of $\bar{\rho}$. Then it follows that,

$$\mathrm{Tr}\left[(I \otimes \Pi) |\psi_t'\rangle \langle\psi_t'|\right] = \mathrm{Tr}[\Pi\bar{\rho}]. \tag{46}$$

Then by combining Eq. (45) and Eq. (46), we obtain:

$$\mathrm{Tr}[\Pi\bar{\rho}] = \mathrm{Tr}\left[(I \otimes \Pi) |\psi_t\rangle \langle\psi_t|\right]. \tag{47}$$

Thus, we have

$$\begin{aligned}
\mathrm{Tr}[\Pi\bar{\rho}] &= \mathrm{Tr}[(I \otimes \Pi) |\psi_t\rangle \langle\psi_t|] && (48)\\
&= \langle\psi_t| \left(|\bar{0}\rangle \langle\bar{0}| \otimes \Pi\right) |\psi_t\rangle + \langle\psi_t| \left[(I - |\bar{0}\rangle \langle\bar{0}|) \otimes \Pi\right] |\psi_t\rangle && (49)\\
&= \frac{1}{\|c\|_1^2} \langle\psi_0| g(H)^\dagger \Pi g(H) |\psi_0\rangle + \langle\Phi^\perp| \left[(I - |\bar{0}\rangle \langle\bar{0}|) \otimes \Pi\right] |\Phi^\perp\rangle && (50)\\
&\geq \frac{1}{\|c\|_1^2} \langle\psi_0| g(H)^\dagger \Pi g(H) |\psi_0\rangle && (51)\\
&\geq \langle\psi_0| f(H)^\dagger \Pi f(H) |\psi_0\rangle - \varepsilon, && (52)
\end{aligned}$$

where in the last line we have invoked Theorem 7, and the fact that $\|c\|_1 \leq 1$. This completes the proof. □

Theorem 2 shows that if we are only interested in the projection of $f(H)\rho_0 f(H)^\dagger$ in some subspace of interest (say Π is a projector onto this subspace), we can drop the ancilla registers of LCU. Instead, we simply sample a unitary V according to $\{c_j/\|c\|_1, U_j\}$. Then, on average, the projection of $\rho = V\rho_0 V^\dagger$ in this subspace is at least as large. A similar result naturally extends to the analog LCU framework as well. While the *Standard LCU* procedure requires $\lceil \log M \rceil$ ancilla qubits and sophisticated multi-qubit control gates to implement $f(H)\rho_0 f(H)^\dagger$ before making the projective measurement, our method establishes that this can be done without any ancilla qubits. At the same time, the average cost per coherent run of this method is given by $\langle\tau\rangle$, which, as discussed previously is upper bounded by τ_Q (the cost of implementing the *select* gate for *Standard LCU*). Note that sampling the unitaries does not guarantee a higher success probability: it is at least as high, on average. However, one can leverage *Ancilla-free LCU* when we only care about the average success probability of the underlying projective measurement. This is indeed the case for quantum spatial search [32, 33] as we shall discuss next.

In the spatial search problem, we are interested in the expected number of steps required to find a subset of nodes (marked nodes) of any reversible Markov chain P. Starting from the stationary distribution of P, the number of steps required by a classical random walk, on average, to find a marked node is known as the hitting time (HT). Analogously, for quantum walks, starting from a quantum state proportional to the stationary distribution of P, we are also interested in finding the expected number of steps after which a

quantum walk has a large overlap with the marked nodes of P. In fact, demonstrating that quantum walks require $O(\sqrt{HT})$ steps for any P and any number of marked nodes, had been a long-standing open problem. Through a series of works [39, 40, 41], a quadratic speedup had been proven only in specific cases (such as particular instances of graphs or when only a single node is marked). The full generic quadratic speedup (up to a log factor) has only been recently proven [32], using the *Standard LCU* procedure. We show that the framework of quantum spatial search fits into the framework of *Ancilla-free LCU*: here we are interested in the average projection of the quantum walk on to the marked subspace. Thus, the *Standard LCU* of Ref. [32] can be bypassed, which leads to saving on ancilla qubits while obtaining the same quadratic speedup. Moreover, we show that new quantum spatial search algorithms can be obtained using this framework which also retain a generic quadratic speedup. These results have been described in detail in Sec. 7.

4 Applying Single-Ancilla LCU: Hamiltonian simulation

In this section, we will use the *Single-Ancilla LCU* procedure to develop a quantum Hamiltonian simulation algorithm that is tailored to early fault-tolerant quantum computers. Consider any Hamiltonian H which is a linear combination of Pauli operators, i.e. $H = \sum_{k=1}^{L} \lambda_k P_k$, where P_k is a sequence of Pauli operators, such that $\beta = \sum_k |\lambda_k|$. Let O be some observable and ρ_0 be some initial state. Then, we outline a procedure using Algorithm 1 (and Theorem 8) that outputs μ such that

$$\left|\mu - \mathrm{Tr}[O\ e^{-iHt}\rho_0 e^{iHt}]\right| \leq \varepsilon \tag{53}$$

Since, $f(H) = e^{-iHt}$ is unitary, $\rho_t = f(H)\rho_0 f(H)^\dagger$ is a normalized quantum state. Consequently, $\mathrm{Tr}[O\rho_t]$ can be obtained by simply using Algorithm 1 as we do not need to estimate the norm separately. Thus it suffices to apply Theorem 8 to output μ.

We decompose e^{-itH} as an approximate linear combination of unitaries, for which we use ideas from the Truncated Taylor series method by Berry et al.[25], as well the LCU decomposition from [80]. We summarize the key ideas here, while the details can be found in the Appendix (Sec. B). Therein, we show that we can divide the time evolution operator e^{-itH} into r segments and truncate the Taylor series expansion of each such segment after K terms. More precisely, we show that if,

$$\tilde{S}_r = \sum_{k=0}^{K} \frac{(-it\tilde{H}/r)^k}{k!},$$

then $\left\|\tilde{S}_r - e^{-itH/r}\right\| \leq \gamma/r$, for some

$$K = O\left(\frac{\log(r/\gamma)}{\log\log(r/\gamma)}\right).$$

Moreover, we prove in the Appendix (Sec. B) that each $\tilde{S}_r$ can be decomposed as a linear combination of unitaries $\sum_{j\in M} \alpha_j U_j$, where M can be defined as a multi-indexed set, i.e.

$$M = \{(k, \ell_1, \ell_2, \cdots \ell_k, m) : 0 \leq k \leq K; \ell_1, \ell_2, \cdots \ell_k, m \in \{1, 2, \cdots, L\}\},$$

and each U_j is a product of $k \leq K$ Pauli operators and a single Pauli rotation, given by

$$U_j = (-i)^k P_{\ell_1} P_{\ell_2} \cdots P_{\ell_k} e^{-i\theta_m P_m}. \tag{54}$$

Moreover, the sum of the coefficients of this LCU decomposition satisfies $\sum_j |\alpha_j| \leq \exp[\beta^2 t^2/r^2]$.

Then, the product of r such segments is also an LCU. That is, $S = \tilde{S}_r^r$ is itself an LCU, i.e. $S = \sum_j c_j W_j, \|c\|_1 = (\sum_j |\alpha_j|)^r$ can be proven to be $O(1)$, for $r = \beta^2 t^2$. We show that for $\gamma = \frac{\varepsilon}{6\|O\|}$, the operator $S/\|c\|_1$ is close to the overall time evolution operator. That is,

$$\left\| e^{-itH} - S/\|c\|_1 \right\| \leq \frac{\varepsilon}{6\|O\|}.$$

This allows us to leverage Theorem 8 and Algorithm 1 to estimate $\mathrm{Tr}[O\rho_t]$ to ε-accuracy. In order to apply Algorithm 1, we intend to sample V_1, V_2 such that $\mathbb{E}[V_1] = \mathbb{E}[V_2] = S/\|c\|_1$. We provide a brief sketch of the sampling strategy next.

Sampling $\mathbf{V_1}$ and $\mathbf{V_2}$: While the sampling strategy is described in detail in the Appendix (Sec. B), we provide a brief outline here. We sample some integer $k \in [0, K]$, according to $\alpha_j / \sum_j |\alpha_j|$, and select $k+1$ unitaries comprising of k Pauli operators and a single Pauli rotation. This is then repeated r times to obtain a product of unitaries $W = W_r \cdots W_1$, such that $\mathbb{E}[W] = S/\|c\|_1$. This allows us to use Algorithm 1 and Theorem 8.

Running time: The circuit corresponding to Algorithm 1 implements controlled (anti-controlled) V_1 (V_2). So, in order to estimate the cost of each run of the Algorithm, we need to estimate the gate depth of V_1 and V_2. The sampling procedure described above, outputs a product of r unitaries $W = W_r \cdots W_1$ such that each W_j is itself a product of at most $K+1$ unitaries - K Pauli operators and a single Pauli rotation. Thus the gate depth of V_1 is at most $(K+1)r$. So, the overall gate depth is given by $2(K+1)r = O(Kr)$. From the choice of r and K, this implies that the gate depth for each run is at most $2\tau_{\max} + \tau_{\rho_0}$, where,

$$\tau_{\max} = O(Kr) = O\left(\beta^2 t^2 \frac{\log(\beta t \|O\|/\varepsilon)}{\log\log(\beta t \|O\|/\varepsilon)}\right). \tag{55}$$

The total number of repetitions needed can be obtained from Theorem 8. In this case, for the choice of $r = \beta^2 t^2$, we have $\|c\|_1 = O(1)$. So,

$$T = O\left(\frac{\|O\|^2 \ln(1/\delta)}{\varepsilon^2}\right), \tag{56}$$

repetitions ensures that Algorithm 1 outputs a μ such that $|\mu - \mathrm{Tr}[O\ e^{-iHt}\rho_0 e^{iHt}]| \leq \varepsilon$, with probability $1-\delta$. The overall gate depth is given by

$$O(\tau_{\max}.T) = O\left(\frac{\beta^2 t^2 \|O\|^2}{\varepsilon^2} \ln(1/\delta) \cdot \frac{\log(\beta t \|O\|/\varepsilon)}{\log\log(\beta t \|O\|/\varepsilon)}\right).$$

We formally state our results via the following Theorem:

Theorem 12. *Let H be a Hamiltonian such that $H = \sum_{k=1}^{L} c_k P_k$, where P_k is a sequence of Pauli operators and $c_k > 0$ such that $\beta = \sum_k c_k$. Suppose ρ_0 be some initial state, prepared in gate depth τ_{ρ_0} and O be any observable. Then provided,*

$$\gamma \leq \frac{\varepsilon}{6\|O\|},$$

such that $\left\|e^{-itH} - S\right\| \leq \gamma$ *and,*

$$T = O\left(\frac{\|O\|^2 \ln(1/\delta)}{\varepsilon^2}\right),$$

Algorithm 1 outputs μ *such that*

$$|\mu - \mathrm{Tr}[O\rho_t]| \leq \varepsilon,$$

using T *repetitions of the quantum circuit in Fig. 2, with probability* $1-\delta$*. Each such run has gate depth at most* $2\tau_{\max} + \tau_{\rho_0}$ *such that*

$$\tau_{\max} = O\left(\beta^2 t^2 \frac{\log(\beta t \|O\| / \varepsilon)}{\log\log(\beta t \|O\| / \varepsilon)}\right).$$

Proof. We use Theorem 8, substituting $f(H) = e^{-itH}$ and so $\|f(H)\| = 1$. For the choice of γ, we require only

$$T = O\left(\frac{\|O\|^2 \ln(1/\delta)}{\varepsilon^2}\right),$$

repetitions of Algorithm 8. The gate depth of each coherent run is at most $2\tau_{\max} + \tau_{\rho_0}$, where $\tau_{\max}$ has been obtained in Eq. (55). □

Comparison with prior works: Let us now compare the cost of our quantum simulation procedure with other established techniques. Throughout, we consider that H is a linear combination of unitaries, as stated in Eq. (A15). We shall compare the cost of each Hamiltonian simulation algorithm to estimate $\mathrm{Tr}[O\rho_t]$, where $\rho_t = e^{-itH}\rho_0 e^{itH}$. Note that in order to estimate this quantity with additive accuracy ε, these methods need to prepare a quantum state that is $\varepsilon/\|O\|$-close to ρ_t (from Lemma 11). We measure the cost in terms of gate depth per coherent run, number of ancilla qubits required, number of classical repetitions needed, and the overall gate depth given by the product of the gate depth per coherent run and the number of classical repetitions. Having prepared ρ_t, there are two ways in which $\mathrm{Tr}[O\rho_t]$ can be estimated. First, by simply measuring O, in which case $O(\|O\|^2/\varepsilon^2)$ repetitions are needed in order to output an ε-accurate estimate of $\mathrm{Tr}[O\rho_t]$, with $\Omega(1)$ probability. The second technique involves estimating this quantity coherently using quantum amplitude estimation. For this, we assume access to a $(\alpha_O, a_O, 0)$-block encoding of O, which can be implemented in gate depth T_O. Then, if T_H is the gate depth of the underlying Hamiltonian simulation procedure (to output ρ_t with $O(\varepsilon/\alpha_O)$ accuracy), quantum amplitude estimation estimates $\mathrm{Tr}[O\rho_t]$ to additive accuracy ε in

$$O\left(\frac{\alpha_O}{\varepsilon}(T_H + T_O)\right),$$

with a constant success probability. We are now in a position to compare the cost of our Hamiltonian simulation algorithm with other methods.

Let us begin by considering the first order Trotter method [48]. This algorithm has a gate depth of $O(L\beta^2 t^2 \|O\| / \varepsilon)$. So the circuit depth of our procedure has an exponentially better dependence on the precision. Additionally, the total number of terms of H, i.e. L can be quite large. For instance for several quantum chemistry Hamiltonians $L = \mathrm{poly}(n)$ for an n-qubit Hamiltonian, while $\beta \ll L$. However, Trotter-based methods such as First and higher order methods, and qDRIFT do not require any ancilla qubit. Even if $\mathrm{Tr}[O\rho_t]$ is

Algorithm	Variant	No. of. Ancilla qubits	Gate depth per coherent run	Classical repetitions
1st order Trotter [48]	Incoherent	0	$O(L\beta^2t^2\\|O\\|/\varepsilon)$	$O\left(\frac{\\|O\\|^2}{\varepsilon^2}\right)$
	Coherent	$O(a_O)$	$O\left(\frac{\alpha_O^2 L\beta^2t^2}{\varepsilon^2}+\frac{\alpha_O T_O}{\varepsilon}\right)$	$O(1)$
qDRIFT [60]	Incoherent	0	$O\left(\beta^2t^2\\|O\\|/\varepsilon\right)$	$O\left(\frac{\\|O\\|^2}{\varepsilon^2}\right)$
	Coherent	$O(a_O)$	$O\left(\frac{\alpha_O^2\beta^2t^2}{\varepsilon^2}+\frac{\alpha_O T_O}{\varepsilon}\right)$	$O(1)$
$2k$-order Trotter [48]	Incoherent	0	$O\left(L(\beta t)^{1+\frac{1}{2k}}\left(\frac{\\|O\\|}{\varepsilon}\right)^{\frac{1}{2k}}\right)$	$O\left(\frac{\\|O\\|^2}{\varepsilon^2}\right)$
	Coherent	$O(a_O)$	$O\left(L(\beta t)^{1+\frac{1}{2k}}\left(\frac{\alpha_O}{\varepsilon}\right)^{1+\frac{1}{2k}}+\frac{\alpha_O T_O}{\varepsilon}\right)$	$O(1)$
Truncated Taylor Series [25]	Incoherent	$O\left(\log(L)\frac{\log(\\|O\\|t/\varepsilon)}{\log\log(\\|O\\|t/\varepsilon)}\right)$	$O\left(L\beta t\frac{\log(\beta t\\|O\\|/\varepsilon)}{\log\log(\beta t\\|O\\|/\varepsilon)}\right)$	$O\left(\frac{\\|O\\|^2}{\varepsilon^2}\right)$
	Coherent	$O\left(a_O+\log(L)\frac{\log(\alpha_O t/\varepsilon)}{\log\log(\alpha_O t/\varepsilon)}\right)$	$O\left(\frac{\alpha_O}{\varepsilon}\left(L\beta t\frac{\log(\beta t\alpha_O/\varepsilon)}{\log\log(\beta t\alpha_O/\varepsilon)}+T_O\right)\right)$	$O(1)$
Qubitization [43]	Incoherent	$O(\log L)$	$O\left(L\left(\beta t+\log(\\|O\\|/\varepsilon)\right)\right)$	$O\left(\frac{\\|O\\|^2}{\varepsilon^2}\right)$
	Coherent	$O(a_O+\log L)$	$O\left(\frac{\alpha_O}{\varepsilon}\left(L\beta t+L\log(\alpha_O/\varepsilon)\right)+\frac{\alpha_O T_O}{\varepsilon}\right)$	$O(1)$
This work	–	1	$O\left(\beta^2t^2\frac{\log(\beta t\\|O\\|/\varepsilon)}{\log\log(\beta t\\|O\\|/\varepsilon)}\right)$	$O\left(\frac{\\|O\\|^2}{\varepsilon^2}\right)$

Table 2: Given a Hamiltonian $H=\sum_{j=1}^{L}\lambda_j P_j$, where P_j's are strings of Pauli matrices and $\beta=\sum_j|\lambda_j|$, we compare the cost of Hamiltonian simulation by *Single-Ancilla LCU* with other methods. In particular, we compare the number of ancilla qubits, gate depth per coherent run, and the number of classical repetitions needed to output an ε-additive estimate of $\mathrm{Tr}[O\rho_t]$, where $\rho_t=e^{-iHt}\rho_0 e^{iHt}$, for some initial state ρ_0 and any measurable observable O. The overall gate depth is the product of the last two columns of the table. Here the *incoherent* approach refers to the method of directly measuring O, having prepared ρ_t. For the *coherent* approach, quantum amplitude estimation is used to estimate $\mathrm{Tr}[O\rho_t]$, given access to an $(\alpha_O, a_O, 0)$ block encoding of O. Finally, Hamiltonian simulation by qubitization [43], assumes a block encoding access to H. In our case, the block encoding can be constructed by an LCU approach requiring $O(\log L)$ ancilla qubits and gate depth $O(L)$.

measured coherently using quantum amplitude estimation, not only does the total circuit depth increase, but so does the overall gate depth, given by $O(\alpha_O^2 L\beta^2t^2\varepsilon^{-2}+\alpha_O T_O\varepsilon^{-1})$. Furthermore, the block encoding of O and the amplitude estimation procedure together require $O(a_O)$ ancilla qubits. Note that this is the worst case complexity of the First order Trotter method. It has been shown that in certain specific instances, when specific bounds on the commutators of the local terms of the Hamiltonian are known, the complexity scales better [81, 48].

The complexity of our procedure retains some similarity to randomized quantum simulation schemes such as qDRIFT [60] in that it does not depend on L. The gate depth for implementing Hamiltonian simulation via qDRIFT is $O(\beta^2t^2\|O\|/\varepsilon)$. Thus, our approach has an exponentially better dependence on precision and $\|O\|$. The incoherent approach to estimate $\mathrm{Tr}[O\rho_t]$ would need $O(\|O\|^2/\varepsilon^2)$ repetitions, which is the same as our method. Even with the coherent approach (using quantum amplitude estimation), the gate depth is $O(\alpha_O^2\beta^2t^2/\varepsilon^2+\alpha_O T_O/\varepsilon)$, which is higher than the overall gate depth of our method.

Just as the First order Trotter method, qDRIFT requires no ancilla qubits in order to prepare ρ_t and then incoherently estimate $\mathrm{Tr}[O\rho_t]$. However, the coherent approach requires $O(a_O)$ ancilla qubits.

For higher order Trotter methods [48], the $2k$-th order Trotter procedure has a gate depth per coherent run, given by

$$O\left(L(\beta t)^{1+\frac{1}{2k}}\left(\frac{\|O\|}{\varepsilon}\right)^{\frac{1}{2k}}\right),$$

in order to prepare ρ_t with $\varepsilon/\|O\|$ accuracy. As compared to the Hamiltonian simulation procedure based on *Single-Ancilla LCU*, the circuit depth of higher order Trotter is sub-exponentially worse in terms of $1/\varepsilon$, but the dependence on t is better. Just as in the previous cases, obtaining an ε-accurate estimate of $\mathrm{Tr}[O\rho_t]$ by measuring O, requires $O(\|O\|^2/\varepsilon^2)$ classical repetitions, each costing T_H. Interestingly, the overall gate depth of the coherent technique using quantum amplitude estimation, given by

$$O\left(L(\beta t)^{1+\frac{1}{2k}}\left(\frac{\alpha_O}{\varepsilon}\right)^{1+\frac{1}{2k}}+\frac{\alpha_O}{\varepsilon}T_O\right),$$

has a better scaling as compared to *Single Ancilla LCU* (in terms of $1/\varepsilon$), at the cost of an increased gate depth and additional $O(a_O)$ ancilla qubits. As with first order Trotter, we have noted the worst case complexity. It has been shown that higher order Trotter methods also perform better in practice for specific local Hamiltonians for which sums of nested commutators of the local terms of H are small [48].

The truncated Taylor series method by Berry et al. makes use of the *Standard LCU* procedure [24]. This requires $O(\log(L)\log(||O||t/\varepsilon)/\log\log(||O||t/\varepsilon))$ ancilla qubits and sophisticated multi-qubit controlled operations. Moreover it needs to implement involved subroutines such as oblivious amplitude amplification for each segment, which cannot be avoided. However, the gate depth per run of this procedure is linear in t and β, which is quadratically better than our method. However, implementing the LCU has an overhead of L in terms of the gate depth, and thus, in the regime where $\beta \ll L$, there is a range of values of t for which our method requires lower gate cost per coherent run. This, despite the fact that our algorithm requires only a single ancilla qubit and does not require operations that are controlled over multiple qubits. The coherent approach to estimate $\mathrm{Tr}[O\rho_t]$ has a better dependence on β, t as well as $1/\varepsilon$ at the cost of an exponentially increased gate depth per coherent run (in terms of $1/\varepsilon$), and $O(a_O + \log(L)\log(\alpha_O t/\varepsilon)/\log\log(\alpha_O t/\varepsilon))$ ancilla qubits.

The state-of-the-art quantum simulation technique makes use of quantum singular value transformation and requires a coherent access to H via a block encoding [43]. The complexity of this procedure is optimal (linear in $t + \log(1/\varepsilon)$) in terms of the number of queries made to any such block encoding. Moreover it requires only one more ancilla qubit as compared to our method. However, when H is a linear combination of unitaries, constructing the block encoding itself requires $O(\log L)$ ancilla qubits, implementing multi-qubit controlled unitaries, and gate depth $O(L)$. Then the gate depth of this procedure to prepare ρ_t with $\varepsilon/\|O\|$ accuracy is given by $O(L\,(\beta t + \log(\,\|O\|\,/\varepsilon)))$. This has optimal scaling in t and $1/\varepsilon$, but has an overhead in terms of L, which can be quite large for several Hamiltonians of interest. Thus, for Hamiltonians satisfying $\beta \ll L$, there is a range of t for which our method provides an advantage in terms of the gate depth per coherent run, even with respect to state-of-the-art algorithms.

In Table 2, we compare the cost of the Hamiltonian simulation procedure using *Single-Ancilla LCU* with other methods.

5 Applications to Ground state preparation and property estimation

The problem of preparing (or extracting useful information from) the ground states of a Hamiltonian finds widespread interest across physics and computer science. Generally, this problem is known to be computationally hard, even for a quantum computer [82]. However, owing to its wide applicability, novel improved quantum algorithms for ground state preparation (GSP), and ground state property estimation are of extreme importance and interest. We will apply the *Analog LCU* and the *Single-Ancilla LCU* approaches, introduced in Sec. 3, to tackle these problems.

This section is organized as follows. We begin by formally describing the ground state preparation (GSP) problem. Next, we discuss an analog quantum algorithm for GSP using the *Analog LCU* framework. Then, we use the *Single-Ancilla LCU* technique to estimate the expectation value of observables with respect to the ground states of a Hamiltonian. We start by describing the ground state preparation problem.

The ground state Preparation problem: The set up of the problem is similar to prior works [35, 36, 33]. Suppose we have a Hamiltonian H with ground state $|v_0\rangle$ and ground energy λ_0, and assume that we are given a lower bound on the gap between the ground state and the first excited state of H (spectral gap), i.e. we have knowledge of Δ such that $|\lambda_1 - \lambda_0| \geq \Delta$.
For clarity of exposition, we assume that the ground space of H is non-degenerate. If this is not the case, e.g. if the degeneracy of the ground space is d and is spanned by mutually orthonormal eigenstates $\{|v_0^{(\ell)}\rangle\}_{l=1}^{d}$, then we will be preparing a quantum state $|v_0\rangle$ which is a projection onto the ground space given by

$$|v_0\rangle = \frac{1}{\sqrt{\sum_{\ell=1}^{d} |c_0^{(\ell)}|^2}} \sum_{\ell=1}^{d} c_0^{(\ell)} |v_0^{(\ell)}\rangle \,.$$

In addition, suppose we have access to some initial state $|\psi_0\rangle$ and a lower bound on the overlap $|\langle\psi_0|v_0\rangle| = c_0 \geq \eta$.
Furthermore, for some desired accuracy $\varepsilon \in (0, 1)$, we will assume that we know the value of the ground energy to some precision parameter ε_g such that $\varepsilon_g = \mathcal{O}\left(\Delta/\sqrt{\log \frac{1}{\eta\varepsilon}}\right)$. That is, we know some E_0 such that

$$|\lambda_0 - E_0| \leq \varepsilon_g. \tag{57}$$

By implementing $H - (E_0 - \varepsilon_g)I$, we ensure that $0 \leq \lambda_0 \leq 2\varepsilon_g$. This transformation also ensures that the lower bound for the spectral gap of H remains Δ. Without loss of generality, we assume that the spectrum of H is in the interval $[0, 1]$. Otherwise, if $\|H\|$ is an upper bound on the maximum eigenvalue of H, we would consider the Hamiltonian $H/\|H\|$, which has a spectral gap of at least $\Delta/\|H\|$. Thus, the complexities of our algorithm would get rescaled by this $\|H\|$ factor.

5.1 Applying Analog LCU: A continuous-time quantum algorithm for ground state preparation

In this section, we will use the *Analog LCU* framework to develop an analog quantum algorithm for the GSP problem. This algorithm was described in the Supplemental Ma-

terial of [33]. Here, we place it in the broader context of the *Analog LCU* framework. Moreover, these results will be used to develop a quantum algorithm for estimating expectation values of observables with respect to the ground states of Hamiltonians in the next Section.

Consider some quantum system in state $|\psi_0\rangle$ coupled to an ancillary system in a Gaussian state

$$|\psi_g\rangle = \int_{-\infty}^{+\infty} \frac{dz}{(2\pi)^{1/4}} e^{-z^2/4} |z\rangle . \tag{58}$$

The Gaussian state is typically easy to prepare in this setting. This state can be seen as the ground state of a one-dimensional quantum harmonic oscillator. The coupling is done via interaction Hamiltonian $H' = H \otimes \hat{z}$, where $\hat{z}$ corresponds to the position (or momentum) operator. Evolving $|\psi_0\rangle |\psi_g\rangle$, under H' for a time t results in the state

$$\begin{aligned} |\eta_t\rangle &= e^{-itH'} |\psi_0\rangle |\psi_g\rangle \\ &= \int_{-\infty}^{+\infty} \frac{dz}{(2\pi)^{1/4}} e^{-z^2/4} e^{-itHz} |\psi_0\rangle |z\rangle \\ &= \int_{-\infty}^{+\infty} \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} e^{-itHz} |\psi_0\rangle |\psi_g\rangle + |\Phi\rangle^{\perp} , \end{aligned} \tag{59}$$

where $|\Phi\rangle^{\perp}$ is a quantum state with the ancillary system being orthogonal to $|\psi_g\rangle$. Now the Fourier transform of a Gaussian is a Gaussian, i.e. we have for any $x \in \mathbb{R}$,

$$e^{-y^2/2} = \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} \, e^{-z^2/2} e^{-iyz}. \tag{60}$$

So using this we obtain

$$|\eta_t\rangle = e^{-t^2H^2/2} |\psi_0\rangle |\psi_g\rangle + |\Phi\rangle^{\perp} . \tag{61}$$

By post-selecting on obtaining $|\psi_g\rangle$ in the second register, we are able to prepare a quantum state proportional to $e^{-t^2H^2/2} |\psi_0\rangle$ in the first register. In [33] it was shown that this state is close to the ground state $|v_0\rangle$, with η^2 probability, provided $t = \widetilde{O}(\Delta^{-1})$. Now we formally state the ground state preparation algorithm of [33] and its complexity, via the following lemma.

Lemma 13 ([33])**.** *Suppose $\varepsilon \in (0,1)$ and $\eta \in (0, 1/\sqrt{2}]$. Furthermore, suppose we have a Hamiltonian H with ground state $|v_0\rangle$ with Δ being a lower bound on the spectral gap. Also, the ground state energy of H is known up to a precision $\varepsilon_g \in \mathcal{O}\left(\Delta/\sqrt{\log \frac{1}{\eta\varepsilon}}\right)$. Then, given an initial state $|\psi_0\rangle$ satisfying $|\langle\psi_0|v_0\rangle| \geq \eta$, we output, with probability $\mathcal{O}(\eta^2)$, a state $|\phi\rangle$ such that $\||\phi\rangle - |v_0\rangle\| \leq \varepsilon$ by evolving the Hamiltonian $H' = H \otimes \hat{z}$ for time*

$$T = \mathcal{O}\left(\frac{1}{\Delta}\sqrt{\log\left(\frac{1}{\eta\varepsilon}\right)}\right).$$

While the detailed proof can be found in [33], the idea is to evolving the overall system according to the interaction Hamiltonian H' for a time $\sqrt{2t)}$ to prepare the quantum state

$$|\eta_t\rangle = e^{-tH^2} |\psi_0\rangle |\psi_g\rangle + |\Phi\rangle^{\perp} , .$$

The next observation is that by choosing any

$$t > \frac{1}{2\Delta^2} \log\left(\frac{1-\eta^2}{\eta^2\varepsilon^2}\right), \tag{62}$$

we ensure that with probability at least η^2, we prepare in the first register, the state

$$|\phi\rangle = \frac{e^{-tH^2}|\psi_0\rangle}{\left\|e^{-tH^2}|\psi_0\rangle\right\|},$$

which is ε-close to the ground state $|v_0\rangle$. For this choice of t, the time evolution of the interaction Hamiltonian should scale as

$$T = \sqrt{2t} = \mathcal{O}\left(\frac{1}{\Delta}\sqrt{\log\left(\frac{1}{\eta\varepsilon}\right)}\right). \tag{63}$$

Overall, this physically motivated quantum algorithm is significantly simpler than implementing standard LCU in the circuit model. Moreover, hybrid qubit-qumode systems are currently being engineered in a number of quantum technological platforms. In the future, we intend to provide an experimental proposal to implement *Analog LCU* on experimental platforms such as ion traps or superconducting systems.

Note that evolving the system according to the interaction Hamiltonian H', we obtain the ground state $|v_0\rangle$ in the first register, with probability η^2 (postselected on measuring $|\psi_g\rangle$ in the ancilla register). Thus, $1/\eta^2$ repetitions of this procedure suffices to prepare $|v_0\rangle$, resulting in a total cost of $T = O\left(\Delta^{-1}\eta^{-2}\sqrt{\log(\eta^{-1}\varepsilon^{-1})}\right)$. Alternatively, $O(1/\eta)$-rounds of quantum amplitude amplification can also be used to prepare $|v_0\rangle$, which would bring down the overall cost by a factor of $1/\eta$. However, amplitude amplification is a discrete procedure with no continuous-time analogue. Indeed in the Appendix (Sec. D), we develop a quantum algorithm for ground state preparation in the circuit model for fully fault tolerant quantum computers whose complexity matches that of the state-of-the-art quantum algorithms for this problem. Therein, we implement a polynomial that approximates the function e^{-tx^2}, using QSVT.

Next, we describe how the *Single-Ancilla LCU* technique can be used to develop a randomized quantum algorithm for ground state property estimation.

5.2 Applying Single-Ancilla LCU: Ground state property estimation

In this section, we assume that we can access the Hamiltonian H through the time evolution operator $U_t = \exp[-itH]$. Furthermore, given access to U_t, we can perform the time evolution controlled on a single ancilla qubit. This is referred to as the Hamiltonian evolution (HE) model as has been used in prior works for ground energy estimation using early fault-tolerant quantum computers [12, 11, 13]. Much like these works, we calculate: (a) the maximal time of evolution of H (controlled by a single ancilla qubit) required in each coherent run, given by $\tau_{\max}$, and (b) the total number of repetitions of the circuit T. The total evolution time is then $O(\tau_{\max} \cdot T)$.

Given any Hamiltonian H with ground state $|v_0\rangle$, we will use Algorithm 2 and Theorem 10 to estimate $\langle v_0|O|v_0\rangle$ to ε-accuracy, for any measurable observable O. The cost of each coherent run of our algorithm will be measured in terms of the maximal time for each H is evolved ($\tau_{\max}$). Additionally, we shall also estimate T, the number of classical repetitions required for our procedure, and the number of ancilla qubits. As mentioned previously in

Sec. 3.2, our method requires only a single ancilla qubit, and we compare the complexity of this procedure (in the HE model) with other methods.

Given any initial state $\rho_0 = |\psi_0\rangle\langle\psi_0|$, with overlap of at least η with the ground state, prepared in cost τ_{ψ_0}, we use Algorithm 2 to obtain an accurate estimate $\mathrm{Tr}[O\rho]$, where

$$\rho = \frac{e^{-tH^2}\rho_0 e^{-tH^2}}{\mathrm{Tr}[e^{-tH^2}\rho_0 e^{-tH^2}]},$$

for some

$$t \in O\left(\frac{1}{\Delta^2}\log\frac{1}{\eta\varepsilon}\right).$$

Using Lemma 13, we know that an accurate enough estimate of $\mathrm{Tr}[O\rho]$ is also an accurate estimate of $\langle v_0|O|v_0\rangle$.

For this, we consider a discretized version of this LCU decomposition, i.e. we approximate e^{-tH^2} as a linear combination of roughly $\sqrt{t}$ terms. This decomposition has already shown up in prior works [37, 76]. We formally state this via the following Lemma.

Lemma 14 (LCU decomposition of e^{-tH^2} [37])**.** *Let $0 < \gamma < 1$ and consider a Hamiltonian H of unit spectral norm. Furthermore, for any $t > 1$, let us define*

$$X_M = \sum_{j=-M}^{M} c_j e^{-ij\delta_t\sqrt{2t}H},$$

where $M = \left\lceil \sqrt{2}\left(\sqrt{t} + \sqrt{\log(5/\gamma)}\right)\sqrt{\log(4/\gamma)}\right\rceil$, $\delta_t = \left(\sqrt{2t} + \sqrt{2\log(5/\gamma)}\right)^{-1}$ *and,*

$$c_j = \frac{\delta_t}{\sqrt{2\pi}} e^{-j^2\delta_t^2/2}.$$

Then,

$$\left\| X_M - e^{-tH^2} \right\| \leq \gamma.$$

We will use this LCU decomposition for our randomized quantum algorithm. First note that the ℓ_1-norm of the LCU coefficients in Lemma 14, can be upper bounded by a constant. In fact,

$$||c||_1 = \sum_{j=-M}^{M} |c_j| \tag{64}$$

$$\leq |c_0| + 2\sum_{j=1}^{\infty} \frac{\delta_t}{\sqrt{2\pi}} e^{-j^2\delta_t^2/2} \tag{65}$$

$$\leq |c_0| + 2\int_0^{\infty} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = 1 + |c_0| \leq 1 + \delta_t = O(1). \tag{66}$$

Second, we will shortly prove that it suffices to consider

$$\gamma = \frac{\varepsilon\eta^2}{30\,\|O\|}.$$

Now, we are in a position to use Algorithm 2. Recall that, Algorithm 2 estimates the expectation value and the norm, by making separate calls to Algorithm 1.

Let us first estimate the cost of each run of our algorithm in the Hamiltonian Evolution model. Each iteration of our randomized quantum algorithm requires implementing (controlled and anti-controlled) versions of unitaries V_1, V_2 which are i.i.d samples from the ensemble $\{U_j, c_j/\|c\|_1\}$, where, from Lemma 14, each $U_j = e^{-ij\delta_t\sqrt{2t}H}$. For this case, it suffices to obtain two integers j_1, j_2 according to $\{c_j/\|c\|_1\}$ and then implement $V_1 = e^{-ij_1\delta_t\sqrt{2t}H}$ and $V_2 = e^{-ij_2\delta_t\sqrt{2t}H}$, respectively. So, the cost of each coherent run can be upper bounded by $2\tau_{\max} + \tau_{\psi_0}$, where $\tau_{\max}$ is the maximum time of evolution for H, which can be obtained from Lemma 14 as

$$\tau_{\max} = M\delta_t\sqrt{2t} = O\left(\sqrt{t\log\left(1/\gamma\right)}\right) = O\left(\frac{1}{\Delta}\log\left(\frac{\|O\|}{\eta\varepsilon}\right)\right). \tag{67}$$

We prove the correctness of our algorithm, as well as the total number of repetitions T, via the following theorem:

Theorem 15. *Let $\varepsilon, \delta, \gamma \in (0,1)$ and $\eta \in (0, 1/\sqrt{2}]$. Suppose $H = \sum_{k=1}^{L}\lambda_k P_k$ is a Hermitian matrix, with ground state $|v_0\rangle$ and let $|\psi_0\rangle$ be some initial state, prepared in cost τ_{ψ_0}, such that $|\langle v_0|\psi_0\rangle| = \eta$. Let O be some observable. Furthermore, for*

$$t = O\left(\frac{1}{\Delta}\log\left(\frac{\|O\|}{\eta\varepsilon}\right)\right),$$

and,

$$\gamma = \frac{\varepsilon\eta^2}{30\ \|O\|},$$

suppose,

$$\left\|e^{-tH^2} - X_M\right\| \leq \gamma.$$

Then for

$$T = O\left(\frac{\|O\|^2\ln(1/\delta)}{\varepsilon^2\eta^4}\right),$$

Algorithm 2 outputs, with probability at least $(1-\delta)^2$, parameters $\mu, \tilde{\ell}$ such that

$$\left|\frac{\mu}{\tilde{\ell}} - \langle v_0|O|v_0\rangle\right| \leq \varepsilon,$$

using T repetitions of the quantum circuit in Fig. 2, and only one ancilla qubit. The maximal time of evolution of H is at most

$$\tau_{max} = O\left(\frac{1}{\Delta}\log\left(\frac{\|O\|}{\varepsilon\eta}\right)\right).$$

Proof. Define

$$|\phi\rangle = \frac{e^{-tH^2}|\psi_0\rangle}{\left\|e^{-tH^2}\ |\psi_0\rangle\right\|}.$$

First observe that $\langle\psi_0|e^{-2tH^2}|\psi_0\rangle \geq \ell_* = \Omega(\eta^2)$. Then, for the chosen value of γ, from Theorem 10, the first call to Algorithm 1 outputs an estimate μ such that

$$\left|\mu - \mathrm{Tr}[O\ e^{-tH^2}\rho_0 e^{-tH^2}]\right| \leq \frac{\varepsilon\eta^2}{5}.$$

The second call to Algorithm 1 outputs $\tilde{\ell}$ such that

$$\left|\tilde{\ell} - \ell^2\right| \leq \frac{\varepsilon\eta^2}{5\|O\|}.$$

Then Algorithm 2 outputs $\mu/\tilde{\ell}$, which from Theorem A1 (substituting $a = b = 5$), guarantees

$$\left|\frac{\mu}{\tilde{\ell}} - \langle\phi|O|\phi\rangle\right| \leq \varepsilon/2.$$

Then, we have:

$$\left|\langle v_0|O|v_0\rangle - \langle\phi|O|\phi\rangle\right| \leq \|O\|_\infty \left\||\phi\rangle\langle\phi| - |v_0\rangle\langle v_0|\right\|_1 \quad [\text{ Using Lemma 6 with } p = \infty,\ q = 1] \tag{68}$$

$$\leq 2\|O\|\sqrt{1 - |\langle v_0|\phi\rangle|}. \tag{69}$$

Now we have

$$\left|\langle\phi|v_0\rangle\right| \geq 1 - \frac{1}{2}\left\||\phi\rangle - |v_0\rangle\right\|^2 \tag{70}$$

$$\geq 1 - \frac{\eta^2}{2(1-\eta^2)}e^{-2t\Delta^2}. \tag{71}$$

So, by choosing some

$$t > \frac{1}{2\Delta^2}\log\left(\frac{8\|O\|^2(1-\eta^2)}{\varepsilon^2\eta^2}\right),$$

we ensure that

$$\left|\langle\phi|v_0\rangle\right| \geq 1 - \frac{\varepsilon^2}{16\|O\|^2}. \tag{72}$$

Substituting this back in Eq. (69) gives us,

$$\left|\langle v_0|O|v_0\rangle - \langle\phi|O|\phi\rangle\right| \leq \varepsilon/2,$$

as desired. By triangle inequality, we obtain,

$$\left|\frac{\mu}{\tilde{\ell}} - \langle v_0|O|v_0\rangle\right| \leq \left|\frac{\mu}{\tilde{\ell}} - \langle\phi|O|\phi\rangle\right| + \left|\langle v_0|O|v_0\rangle - \langle\phi|O|\phi\rangle\right| \leq \varepsilon. \tag{73}$$

The maximal time of evolution,

$$\tau_{\max} = O(\sqrt{t}) = O\left(\frac{1}{\Delta}\log\left(\frac{\|O\|}{\varepsilon\eta}\right)\right).$$

The total number of repetitions of the underlying quantum circuit can be obtained by simply substituting in the value of T (in Theorem 2), $\ell_* = \eta^2$, and $\|c\|_1 = O(1)$ to obtain

$$T = O\left(\frac{\|O\|^2\ln(1/\delta)}{\varepsilon^2\eta^4}\right), \tag{74}$$

which completes the proof. □

Thus, the total evolution time is $O(\tau_{\max}.T) = \widetilde{O}(\Delta^{-1}\eta^{-4}\|O\|^2/\varepsilon^2)$.

Comparison with prior works: Here, we compare our method with prior works on ground state preparation: (a) algorithms that make use of Standard LCU, and also, (b) state-of-the-art quantum algorithms for ground state preparation by using QSVT, and (c) early fault-tolerant quantum algorithms for ground state preparation and ground energy estimation. While this is discussed at length in the subsequent paragraphs, the comparison has been summarized in Table 3.

Let us consider ground state preparation algorithms that make use of the *Standard LCU* procedure. We will estimate the performance of these algorithms in the Hamiltonian evolution (HE) model. It is important to note that for *Standard LCU*, we will be dealing with a multi-qubit controlled Hamiltonian evolution oracle: we calculate the maximal time evolution of the multi-qubit controlled unitary c_m-U_t ($U_t = \exp[-itH]$ controlled over m-ancilla qubits), in addition to the number of classical repetitions (overall complexity is measured in terms of the total evolution time which is the product of these two quantities), and the number of ancilla qubits needed. Also, as outlined in Table 1, any generic LCU procedure for preparing $|v_0\rangle$ has three ways which it can estimate $\langle v_0|O|v_0\rangle$. We will consider each of them.

- Ge et al. [34] use the standard LCU procedure to prepare $|v_0\rangle$. Implementing the LCU requires $O(\log(\log(||O||\eta^{-1}\varepsilon^{-1})/\Delta))$ qubits as ancillae, and multi-qubit controlled operations, which prepares $|v_0\rangle$ with $\varepsilon/\|O\|$-accuracy. It is possible to use the LCU algorithm to first prepare $|v_0\rangle$ with a constant probability using quantum amplitude amplification, and then measure O, using $O(\ ||O||^2/\varepsilon^2)$ classical repetitions. The maximal time evolution of H per coherent run is $\widetilde{O}(\eta^{-1}\Delta^{-1})$, which is higher than our method. However, the number of classical repetitions needed (and also the total evolution time) is lower than *Single-Ancilla LCU*. Overall, the advantage of our method is that requires only a single ancilla qubit, no multi-qubit controlled gates, and lower maximal time of evolution of H per coherent run.

- The number of classical repetitions can be reduced to a constant if $\langle v_0|O|v_0\rangle$ is estimated using quantum amplitude estimation. This technique also reduces the dependence on precision to $1/\varepsilon$ instead of $1/\varepsilon^2$. In this case, we consider an $(\alpha_O, a_O, 0)$-block encoding of the observable O. The maximal time of evolution per coherent run is $\widetilde{O}\left(\alpha_O\varepsilon^{-1}\eta^{-1}\Delta^{-1}\right)$, which is also the total evolution time. Thus, the total evolution time of this approach is lower than our method. However, the maximal evolution time of H per coherent run is exponentially higher (in terms of $1/\varepsilon$). Furthermore, this approach requires $O(a_O + \log(\log(||O||\eta^{-1}\varepsilon^{-1})/\Delta))$ ancilla qubits, while our method requires only a single ancilla qubit and no multi-qubit controlled operations.

- Finally, if simply standard LCU is used without quantum amplitude amplification or estimation, the maximal time of evolution matches our method while the number of classical repetitions (and hence the total evolution time) is quadratically lower in terms of $1/\eta$. However, as with the previously mentioned approaches, *Standard LCU* requires: $O(\log(\log(\eta^{-1}\varepsilon^{-1})/\Delta))$ ancilla qubits, and implementing multi-qubit controlled gates.

The quantum algorithm for ground state preparation by Lin and Tong [64] uses the

framework of Quantum Singular Value Transformation. Consequently, it does not require the Hamiltonian evolution operator, and hence cannot be directly compared with our approach. Its complexity however is measured in terms of the number of queries to a block encoding of H. Given access to an $(\alpha_H, a, 0)$-block encoding of H, their algorithm requires $O(\alpha_H \Delta^{-1} \eta^{-1} \log(1/\varepsilon))$ queries and only $O(1)$ ancilla qubits to prepare $|v_0\rangle$. However, constructing a block encoding of H can be resource demanding and may lead to the use of multiple ancilla qubits and multi-qubit controlled gates. For instance if H is a linear combination of L Pauli terms with β being the total weight of the coefficients, the block encoding of H requires $O(\log L)$ ancilla qubits and gate depth $O(L)$ (also, $\alpha_H = \beta$). Thus, this method is not suitable in the early fault-tolerant regime. Nevertheless, much like the *Standard LCU* approach, this algorithm can also be adapted to estimate $\langle v_0|O|v_0\rangle$ in three ways (as shown in Table 3). The overall query complexity is always lower than our total evolution time. However, besides needing more ancilla qubits, the number of queries per coherent run of the Lin and Tong algorithm can be higher (as $\alpha_H \gg 1$) than the maximal time of evolution H in the *Single-Ancilla LCU* method.

A number of quantum algorithms have been developed for estimating the ground energy of Hamiltonians in the Hamiltonian evolution model, tailored to early fault-tolerant quantum computers. Most of these algorithms also make use of a single ancilla qubit. For instance, in Ref. [12], Lin and Tong use the Hadamard test (and classical post processing) to achieve the so-called Heisenberg scaling ($1/\varepsilon$ - dependence) for estimating the ground energy to ε-additive accuracy. Their algorithm requires a maximal evolution time $\widetilde{O}(\varepsilon^{-1}\text{polylog}\,(1/\eta))$, while the total evolution time scales as $\widetilde{O}(\varepsilon^{-1}\eta^{-2})$. The algorithm of Wang et al. [13] on the other hand, uses the same circuit but a different post-processing methodology through which they are able to exponentially improve the dependence on precision with respect to the Hamiltonian evolution time per coherent run. Their result requires a maximal evolution time of $O(\Delta^{-1}\text{polylog}\left(\varepsilon^{-1}\eta^{-1}\Delta\right))$, and a total evolution time scaling as $O(\eta^{-2}\varepsilon^{-2}\Delta\text{polylog}\left(\eta^{-1}\varepsilon^{-1}\Delta\right))$.

The *Single-Ancilla LCU* method for ground state property estimation assumes that the ground energy of H is known to $O(\Delta/\sqrt{\log(\eta^{-1}\varepsilon^{-1})})$ precision. Thus, one can make use of the algorithm of Wang et al. [13] to first estimate the ground energy and then run our algorithm, without adding any asymptotic overhead either in terms of the maximal time evolution per coherent run or the total evolution time.

Dong, Lin and Tong [10] provide ground energy estimation and ground state preparation algorithms for early fault-tolerant quantum computers. Their access model is slightly different: it measures the complexity in terms of the number of queries to $U = e^{-iH}$ (and $U^\dagger$). The underlying approach is reminiscent of quantum signal processing: interleaved applications of c-U and c-$U^\dagger$ along with single qubit rotations (with adjustable phases). They refer to this as Quantum Eigenvalue Transformation of Unitaries (QETU). Thus, as compared to other early fault-tolerant approaches, this approach has an overhead in terms of the number of single qubit gates needed, which scales linearly with the number of queries made to U and $U^\dagger$. The authors provide two algorithms for ground energy estimation: (a) The first one requires $\widetilde{O}(\varepsilon^{-1}\log(1/\eta))$ queries to c-U and c-$U^\dagger$ per coherent run, while using only a single ancilla qubit. The total number of queries needed is $\widetilde{O}(\varepsilon^{-1}\eta^{-2})$. (b) The second one makes use of quantum amplitude amplification and binary amplitude estimation to improve the overall query complexity by a factor of $1/\eta$, but at the same time the maximal query-depth per run, also increases by this factor. Furthermore, this approach requires three ancilla qubits and hence, multi-qubit controlled operations.

Similarly, for ground state preparation there are two algorithms: the first is a near-optimal ground state preparation algorithm has a maximal query depth of $\widetilde{O}(\eta^{-1}\Delta^{-1})$,

Algorithm	Access	Variant	Ancilla	Cost per coherent run	Classical repetitions
Standard LCU [34]	HE (MQC)	QAA + classical repetitions	$O\left(\log\left(\log\left(\frac{\|O\|}{\eta\varepsilon}\right)/\Delta\right)\right)$	$\tilde{O}\left(\Delta^{-1}\eta^{-1}\right)$	$O\left(\frac{\|O\|^2}{\varepsilon^2}\right)$
		QAE	$O\left(a_O+\log\left(\log\left(\frac{\|O\|}{\eta\varepsilon}\right)/\Delta\right)\right)$	$\tilde{O}\left(\frac{\alpha_O}{\varepsilon\eta\Delta}\right)$	$O(1)$
		Without QAA or QAE	$O\left(\log\left(\log\left(\frac{\|O\|}{\eta\varepsilon}\right)/\Delta\right)\right)$	$\tilde{O}\left(\frac{1}{\Delta}\right)$	$O\left(\frac{\|O\|^2}{\varepsilon^2\eta^2}\right)$
QSVT [64]	BE (MQC)	QAA + classical repetitions	$O(a_H)$	$\tilde{O}\left(\frac{\alpha_H}{\Delta\eta}\right)$	$O\left(\frac{\|O\|^2}{\varepsilon^2}\right)$
		QAE	$O(a_O+a_H)$	$\tilde{O}\left(\frac{\alpha_O\alpha_H}{\varepsilon\eta\Delta}\right)$	$O(1)$
		Without QAA or QAE	$O\left(a_H\right)$	$\tilde{O}\left(\frac{\alpha_H}{\Delta}\right)$	$O\left(\frac{\|O\|^2}{\varepsilon^2\eta^2}\right)$
Early Fault-tolerant [10]	HE*	Without QAA	1	$\tilde{O}\left(\frac{1}{\Delta}\right)$	$O\left(\frac{\|O\|^2}{\varepsilon^2\eta^2}\right)$
	HE* (MQC)	With QAA	2	$\tilde{O}\left(\Delta^{-1}\eta^{-1}\right)$	$O\left(\frac{\|O\|^2}{\varepsilon^2}\right)$
Early Fault-tolerant [11]	HE+BE (MQC)	–	$O(a_O)$	$\tilde{O}\left(\Delta^{-1}\right)$	$O\left(\frac{\alpha_O^2}{\eta^4\varepsilon^2}\right)$
This work	HE	–	1	$\tilde{O}\left(\frac{1}{\Delta}\right)$	$O\left(\frac{\|O\|^2}{\varepsilon^2\eta^4}\right)$

Table 3: Consider any Hamiltonian H with ground state $|v_0\rangle$, an initial quantum state with overlap at least η with $|v_0\rangle$, and any measurable observable O. Furthermore, suppose we have knowledge of the ground energy with precision $\varepsilon_g = O(\Delta/\sqrt{\log(\eta^{-1}\varepsilon^{-1})})$. In this table, we compare with our technique, the cost of various quantum algorithms for estimating $\langle v_0|O|v_0\rangle$ to additive accuracy ε. For both Standard LCU [34] and our method (*Single-Ancilla LCU*), we consider that H can only be accessed through the time evolution operator $U_t = \exp[-iHt]$. This is the Hamiltonian Evolution (HE) model, where the cost of each coherent run is estimated in terms of the maximal time evolution of H. The number of repetitions refers to the total number of times the circuit is run. The total evolution time is then the product of these two quantities. The Standard LCU approach requires implementing multi-qubit controlled (MQC) time evolution operators, which is indicated in the Table along with the access model. In the quantum algorithm by Dong et al. [10], the authors estimate the complexity in terms of the number of queries made to the unitary $U = e^{-iH}$. The algorithm involves implementing interleaved U and $U^\dagger$, controlled over a single qubit ancilla, which implements single qubit phase rotations. This is slightly different from the HE model considered earlier, and hence is denoted as HE* in the Table. Furthermore, the number of queries made to this operator in one run of the underlying circuit determines the cost per coherent run. The QSVT-based algorithm by Lin and Tong [64] considers the block encoding model (denoted by BE in the Table). More precisely, the cost per coherent run is estimated in terms of the number of queries made to an $(\alpha_H, a_H, 0)$-block encoding of H. The product of the number of queries per run and the total number of classical repetitions is the total number of queries to the block encoding of H. In order to estimate the desired expectation value directly via quantum amplitude estimation (QAE), we assume that O is accessed via an $(\alpha_O, a_O, 0)$-block encoding. Also the algorithm of Zhang et al. [11] assumes this block encoding for estimating $\langle v_0|O|v_0\rangle$.

which is also the total number of queries, while requiring only two ancilla qubits. Unlike Ref. [64], this does not assume a block encoding access to H, but makes us of quantum amplitude amplification, which is hard to implement in the early fault-tolerant regime. This algorithm can be used to estimate $\langle v_0|O|v_0\rangle$, requiring additional queries scaling as $\|O\|^2/\varepsilon^2$. Thus, this has a lower overall evolution time as compared to our method, at the cost of requiring only one additional ancilla qubit. However, a higher maximal evolution time per coherent run is needed as compared to *Single-Ancilla LCU*.

The second algorithm requires shorter query depth per coherent run and uses only a single ancilla qubit. The algorithm requires a maximal query depth of $\widetilde{O}(1/\Delta)$, to prepare the ground state with probability η^2. In order to estimate $\langle v_0|O|v_0\rangle$, this algorithm will require $O(\|O\|^2 \eta^{-2}/\varepsilon^2)$ classical runs of their circuit, which is quadratically better than our approach.

However, besides the single qubit gate overhead, there is another drawback of this algorithm which concerns translating query depth to actual gate depth. For this, a specific Hamiltonian simulation procedure must be chosen to implement $U = \exp[-iH]$. State-of-the-art techniques require access to a block encoding of H which leads to an increase in the number of ancilla qubits, as well as the overall gate depth. Thus, Trotter-based methods or Hamiltonian simulation by *Single-Ancilla LCU* need to be leveraged in order to keep the overall ancilla qubit count to one. However, the later method cannot be efficiently incorporated into the framework of Dong et al. This is primarily because our Hamiltonian simulation procedure implements some S such that $S/\|c\|_1 \approx e^{-iH}$, where $\|c\|_1 = O(1)$. Since the subnormalization factor is not unity, many queries to U and $U^\dagger$ would lead to an exponential overhead (d queries lead to an overhead of $\|c\|_1^d$) in the final complexity. Consequently, those techniques can be employed which implements U without any (even constant) sub-normalization factor while still keeping the overall ancilla qubits to one. Thus, only Trotter-based Hamiltonian simulation techniques can be incorporated which has a sub-exponentially worse dependence on the gate depth per coherent run (in terms of $1/\varepsilon$). On the other hand, the gate depth per coherent run of our ground state property estimation algorithm which uses the Hamiltonian simulation algorithm by *Single-Ancilla LCU* to implement the time evolution operator (proven in Appendix Sec. C.2) has a $O(\text{polylog}\,(1/\varepsilon))$ dependence.

Finally, the recent early fault-tolerant quantum algorithm by Zhang et al. [11] also estimates properties of ground states of Hamiltonians, i.e. $\langle v_0|O|v_0\rangle$. Their algorithm assumes (i) access to H in the Hamiltonian evolution model, and (ii) for any generic observable O, an $(\alpha_O, a_O, 0)$-block encoding of O. While the maximal time evolution of each run is $\tilde{O}(1/\Delta)$ (same as our method), the number of classical repetitions needed is $O(\alpha_O^2 \eta^{-4} \varepsilon^{-2})$, can be higher depending on the specific block encoding. Moreover, unlike our algorithm, the technique of Zhang et al. [11] cannot estimate $\langle v_0|O|v_0\rangle$ using a single ancilla qubit. This is because constructing the block encoding of O requires several ancilla qubits and multi-qubit controlled operations.

We reiterate that $\tau_{\max}$ is different from the actual circuit depth, which depends on how the Hamiltonian evolution unitary is implemented. Recall that in the early fault-tolerant regime, we are limited by a small ancillary qubit space and the inability to perform multi-qubit controlled operations. This restricts the choice of the underlying simulation algorithm. If H is a linear combination of strings of Pauli operators, i.e. $H = \sum_{k=1}^{L} \lambda_k P_k$, (first and higher order) Trotter methods, as well as the *Single Ancilla LCU*- based Hamiltonian simulation algorithm are suitable options. This is because $c\text{-}U_t$ can be implemented using only a single ancilla qubit and no multi-qubit controlled gates. However, these methods require a circuit depth which is super-linear in τ, which would in turn increase the circuit depth of our algorithm, and also the overall cost. On the other hand, state-of-the-art algorithms such as qubitization have an optimal dependence on t (measured in terms of the number of queries made to a block encoding of the Hamiltonian H), which would mean that the cost to implement $c\text{-}U_\tau$ would be $\tilde{O}(\tau)$. Moreover, the procedure itself requires only two ancilla qubits. However, constructing a block encoding of H could require several ancilla qubits, multi-qubit controlled operations, and also adds to the overall gate depth. For instance when H is a linear combination of Paulis as defined above, the block encoding

would require $\lceil \log_2 L \rceil + 2$ ancilla qubits, and has a gate depth of $O(L)$.

In fact, in Appendix C.2, we analyze the cost (in terms of the gate depth, ancilla qubits, and number of classical repetitions) of our ground state property estimation algorithm when U_τ is implemented according to the Hamiltonian simulation algorithm in Sec. 4, as well as $2k$-order Trotter [48]. We demonstrate that our algorithm can still be implemented using a single ancilla qubit and no multi-qubit controlled gates. Moreover, despite these restrictions, we show that even when compared to state-of-the art techniques, there are regimes where our method has a shorter gate depth per coherent run (See Table A2).

6 Applications to Quantum linear systems

The quantum linear systems algorithm can be stated as follows: Given access to a Hermitian matrix $H \in \mathbb{C}^{N\times N}$ and some initial state $|b\rangle$, prepare the quantum state $|x\rangle = H^{-1}|b\rangle / \left\| H^{-1}|b\rangle \right\|$. Ever since the first quantum algorithm for this problem by Harrow, Hassidim, and Lloyd [65], the quantum linear system algorithm has been widely studied. The complexity of this algorithm has been progressively improved through a series of results [26, 27, 45]. Recently, adiabatic-inspired approaches have also been reported [68, 69], which optimally solve this problem [51]. For several applications, simply preparing the state $|x\rangle$ may not be useful. Rather, one is often interested in extracting useful information from this state, such as estimating the expectation value of an observable O, i.e. $\langle x|O|x\rangle$.

Just like in the previous section, we apply *Analog LCU* to develop two quantum linear systems algorithms in continuous-time (Sec 6.1): the first one is an analog variant of the direct approach in [26] while the second one is more amenable to near-term implementation. Following this, we use the *Single-Ancilla LCU* approach to develop a randomized quantum algorithm for estimating $\langle x|O|x\rangle$ which is implementable on early fault-tolerant quantum computers (Sec. 6.2).

Let us begin by formally stating the quantum linear systems problem.

Quantum linear systems: Suppose we have access to a Hermitian matrix $H \in \mathbb{C}^{N\times N}$ such that its eigenvalues lie in the interval $[-1, -1/\kappa] \cup [1/\kappa, 1]$. Then, given a procedure that prepares the N-dimensional quantum state $|b\rangle$, a quantum linear systems algorithm prepares a quantum state that is $O(\varepsilon)$ - close to

$$|x\rangle = \frac{H^{-1}|b\rangle}{\left\| H^{-1}|b\rangle \right\|}.$$

It is worth noting that the quantum linear systems algorithm is different from its classical counterpart in that by preparing $|x\rangle$, one does not have access to the entries of the classical vector $\vec{x}$. To extract the entire solution vector $\vec{x}$ from the quantum state $|x\rangle$, would require $\Omega(N)$ cost (via tomography). In quantum linear systems, thus, often one is interested in extracting useful information out of the state $|x\rangle$, such as estimating the expectation value $\langle x|O|x\rangle$, for some observable O.

Also, the assumption that H is a Hermitian matrix is without loss of generality. Given any non-Hermitian $H \in \mathbb{C}^{M\times N}$, there exist efficient procedures to obtain a Hermitian matrix $\tilde{H}$ of dimension $(M+N)\times(M+N)$ [65]. Then, one may instead implement quantum linear systems with $\tilde{H}$ instead of H.

6.1 Applying Analog LCU: Continuous-time quantum linear systems algorithms

In this section, we develop analog quantum algorithms for solving quantum linear systems. Following the exposition in Sec. 3.3, we shall assume that we are given a system Hamiltonian H. We couple this Hamiltonian (the primary system) to two ancillary continuous-variable systems via the interaction Hamiltonian

$$H' = H \otimes \hat{y} \otimes \hat{z}. \tag{75}$$

The primary system will be initialized in the quantum state $|b\rangle$ while the two ancillary systems will be in some continuous-variable states. The quantum algorithms developed in this subsection involve evolving the overall system according to H' for some time. Following this, we shall show that the primary system is in the state $|x\rangle$ (or close to it) with an amplitude of $\Omega(1/\kappa)$.

We begin with the first quantum algorithm, which is an analog analogue of the quantum linear systems algorithm of [26].

Continuous-time quantum linear systems algorithm: Consider the function $f(y) = ye^{-y^2/2}$, where $y \in \mathbb{R}$. As

$$\int_0^\infty dy\ f(y) = 1 \tag{76}$$

$$\implies \int_0^\infty dy\ f(xy) = 1/x, \tag{77}$$

which holds for any $x \neq 0$. For any function $g(y)$, suppose its Fourier transform is $\mathcal{F}(g(y)) = F(\omega)$, then $\mathcal{F}(g'(y)) = i\omega F(\omega)$. If $g(y) = e^{-y^2/2}$, we have that $g'(y) = -ye^{-y^2/2} = -f(y)$. This implies,

$$\frac{i}{\sqrt{2\pi}} \int_{-\infty}^\infty dz\ ze^{-z^2/2} e^{-izy} = ye^{-y^2/2}, \tag{78}$$

and,

$$\frac{1}{x} = \frac{i}{\sqrt{2\pi}} \int_0^\infty dt \int_{-\infty}^\infty dz\ ze^{-z^2/2} e^{-izxt}.$$

Next, we will prove via a lemma that the upper limit of the outer integral can be truncated at $T = \widetilde{O}(\kappa)$, without introducing significant error.

Lemma 16. *Suppose $\varepsilon > 0, z \in \mathbb{R}$, and $x \in \mathbb{R}\backslash\{0\}$. Then there exists $T \in \Theta\left(\kappa\sqrt{\log(\kappa/\varepsilon)}\right)$, such that on the domain $[-1, -1/\kappa] \cup [1/\kappa, 1]$,*

$$\left| \frac{1}{x} - \frac{1}{\sqrt{2\pi}} \int_0^T dt\ \int_{-\infty}^\infty dz\ ze^{-z^2/2} e^{-izxt} \right| \leq \varepsilon. \tag{79}$$

Proof. We have to evaluate the quantity

$$\left| \frac{1}{\sqrt{2\pi}} \int_T^\infty dt\ \int_{-\infty}^\infty dz\ ze^{-z^2/2} e^{-izxt} \right|$$

We first evaluate the outer integral and obtain,

$$\left|\frac{1}{\sqrt{2\pi}}\int_T^\infty dt\ \int_{-\infty}^\infty dz\ ze^{-z^2/2}e^{-izxt}\right| = \left|\int_T^\infty dt\ xt\ e^{-x^2t^2/2}\right| \qquad [\text{ Using Eq. (78) }] \tag{80}$$

$$= \left|\frac{1}{x}\int_{x^2T^2/2}^\infty dy\ e^{-y}\right| \qquad [\ y = x^2t^2/2\ ] \tag{81}$$

$$= \left|\frac{1}{x}\cdot e^{-x^2T^2/2}\right| \tag{82}$$

$$\leq \frac{1}{|x|}\left|e^{-x^2T^2/2}\right|. \tag{83}$$

Now for $T = \kappa\sqrt{2\log(\kappa/\varepsilon)}$, we have $\left|e^{-x^2T^2/2}\right| \leq \varepsilon/\kappa$. Now as $|x| \geq 1/\kappa$, we have that Eq. (83) is upper bounded by ε. So finally,

$$\left|\frac{1}{x} - \frac{1}{\sqrt{2\pi}}\int_0^T dt\ \int_{-\infty}^\infty dz\ ze^{-z^2/2}e^{-izxt}\right| = \left|\frac{1}{\sqrt{2\pi}}\int_T^\infty dt\ \int_{-\infty}^\infty dz\ ze^{-z^2/2}e^{-izxt}\right| \leq \varepsilon. \tag{84}$$

□

In order to design the analog quantum algorithm, consider that the effective interaction Hamiltonian is $H' = H\otimes\hat{y}\otimes\hat{z}$. While the system Hamiltonian H is prepared in some input state $|b\rangle$, the first ancilla system is prepared in the first-excited state of a one-dimensional quantum Harmonic oscillator

$$|\psi_h\rangle = \frac{1}{(2\pi)^{1/4}}\int_{-\infty}^\infty dy\ ye^{-y^2/4}|y\rangle\,. \tag{85}$$

The second ancilla system is in the ground state of a "particle in a ring" of diameter 1, given by

$$|\tau\rangle = \int_0^1 dz\ \ |z\rangle\,. \tag{86}$$

Then evolving the overall system according to H' for time T, we obtain

$$|\eta_t\rangle = e^{-i\tilde{H}T}|b\rangle|\psi_h\rangle|\tau\rangle \tag{87}$$

$$= \int_0^1 dz\ \int_{-\infty}^\infty \frac{dy}{(2\pi)^{1/4}}\ ye^{-y^2/4}e^{-iyzHT}|b\rangle|y\rangle|z\rangle \tag{88}$$

$$= \frac{1}{T}\int_0^T dt\ \int_{-\infty}^\infty \frac{dz}{\sqrt{2\pi}}\ ze^{-z^2/2}e^{-iztH}|b\rangle|\psi_h\rangle|\tau\rangle + |\Phi\rangle^\perp \qquad [\text{ Change of variable } t = Ty\ ] \tag{89}$$

Now, by choosing time $T = \Theta\left(\kappa\sqrt{\log(\kappa/\varepsilon)}\right)$, from Lemma 16, we obtain a quantum state that is $O(\varepsilon/T)$-close to

$$|\eta_t\rangle = \frac{H^{-1}}{T}|b\rangle|\psi_h\rangle|\tau\rangle + |\Phi\rangle^\perp\,. \tag{90}$$

The cost of preparing this state is thus, linear in κ (upto log factors), which is optimal. For fully fault tolerant quantum computers, the state $|x\rangle$ is obtained by using variable time amplitude amplification which is a complicated subroutine, requiring a large number of controlled operations [26, 27, 67]. However, this procedure ensures that the overall query

complexity of the quantum linear systems algorithm is still $\widetilde{O}(\kappa)$. Alternatively, in the circuit model, $\widetilde{O}(\kappa)$-rounds of amplitude amplification can yield a quantum state $O(\varepsilon)$-close to $|x\rangle$. The overall cost of this procedure is $\tilde{O}(\kappa^2)$. Note that both these procedures are suitable for implementation on fully fault-tolerant quantum computers. Moreover, procedures such as amplitude amplification have no known continuous-time analogues.

Thus, for our analog procedure, after preparing the state $|\eta_t\rangle$, we simply post-selecting on obtaining $|\psi_h\rangle$ in the second register. This allows us to obtain $|x\rangle$ in the first register with probability $\widetilde{\Omega}(1/\kappa^2)$. Thus $O(\kappa^2)$ repetitions of the continuous-time procedure would allow us to obtain $|x\rangle$.

Although this procedure works in general, the quantum state $|\tau\rangle$ might be difficult to prepare experimentally. In fact, for continuous-variable systems, Gaussian states are the easiest to prepare and manipulate [56]. So, next, we provide a quantum algorithm for which it suffices to prepare both the ancillary registers in Gaussian states.

Continuous-time quantum linear systems algorithm using only Gaussian states: The previous quantum algorithm requires us to prepare the non-Gaussian continuous-variable state

$$|\tau\rangle = \frac{1}{\sqrt{T}} \int_0^T dz \ |z\rangle \, .$$

Since Gaussian states are typically easier to generate and manipulate, let us design alternative algorithms using Gaussian states only. The general idea is to approximate $\int_{-\infty}^{+\infty} dt$ by $\int_{-\infty}^{+\infty} dt \ e^{-t^2/2T^2}$ (rather than $\int_{-T}^{T} dt$) for large enough T. The analogue of Lemma 16 becomes

Lemma 17. *Suppose $\varepsilon > 0, z \in \mathbb{R}$, and $x \in \mathbb{R} \setminus \{0\}$. Then, there exists $T \geq \kappa^{3/2}/\sqrt{\varepsilon}$, such that on the domain $[1/\kappa, 1]$,*

$$\left| \frac{1}{x} - \frac{1}{2\pi} \int_{-\infty}^{+\infty} dt \ e^{-t^2/2T^2} \int_{-\infty}^{+\infty} dz \ e^{-z^2/2} e^{-ixtz} \right| \leq \Theta(\varepsilon). \tag{91}$$

Proof. We have, using the fact that the Fourier transform of a Gaussian is a Gaussian

$$\begin{aligned} \frac{1}{2\pi} \int_{-\infty}^{+\infty} dt \ e^{-t^2/2T^2} \int_{-\infty}^{+\infty} dz \ e^{-z^2/2} e^{-ixtz} &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} dt \ e^{-t^2/2T^2} \ e^{-x^2t^2/2} \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} dt \ e^{-\left(x^2+1/T^2\right)t^2/2} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} dt \ e^{-\tilde{x}^2t^2/2} \\ &= \frac{1}{\tilde{x}} \end{aligned}$$

where, we have set $\tilde{x} = \sqrt{x^2 + 1/T^2}$. Therefore, it remains to bound

$$\left| \frac{1}{x} - \frac{1}{\tilde{x}} \right| = \left| \frac{1}{x} \left(1 - \frac{x}{\tilde{x}} \right) \right| \leq \frac{1}{|x|} \left| 1 - \frac{1}{\sqrt{1 + 1/x^2T^2}} \right| \leq \frac{1}{|x|} \cdot \frac{1}{x^2T^2} \leq \varepsilon$$

□

Unfortunately, the scaling of T is worse than for the non-Gaussian approach since T scales as $\kappa^{3/2}$ (instead of linear) and the dependence of precision is $1/\sqrt{\varepsilon}$ (rather than inverse-logarithmic). Moreover, as the Gaussian function is even, the procedure works only for positive semi-definite Hamiltonians. Nevertheless, this allows us to design a quantum linear systems algorithm using only Gaussian states as ancillae.

Let us again consider the interaction Hamiltonian $H' = H \otimes \hat{y} \otimes \hat{z}$, where H is now some positive definite Hamiltonian with its eigenvalues lying in $[1/\kappa, 1]$. We prepare both the ancilla registers in a Gaussian state which, similarly to Sec. 5.1, is defined as follows

$$|\psi_g\rangle = \frac{1}{(2\pi)^{1/4}} \int_{-\infty}^{\infty} dz \; e^{-z^2/4} |z\rangle .$$

Indeed, it suffices to let the state $|b\rangle |\psi_g\rangle |\psi_g\rangle$ evolve under Hamiltonian H' for time T to obtain

$$e^{-iH'T} |b\rangle |\psi_g\rangle |\psi_g\rangle = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dz \int_{-\infty}^{\infty} dy \; e^{-(y^2+z^2)/4} e^{-iyzHT} |b\rangle |y\rangle |z\rangle .$$

If we choose some $T \geq \kappa^{3/2}/\sqrt{\varepsilon}$, we have

$$\left(I \otimes |\psi_g\rangle\langle\psi_g| \otimes |\psi_g\rangle\langle\psi_g|\right) e^{-i\tilde{H}T} |b\rangle |\psi_g\rangle |\psi_g\rangle = \frac{1}{2\pi} \int_{-\infty}^{+\infty} dz \int_{-\infty}^{+\infty} dy \; e^{-(y^2+z^2)/2} e^{-iyzHT} |b\rangle |\psi_g\rangle |\psi_g\rangle \tag{92}$$

$$= \frac{1}{2\pi T} \int_{-\infty}^{+\infty} dt \; e^{-t^2/2T^2} \int_{-\infty}^{+\infty} dz \; e^{-z^2/2} e^{-itzH} |b\rangle |\psi_g\rangle |\psi_g\rangle \tag{93}$$

where we have used the change of variable $t = Ty$. So,

$$e^{-i\tilde{H}T} |b\rangle |\psi_g\rangle |\psi_g\rangle = \frac{1}{2\pi T} \int_{-\infty}^{+\infty} dt \; e^{-t^2/2T^2} \int_{-\infty}^{+\infty} dy \; e^{-z^2/2} e^{-itzH} |b\rangle |\psi_g\rangle |\psi_g\rangle + |\phi\rangle^{\perp} \tag{94}$$

$$= \frac{H^{-1}}{T} |b\rangle |\psi_g\rangle |\psi_g\rangle + |\Phi\rangle^{\perp} + O(\varepsilon/T) \qquad [\text{ From Lemma 17 }]. \tag{95}$$

So, by post-selecting on obtaining $|\psi_g\rangle$in the second register, we obtain a state that is $O\left(\varepsilon/\kappa\right)^{3/2}$ -close to

$$|x\rangle = \frac{H^{-1} |b\rangle}{\left\| H^{-1} |b\rangle \right\|},$$

with amplitude $\tilde{\Omega}\left(\sqrt{\varepsilon}/\kappa^{3/2}\right)$. Although the complexity of this algorithm is worse than the continuous-time quantum algorithm in the previous section, it requires only Gaussian states. Consequently, it is more suitable for being implementable in the near term.

One can improve the complexity of this quantum linear systems algorithm by replacing the Gaussian state in the second register with the flat state $|\tau\rangle$. For positive definite Hamiltonians, if the first ancillary system is in a Gaussian state while the second one is in $|\tau\rangle$, we can still obtain a quantum state that is $O(\varepsilon/\kappa)$-close to the solution of the quantum linear systems in time $\widetilde{O}(\kappa)$. This follows from observing

$$\frac{1}{x} = \int_{-\infty}^{\infty} \frac{dt}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} e^{-ixtz} = 2 \int_{0}^{\infty} \frac{dt}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} e^{-ixtz}. \tag{96}$$

We can truncate the outer integral to $T = \Theta(\kappa\sqrt{\log(\kappa/\varepsilon)})$, and introduce only an additive ε-error.

Our analog approach provides a more physically motivated model for implementing quantum linear systems. We believe several existing quantum technological platforms might already be able to engineer these interactions for system Hamiltonians of small dimensions. It would be interesting to explore whether one can obtain a quantum linear systems algorithm by using just a single continuous-variable ancilla.

Next, we move on to the problem of estimating expectation values of observables with respect to the solution of quantum linear systems. For this, we make use of the *Single-Ancilla LCU* technique.

6.2 Applying Single-Ancilla LCU: estimating expectation values of observables

In the *Single-Ancilla LCU* framework, we shall consider that the Hamiltonian H can be accessed through the time evolution operator $U_t = \exp[-itH]$. Much like the ground state property estimation algorithm in this framework, we estimate: (a) the maximal time of evolution per coherent run ($\tau_{\max}$), and (b) the total number of repetitions T. The total time of evolution is then the product of T and $\tau_{\max}$. Our randomized quantum algorithms for estimating $\langle x|O|x\rangle$, for any measurable observable O. We consider the discrete LCU decomposition of H^{-1} of Ref. [26]. We begin by stating the discretized version of the expression in Lemma 16.
Let,

$$g(x) = \frac{i}{\sqrt{2\pi}} \sum_{j=0}^{J-1} \Delta_y \sum_{k=-K}^{K} \Delta_z z_k e^{-z_k^2/2} e^{-ixy_j z_k}, \tag{97}$$

where $y_j = j\Delta_y$ and $z_k = k\Delta_z$, for some $J \in \Theta(\frac{\kappa}{\gamma}\log(\kappa/\gamma))$, $K = \Theta(\kappa\log(\kappa/\gamma))$, $\Delta_y = \Theta(\gamma/\sqrt{\log(\kappa/\gamma)})$ and $\Delta_z = \Theta((\kappa\sqrt{\log(\kappa/\gamma)})^{-1})$. Then, Childs et al. [26] proved that $|1/x - g(x)| \leq \gamma$ in the domain $[-1,-1/\kappa] \cup [1/\kappa, 1]$. From this LCU it is clear that in order to approximate H^{-1} in this domain, the time parameter of the Hamiltonian simulation is at most,

$$t = \Theta(y_J z_K) = \Theta\left(\kappa\log(\kappa/\gamma)\right). \tag{98}$$

Furthermore, from [26], the ℓ_1-norm of the LCU coefficients were shown to be upper bounded by $\|c\|_1 = \Theta(\kappa\sqrt{\log(\kappa/\gamma)})$. This LCU decomposition allows us to use Algorithm 2 to estimate $\langle x|O|x\rangle$. We will prove that it suffices to choose

$$\gamma = \frac{\varepsilon}{18\|O\|}.$$

From Eq. (98) and the aforementioned choice of γ, the maximal time evolution for each coherent run of our algorithm is

$$\tau_{\max} = O\left(\kappa\log\left(\frac{\|O\|\,\kappa}{\varepsilon}\right)\right).$$

We formally prove the correctness of our method via the following theorem

Theorem 18 (Expectation values of observables with respect to the solution of quantum linear systems)**.** *Let H be a Hermitian matrix such that its non zero eigenvalues lie in $[-1,-1/\kappa] \cup [1/\kappa, 1]$. Let O be an observable, and $\varepsilon, \delta, \gamma \in (0,1)$. Then if*

$$\gamma = \frac{\varepsilon}{18\|O\|},$$

such that

$$\left\|H^{-1} - g(H)\right\| \leq \gamma,$$

and,

$$T = O\left(\frac{\|O\|^2\,\kappa^4\log^2\left(\frac{\|O\|\kappa}{\varepsilon}\right)\ln(1/\delta)}{\varepsilon^2}\right).$$

then Algorithm 2 outputs, with probability at least $(1-\delta)^2$, parameters μ and $\tilde{\ell}$ such that

$$\left|\frac{\mu}{\tilde{\ell}} - \langle x|O|x\rangle\right| \leq \varepsilon,$$

using T repetitions of the quantum circuit in Fig. 2, and only one ancilla qubit. The maximal time evolution of H in each coherent run is,

$$\tau_{max} = O\left(\kappa \log\left(\frac{\|O\|\,\kappa}{\varepsilon}\right)\right).$$

Proof. First observe that $\kappa \geq \|f(H)\| = \left\|H^{-1}\right\| \geq 1$, and similarly $\kappa^2 \geq \ell^2 = \left\|H^{-1}\,|b\rangle\right\|^2 \geq \ell_* = 1$. So, after the substitution the appropriate parameters, we find that choice of γ is the same as in Theorem 10. Also, for this choice of γ, ℓ_1-norm of the LCU coefficients of $g(H)$ is

$$\|c\|_1 = O\left(\kappa\sqrt{\log\left(\frac{\kappa\|O\|}{\varepsilon}\right)}\right).$$

So, for $\ell_* = 1$, Algorithm 2 outputs parameters μ and $\tilde{\ell}$ such that

$$\left|\mu - \mathrm{Tr}[O\; H^{-1}\,|b\rangle\,\langle b|\,H^{-1}]\right| \leq \varepsilon/3,$$

and,

$$\left|\tilde{\ell} - \ell^2\right| \leq \frac{\varepsilon}{3\|O\|}.$$

From Theorem 9, the parameters μ and $\tilde{\ell}$ satisfy,

$$\left|\frac{\mu}{\tilde{\ell}} - \langle x|O|x\rangle\right| \leq \varepsilon.$$

Each coherent run quantum circuit costs no more than $2\tau_{\max} + \tau_b$, where

$$\tau_{\max} = O\left(\kappa \log\left(\frac{\|O\|\,\kappa}{\varepsilon}\right)\right), \tag{99}$$

The total number of iterations required can be obtained from Theorem 10 by substituting the appropriate values of $\|c\|_1$ and ℓ^* as

$$T = O\left(\frac{\|O\|^2\,\kappa^4 \log^2\left(\frac{\|O\|\kappa}{\varepsilon}\right)\ln(1/\delta)}{\varepsilon^2}\right). \tag{100}$$

□

It is important to note that our quantum algorithm extracts useful information about the quantum state $|x\rangle$. It outputs a number as opposed to a quantum state, which is distinct from the quantum linear systems problem in general. The total time evolution of H is $\widetilde{O}(\kappa^5\|O\|^2/\varepsilon^2)$.

Comparison with prior works: Let us now compare the complexity of our procedure with that of other quantum linear systems algorithms. Much like the *Single-Ancilla LCU* algorithm for ground state property estimation, we compare our algorithm with quantum

linear systems algorithms making use of (a) the *Standard LCU* method [26], (b) QSVT [45, 67] and the (c) discrete adiabatic theorem [51]. We summarize the comparison in Table 4.

First note that the HHL algorithm [65] requires access to a time-evolution oracle and can estimate $\langle x|O|x\rangle$. The algorithm also makes use of quantum phase estimation, which using modern methods can be implemented with $O(1)$ ancilla qubits. This algorithm can estimate the expectation value in three ways: Preparing $|x\rangle$ first using quantum amplitude amplification requiring a maximal time evolution of $\widetilde{O}(\kappa^2\|O\|/\varepsilon)$, followed by $O(\ ||O||^2/\varepsilon^2)$ measurements of O. Thus, the maximal time evolution per coherent run is exponentially worse than *Single-Ancilla LCU*. However, the total evolution time has a better dependence on κ, but a worse dependence on other parameters. Moreover, the coherent estimation of the expectation value by quantum amplitude estimation cannot reduce the dependence on ε, which continues to be $O(1/\varepsilon^2)$, while needing more ancilla qubits (owing to the block encoding of O). Finally, if quantum amplitude amplification or estimation are not used, the maximal time evolution of H per coherent run is $O(\kappa\|O\|/\varepsilon)$, which is exponentially worse than our method. On the other hand, the total evolution time is $O(\kappa^3\|O\|^3/\varepsilon^3)$ has a better dependence on κ, but a worse dependence on all other parameters.

As before, for Standard LCU, we consider access to the Hamiltonian evolution oracle (HE access model) while for QSVT and the adiabatic approaches, we consider the block encoding framework (BE access model). For Standard LCU, we consider time evolution of a multi-qubit controlled time evolution operator and compare the maximal time of evolution of H, the number of classical repetitions with our method. Although a direct comparison cannot be made between the HE model and the BE model, for QSVT and adiabatic based approaches, we consider the number of queries made to an $(\alpha_H, a_H, 0)$-block encoding of H in one coherent run of the algorithm, and the total number of runs required. In addition to this, for both these access models, we compare the number of ancilla qubits needed.

The LCU-based procedure by Childs, Kothari and Somma [26] requires $O(\log(\kappa\|O\|/\varepsilon))$ ancilla qubits and sophisticated multi-qubit controlled operations. There are three ways to estimate $\langle x|O|x\rangle$:

- If $|x\rangle$ is prepared first using quantum amplitude amplification, the maximal time evolution of (multi-qubit controlled) H per coherent run is $O(\kappa^2\log^2(\kappa\|O\|/\varepsilon))$, which is higher than our method by a factor of κ. However, the total evolution time is lower by a factor of $O(\kappa^3)$, as only $O(||O||^2/\varepsilon^2)$ repetitions of this procedure is needed.

- Given access to an $(\alpha_O, a_O, 0)$-block encoding of O, the desired expectation value can be measured coherently using quantum amplitude estimation. This reduces the overall dependence on the precision to $1/\varepsilon$, at the cost of increasing the maximal time of evolution of H to $\widetilde{O}(\kappa^2\alpha_O/\varepsilon)$, which is also the total evolution time. Furthermore, the number of ancilla qubits needed increases by a_O, owing to the implementation of the block encoding of O.

- The maximal evolution time per coherent run can be minimized by avoiding the use of quantum amplitude amplification or estimation. Standard LCU followed by a direct measurement of O leads to a maximal time evolution of $O(\kappa\log(||O||\kappa/\varepsilon))$ which matches that of *Single-Ancilla LCU*. On the other hand, the total number of repetitions is $\widetilde{O}(\kappa^2\|O\|^2/\varepsilon^2)$ which is lower than our method by a factor of κ^2. However, the overhead due to ancilla qubits and multi-qubit controlled operations remain.

The QSVT based approach [45] queries an $(\alpha_H, a_H, 0)$-block encoding of H (say U_H), and implements a polynomial approximation of H^{-1} using queries to U_H (and $U_H^\dagger$), interleaved with a single qubit phase rotations. So, we will measure the complexity in terms of the query complexity per coherent run as well as the overall queries. Much like standard LCU, it can estimate $\langle x|O|x\rangle$ in three ways:

- Preparing $|x\rangle$ by QSVT followed by amplitude amplification requires $\widetilde{O}(\alpha_H \kappa^2 \log(\kappa \|O\| / \varepsilon))$ queries to U_H per coherent run, followed by $O(||O||^2/\varepsilon^2)$ classical repetitions. The procedure requires $O(a_H)$ ancilla qubits and multi-qubit controlled operations to construct the block encoding.

- Directly using quantum amplitude estimation to estimate the desired expectation value assumes access to an $(\alpha_O, a_O, 0)$-block encoding of O, and requires $\widetilde{O}\left(\alpha_H \alpha_O \kappa^2 \varepsilon^{-1}\right)$ queries per coherent run which is also the overall query complexity. The number of ancilla qubits needed in the overall procedure scales as the number of ancilla qubits required to construct the respective block encodings, i.e. $O(a_H + a_O)$. This can be quite large depending on the way this block encoding is constructed. Thus, the overall query complexity of this procedure is lower than the total time of evolution of our algorithm.

- If quantum amplitude amplification or estimation is not used, $\langle x|O|x\rangle$ can be estimated by implementing the polynomial approximation of H^{-1} by querying the block encoding followed by a measurement of O. This method has a reduced query complexity per coherent run given by $O(\alpha_H \kappa \log(\kappa||O||/\varepsilon))$, which matches the $\tau_{\max}$ of our method (in terms of κ) but the linear dependence on α_H means there are regimes where our method requires less cost per coherent run. The total number of classical repetitions has a quadratically better dependence on κ as compared to our method given by $\widetilde{O}(\kappa^2 \|O\|^2 / \varepsilon^2)$. However constructing the block encoding requires $O(a_H)$ ancilla qubits and the cost per coherent run has a dependence on α_H (which can be $O(\log L)$ and β respectively, when H is a linear combination of Pauli operators as described in Eq. (A15)).

It is important to note that for both the aforementioned approaches the query complexity per coherent run can be brought down to $\widetilde{O}(\kappa)$ by using variable time amplitude amplification (VTAA) [83] instead of standard amplitude amplification. However even with recent improvements [67], this procedure requires an additional $O(\log \kappa)$ ancilla qubits (overall $O(a_H + \log \kappa)$ ancilla qubits are needed), and even more multi-qubit controlled operations to be implemented. Hence this is significantly out of reach for early fault tolerant quantum computers. We do not compare our method with quantum linear systems algorithms making use to VTAA.

Instead we consider the state-of-the-art quantum linear systems algorithm by Costa et. al. [51], which assumes an $(\alpha_H, a_H, 0)$-block encoding of H, and implements a block encoding of the some interpolated Hamiltonian $H(s)$, similar to other adiabatic approaches for this problem [68, 69]. It then proceeds to construct an interpolated quantum walk operator out of the block encoding, and carries out a fine grained analysis of the spectrum of this operator for discrete time steps, in accordance with the discrete adiabatic theorem, followed by eigenstate filtering. Their method prepares $|x\rangle$ using $O(\alpha_H \kappa \log(1/\varepsilon))$ queries to the block encoding of H. Thus, this algorithm uses more ancilla qubits and also needs several multi-qubit controlled gates. However, it achieves a $\log \kappa$ improvement in the error dependence over LCU and QSVT based approaches, and at the same time has a linear

Algorithm	Access	Variant	Ancilla	Cost per coherent run	Classical repetitions
Standard LCU [26]	HE (MQC)	QAA + classical repetitions	$O\left(\log(\kappa\|O\|/\varepsilon)\right)$	$\tilde{O}(\kappa^2)$	$O\left(\frac{\|O\|^2}{\varepsilon^2}\right)$
		QAE	$O\left(a_O + \log(\kappa\|O\|/\varepsilon)\right)$	$\tilde{O}\left(\alpha_O\kappa^2/\varepsilon\right)$	$O(1)$
		Without QAA or QAE	$O\left(\log(\kappa\|O\|/\varepsilon)\right)$	$O\left(\kappa\log(\kappa\|O\|/\varepsilon)\right)$	$\tilde{O}\left(\frac{\kappa^2\|O\|^2}{\varepsilon^2}\right)$
QSVT [45]	BE (MQC)	QAA + classical repetitions	$O(a_H)$	$\tilde{O}\left(\alpha_H\kappa^2\right)$	$O\left(\frac{\|O\|^2}{\varepsilon^2}\right)$
		QAE	$O(a_O + a_H)$	$\tilde{O}\left(\frac{\alpha_O\alpha_H\kappa^2}{\varepsilon}\right)$	$O(1)$
		Without QAA or QAE	$O\left(a_H\right)$	$O\left(\alpha_H\kappa\log(\kappa\|O\|/\varepsilon)\right)$	$\tilde{O}\left(\frac{\kappa^2\|O\|^2}{\varepsilon^2}\right)$
Discrete adiabatic theorem [51]	BE (MQC)	Classical repetitions	$O(a_H)$	$O\left(\alpha_H\kappa\log(\|O\|/\varepsilon)\right)$	$O\left(\frac{\|O\|^2}{\varepsilon^2}\right)$
		QAE	$O(a_O + a_H)$	$\tilde{O}\left(\frac{\alpha_O\alpha_H\kappa}{\varepsilon}\right)$	$O(1)$
This work	HE	–	1	$O\left(\kappa\log(\kappa\|O\|/\varepsilon)\right)$	$\tilde{O}\left(\frac{\kappa^4\|O\|^2}{\varepsilon^2}\right)$

Table 4: Consider any Hamiltonian H with eigenvalues in $[-1,-1/\kappa]\cup[1/\kappa,1]$. Then given an input state $|b\rangle$, define $|x\rangle = H^{-1}|b\rangle / \left\|H^{-1}|b\rangle\right\|$, and suppose O is any measurable observable. In this table, we compare with our technique, the cost of various quantum algorithms for estimating $\langle x|O|x\rangle$ to additive accuracy ε. For both Standard LCU [34] and our method (*Single-Ancilla LCU*), we consider that H can only be accessed through the time evolution operator $U_t = \exp[-iHt]$. This is the Hamiltonian Evolution (HE) model, where the cost of each coherent run is estimated in terms of the maximal time evolution of H. The number of repetitions refers to the total number of times the circuit is run. The total evolution time is then the product of these two quantities. The Standard LCU approach requires implementing multi-qubit controlled (MQC) time evolution operators, which is indicated in the Table along with the access model. Both the QSVT-based algorithm in Ref. [45], and the state-of-the-art algorithm by Costa et al. [51], considers the block encoding access model (denoted by BE in the Table). The cost is estimated in terms of the number of queries made to an $(\alpha_H, a_H, 0)$-block encoding of H. The cost per coherent run is the number of queries made, while the product of this quantity and the number of classical repetitions is the overall query complexity. Finally, in order to estimate the desired expectation value directly via quantum amplitude estimation (QAE), we assume that O is accessed via an $(\alpha_O, a_O, 0)$-block encoding.

dependence on κ, without requiring the complicated VTAA procedure. Additionally, this technique requires four extra ancilla qubits (in addition to a_H needed for implementing the block encoding of H). The query complexity per run is thus slightly lower than the maximal time evolution of H needed by our method in terms of the precision (by a factor of $\log(\kappa)$), but the normalization of the block encoding α_H is multiplied. Moreover, this method would need $O(\ ||O||^2/\varepsilon^2)$ classical repetitions to estimate $\langle x|O|x\rangle$ incoherently. Using quantum amplitude estimation, this would require $\widetilde{O}(\alpha_H\alpha_O\kappa\|O\|/\varepsilon)$ queries per coherent run, which is also the total query complexity. This is lower than the total time evolution of our algorithm, but the maximal time evolution of H per coherent run is exponentially higher (in terms of $1/\varepsilon$).

Huang et al. [70] analyze the possibility of solving quantum linear systems using near-term quantum devices. In particular, their algorithms can be adapted to estimate expectation values of observables in the Hamiltonian evolution model. If x is the (unnormalized) state that minimizes the loss function $\left\|x - H^{-1}|b\rangle\right\|^2$, the authors devise a

strategy to write down the description of x as a linear combination

$$x = \sum_{j=1}^{m} \alpha_j \left|\psi_{U_j}\right\rangle ,$$

where $\alpha_j \in \mathbb{C}$ and $|\psi_{U_j}\rangle$ is any *good quantum state* such that $H^{-1}|b\rangle \in \text{Span}\{|\psi_{U_j}\rangle\}$ for $j \in [1, m]$. If $H = \sum_{k=1}^{L} \lambda_k U_k$ is a linear combination of unitaries U_k, then given access to a set of such *good quantum states*, the authors first estimate the optimal α_i's that minimize the aforementioned loss function (up to ε) accuracy, through a hybrid quantum classical algorithm: a combination of the Hadamard test and quadratic optimization. The quantum part of the algorithm requires $O(L^2 m^3/\varepsilon)$ classical repetitions, and the cost of each such run depends on the cost of implementing the quantum circuit that prepares $|\psi_{U_i}\rangle$. Note that the Hadamard test requires a single ancilla qubit. Finally, the outcomes of the quantum procedure serve as inputs to the classical optimization problem that finds the optimal α_i's. Then having obtained such an x they re-use the Hadamard test (a slightly modified version, to be precise) to estimate $\langle\psi_{U_j}|O|\psi_{U_i}\rangle$ for each pair $i, j \in [1, m]$, and then add up these quantities classically, weighted by the optimal α_i's to estimate $x^\dagger O x$. The authors show that one way to obtain a *good quantum state* is by applying the time evolution operator to $|b\rangle$, i.e. each $|\psi_{U_j}\rangle = e^{-it_j H}|b\rangle$. Here t_j is chosen as per the quantum linear systems algorithm of Ref. [26], and so any such $t_j \leq O(\kappa \log(\kappa/\varepsilon))$. Thus, the procedure requires using a single ancilla qubit and a maximal Hamiltonian evolution time $\tau_{\max} = \widetilde{O}(\kappa)$ in each coherent run, which matches our method. However, the number of classical repetitions is exponentially higher than our procedure. First, the estimation of the α_i's require $O(L^2 m^3/\varepsilon)$ classical repetitions, such that the maximal time evolution of H per run is $\widetilde{O}(\kappa)$. Note that m can be $O(N)$ in general for a $N \times N$ Hamiltonian H. Furthermore, for each $i, j \in [1, m]$, the quantity $\langle\psi_{U_j}|O|\psi_{U_k}\rangle$ needs to be estimated to $O(\varepsilon/m^2)$ accuracy. So the total number of classical repetitions is $O(m^4/\varepsilon^2)$. Thus, overall, the total evolution time is exponentially higher than our method. There are other issues: x is not normalized and so accounting for the normalization factor to ultimately estimate $\langle x|O|x\rangle$ adds to the overall cost. Moreover this estimation scheme does not work for general observables. The authors consider observables which can be diagonalized as $O = UDU^\dagger$, such that U is efficiently implementable and the entries of the diagonal matrix D are also efficiently computable. Finally, there are other (variational) strategies in this work to obtain x, which are largely heuristic.

The quantum algorithm of Zhang et al. [11] also estimates $\langle x|O|x\rangle$. It assumes the availability of a quantum state $|\phi_0\rangle$ with a constant overlap with $|x\rangle$, requiring a maximal evolution time of $\widetilde{O}(\kappa)$. This can be achieved by using the adiabatic-based framework of [68]. Moreover, they also assume a $(\alpha_O, a_O, 0)$-block encoding of O which increases the ancilla space. Having prepared such a $|\phi\rangle$, any ground state property estimation algorithm (including *Single-Ancilla LCU*) can be employed to estimate the desired expectation value: the problem reduces to a particular case where both $\|c\|_1$ and η are constant. Their procedure requires a maximal time evolution of $\widetilde{O}(\kappa)$ per coherent run and $O(\alpha_O^2/\varepsilon^2)$ repetitions. Similar assumptions of access to such a state preparation procedure, would also result in our method estimating the desired expectation value with a matching maximal time evolution per coherent run, as well as the same total time evolution, without requiring a block encoding access to O, thereby requiring fewer ancilla qubits.

Finally, in the Appendix (Sec. C.2), we have provided explicit gate depths of our algorithm while implementing the time evolution operator using (a) Hamiltonain simulation by *Single-Ancilla LCU* and (b) $2k$-order Trotter (See Table A1). Overall, our algorithm still requires only a single ancilla qubit and no multi-qubit controlled gates. Despite this,

while comparing our method with other algorithms (See Table A3), we find that there are regimes where our method requires a shorter gate depth per coherent run as compared to even state-of-the-art methods.

7 Applications to quantum walks

So far, we have seen applications of the *Single-Ancilla LCU* and the *Analog LCU* approaches. In this section, we will show that the *Ancilla-free LCU* can be applied to the framework of quantum walks. Recall from Sec. 3.4, that this approach is useful when we are interested in the average projection of the LCU state $f(H)\,|\psi_0\rangle$ in some subspace of interest. In such scenarios, it suffices to sample the unitaries U_j according to the distribution of the LCU coefficients. This is because the projection of resulting density matrix on to this subspace is at least as large on average (See Theorem 2). We will show this is precisely the case for spatial search by quantum walks, where we are interested in finding out the expected number of steps after which the projection of the state of the quantum walk is high on some subset of the nodes (marked nodes) of the underlying Markov chain. Using *Ancilla-free LCU* allows us to retain the same quadratic speedup as in prior works while requiring no ancilla qubits (other than the quantum walk space).

We provide two quantum algorithms for spatial search by discrete-time quantum walks. The first one adapts the recent optimal algorithm of [32], wherein the authors used quantum fast-forwarding via *Standard LCU* [31] to obtain a generic quadratic speedup (up to a log factor). We show that by using *Ancilla-free LCU* instead, we can obtain the same quadratic speedup while requiring fewer ancilla qubits. This formalizes the observation of Ref. [42] - where the authors stated that the LCU of [32] could indeed be bypassed. Our second quantum algorithm relies on fast-forwarding continuous-time random walks, which also fits nicely in the *Ancilla-free LCU* framework. This algorithm too, achieves the same generic speedup, using fewer ancilla qubits as compared to [32]. For completeness, we would like to mention that the recent optimal spatial search algorithm by continuous-time quantum walk [33], can also be seen as an exposition of a continuous-time variant of *Ancilla-free LCU*.

Similar to the previous sections, here too, we shall present our results based on generic Hamiltonians. We will refer to quantum walks only as a particular case of our general results, which we believe are more broadly applicable. We begin with a very brief review of random and quantum walks.

7.1 Random and quantum walks: A very brief overview

Consider any ergodic, reversible Markov chain P defined on a vertex space X with $|X| = n$ nodes. One can think of such chains as a weighted graph of n nodes (For detailed definitions of these terms, refer to the Appendix of [41]). Then P is an $n \times n$ stochastic matrix. Let $p_{x,y}$ be the (x, y) - th entry of P. We shall consider that the singular values of P lie in $[0, 1]$. This is without loss of generality: one can implement the transformation $P \mapsto (I + P)/2$ to always ensure this.

Then starting from any initial probability distribution over X, represented by the row vector v_0, t-steps of a classical random walk, results in distribution $v_t = v_0 P^t$ over X. For any such P there exists a stationary distribution $\pi = (\pi_1, \pi_2, \cdots, \pi_n)$ such that $\pi = \pi P$. From any P one obtains a continuous-time random walk by using $Q = I - P$ (under

fairly general conditions). A continuous-time random walk, starting from v_0, evolves to $v_t = v_0 e^{Qt}$.

Since P is not symmetric in general, it would be useful to work with the Discriminant matrix D of P. D is an $n \times n$ symmetric matrix such that its $(x,y)^{\text{th}}$ - entry is $\sqrt{p_{xy}p_{yx}}$. The singular values of P are the same as the eigenvalues of D. Moreover, the state $|\sqrt{\pi}\rangle = \sum_{x \in X} \sqrt{\pi_x} |x\rangle$ is the eigenstate of D with eigenvalue 1.

In order to define a discrete-time quantum walk, define the unitary U_P such that

$$U_P |\bar{0}\rangle |x\rangle = \sum_{y=1}^{n} \sqrt{p_{xy}} |y, x\rangle .$$

where $|\bar{0}\rangle$ is some reference state. Let S be the swap operation such that $S|x,y\rangle = |y,x\rangle$, and $\Pi_0 = |\bar{0}\rangle\langle\bar{0}| \otimes I$. Then the unitary defined by

$$V_P = [(2\Pi_0 - I) \otimes I] U_P^\dagger S U_P, \tag{101}$$

is a discrete-time quantum walk on the edges of P. For details on these discrete-time quantum walks, we refer the reader to Refs. [39, 41, 32, 42]. We now describe the spatial search problem, which we shall deal with in the subsequent sections.

Suppose a subset M of the n nodes of P are marked. That is, its state space $X = U \cup M$, where U is the set of unmarked nodes. Then, the spatial search problem can be defined as follows: suppose the random walk starts from the stationary distribution π of P. What is the expected number of steps needed by the random walk to find some node $v \in M$? For random walks (both discrete and continuous-time), this is known as the hitting time (HT). Whether quantum walks can provide a quadratic advantage for the spatial search problem for any P and any number of marked nodes, was open until recently. Ambainis et al. proved that discrete-time quantum walks solve the problem in $\tilde{O}(\sqrt{HT})$ steps, on average [32]. A similar result was shown for continuous-time quantum walks in [33].

Both these quantum algorithms make use of the so-called interpolated Markov chains framework. Let P' be the absorbing Markov chain, obtained from P by replacing all outgoing edges from M with self-loops. Then, the interpolated Markov chain is defined as $P(s) = (1-s)P + sP'$, where $s \in [0,1)$. One can define a Discriminant matrix $D(s)$ for $P(s)$, analogous to P. The relationship between $D(s)$ and $P(s)$ is also analogous to the non-interpolated case. In addition to interpolated Markov chains, the optimal quantum spatial search algorithms in [32] made use of *Standard LCU*-based techniques. Here, we will show that the framework of *Ancilla-free LCU* quite naturally leads to new optimal quantum algorithms for spatial search while saving on the number of ancilla qubits needed. As mentioned previously, we will work with general Hermitian operators (Hamiltonians) and only invoke quantum (or random) walks as particular cases.

7.2 Applying Ancilla-free LCU: Optimal quantum spatial search by fast-forwarding discrete-time random walks

We begin by considering any Hamiltonian H of unit spectral norm. Consider U_H, which is a $(1, a, 0)$-block encoding of H, such that $U_H^2 = I$. Then, it is well known that we can implement a block encoding of H^t using LCU in cost scaling as $O(\sqrt{t})$. This has been implicit in Ref. [26], and also appeared in Ref. [32]. For this, generally one uses the fact that for any x such that $|x| \leq 1$, x^t can be expressed as a linear combination of Chebyshev

polynomials. The following d-degree polynomial $p_{t,d}(x)$, defined as

$$p_{t,d}(x) = \begin{cases} \frac{1}{2^t}\sum_{j=-d/2}^{d/2}\binom{t}{j+t/2}T_{2j}(x) & t,d \text{ are even} \\ \\ \frac{2}{2^t}\sum_{j=0}^{(d-1)/2}\binom{t}{\frac{t+1}{2}+j}T_{2j+1}(x) & t,d \text{ are odd.} \end{cases}, \tag{102}$$

approximates $f(x) = x^t$, for any $t \in \mathbb{Z}$. This has been formally proven in Ref. [84] which we restate here:

Lemma 19. *[84] Suppose $\varepsilon > 0$, $x \in [-1,1]$, $q \geq 1$ and $t \in \mathbb{R}^+$, then there exists a polynomial $p_{t,d}(x)$ of degree $d = \lceil\sqrt{2t\ln(2q/\varepsilon)}\rceil$ such that,*

$$\sup_{x\in[-1,1]}\left|x^t - p_{t,d}(x)\right| \leq 2e^{-d^2/2t} \leq \varepsilon/q.$$

If $R = (2\left|\bar{0}\right\rangle\left\langle\bar{0}\right| - I)\otimes I = 2\Pi_0 - I\otimes I$, is the reflection operator. Then V^t is a $(1,a,0)$-block encoding of $T_t(H)$, where $V = R.U_H$. That is, $\left(\left\langle\bar{0}\right|\otimes I\right)V^t\left(\left|\bar{0}\right\rangle\otimes I\right) = T_t(H)$. Then H^t can be approximated by a linear combination of powers of V. In fact,

$$\left\| H^t - \sum_{\ell=0}^{d/2}\frac{c_\ell}{\|c\|_1}V^\ell \right\| \leq \varepsilon, \tag{103}$$

for $d = \lceil\sqrt{2t\ln(8/\varepsilon)}\rceil$. Here, for even t, the LCU coefficients are defined as

$$c_\ell = \begin{cases} 2^{1-t}\binom{t}{\frac{t}{2}+\ell}, & \ell > 0 \\ 2^{-t}\binom{t}{t/2}, & \ell = 0, \end{cases}, \tag{104}$$

while for odd t,

$$c_\ell = 2^{1-t}\binom{t}{\frac{t+1}{2}+\ell}, \tag{105}$$

The ℓ_1-norm of the LCU coefficients can be easily obtained by observing for $x \in [-1,1]$

$$\|c\|_1 = \left| x^t - \sum_{\ell=d/2+1}^{t} 2^{1-t}\binom{t}{\frac{t}{2}+\ell}\right| \tag{106}$$

$$\geq 1 - \left|\sum_{\ell=d/2+1}^{t} 2^{1-t}\binom{t}{\frac{t}{2}+\ell}\right| \tag{107}$$

$$\geq 1 - \varepsilon/4, \tag{108}$$

for even t, while an analogous bound can also be obtained for odd t. Now, implementing the linear combination of powers of V in Eq. (103) via *Standard LCU* requires $O(a+\log t+\log\log(1/\varepsilon))$ ancilla qubits, and $O(\sqrt{t\ln(1/\varepsilon)})$ cost, which has been used in quantum fast-forwarding [31, 32].

Now suppose we are concerned about the average projection of H^t in some subspace of interest. Then by *Ancilla-free LCU*, we do not need the additional $O(\log t + \log\log(1/\varepsilon))$

ancilla qubits (See Theorem 2): we can simply sample some ℓ according to $c_\ell/\|c\|_1$ and apply V^ℓ to the initial state. On average, the projection of the resulting density matrix would be at least as large as the projection of H^t. This is what we show next.

Fast-forwarding by Ancilla-free LCU: Given access to U_H, which is a $(1, a, \delta)$-block encoding of some Hamiltonian H, we want to prepare a state whose projection on to the subspace of interest (spanned by Π) is at least $\left\|\Pi H^t |\psi_0\rangle\right\|^2$, on average. In such a scenario, we can implement *Ancilla-free LCU* by replacing *Standard LCU* with importance sampling. Consider Algorithm 3, where $V = R.U_H$. If the initial state is $\rho_0 = |\psi_0\rangle\langle\psi_0|$, Algorithm 3 outputs a density matrix, which, on average, is

$$\bar{\rho} = \sum_{\ell=0}^{d/2} \frac{c_\ell}{\|c\|_1} V^{2\ell} \rho_0 V^{-2\ell}, \tag{109}$$

if t is even (an analogous expression is obtained when t is odd). Then, we can use

Algorithm 3: POW-HAM$(t, d, V, |\psi_0\rangle)$

Inputs: The unitary V, an initial state $|\psi_0\rangle$ and parameters $t \in \mathbb{R}^+$ and $d \in \mathbb{N}$.

1. If t is even,
   (a) Pick $\ell \in [0, d/2]$ according to $c_\ell/\|c\|_1$, where $c_\ell = 2^{1-t}\binom{t}{\frac{t}{2}+\ell}$.
   (b) Apply $V^{2\ell}$ to $|\psi_0\rangle$.

2. If t is odd,
   (a) Pick $\ell \in \left[0, \frac{d-1}{2}\right]$ according to $c_\ell/\|c\|_1$, where $c_\ell = 2^{1-t}\binom{t}{\frac{t+1}{2}+\ell}$.
   (b) Apply $V^{2\ell+1}$ to $|\psi_0\rangle$.

Theorem 2 to prove that $\mathrm{Tr}[\Pi\bar{\rho}] \geq \mathrm{Tr}[\Pi H^t \rho_0 H^t] - \varepsilon$. However, there are still issues to consider before we can do so. For instance, U_H is not a perfect block encoding of H. How should the precision in block encoding, δ, scale for this to hold? We formally state this, and prove the algorithmic correctness via the following Theorem:

Lemma 20. *Suppose $\varepsilon \in (0, 1)$ and we have access to U_H, which is a $(1, a, \delta)$-block encoding of a Hamiltonian H such that $\|H\| = 1$ and $U_H^2 = I$. Then, provided $d = \lceil\sqrt{2t\ln(24/\varepsilon)}\rceil$ and,*

$$\delta \leq \frac{\varepsilon^2}{1152\ t\ln(24/\varepsilon)},$$

for any $t \in \mathbb{R}^+$, initial state $\rho_0 = |\psi_0\rangle\langle\psi_0|$, and projector Π on to the space of ρ_0, then the sampling in Algorithm 3 prepares some density matrix ρ such that $\mathbb{E}[\rho] = \bar{\rho}$, where $\bar{\rho}$ is defined in Eq. (109), and satisfies

$$\mathrm{Tr}[\Pi\bar{\rho}] \geq \mathrm{Tr}[\Pi H^t \rho_0 H^t] - \varepsilon,$$

using $O\left(\sqrt{t\log(1/\varepsilon)}\right)$ queries to $V = R.U_H$.

Proof. Let H' be a $(1, a, 0)$-block encoding of U_H. Then, by definition $\|H - H'\| \leq \delta$. Let us choose the degree of the polynomial $p_{t,d}(H')$ to be $d = \lceil \sqrt{2t \ln(24/\varepsilon)} \rceil$, which ensures that $\left\| x^t - p_{t,d}(x) \right\| \leq \varepsilon/12$ (from Lemma 19).

Now, the *Standard LCU* procedure would implement the state

$$|\psi_t\rangle = |\bar{0}\rangle \frac{p_{t,d}(H')}{\|c\|_1} |\psi_0\rangle + |\Phi\rangle^{\perp}.$$

From the choice of d, we ensure that $\alpha = \|c\|_1 \geq 1 - \varepsilon/12$. Also,

$$\begin{aligned}
\left\| H^t - p_{t,d}(H')/\alpha \right\| &\leq \left\| H^t - p_{t,d}(H') \right\| + (1-\alpha) \left\| p_{t,d}(H')/\alpha \right\| && (110)\\
&\leq \varepsilon/12 + \left\| H^t - p_{t,d}(H) \right\| + \left\| p_{t,d}(H) - p_{t,d}(H') \right\| && (111)\\
&\leq \varepsilon/12 + \varepsilon/12 + 4d\sqrt{\delta} && [\text{ From Lemma 5 }] \quad (112)\\
&\leq \varepsilon/6 + \varepsilon/6 && \left[\text{As } \delta \leq \frac{\varepsilon^2}{576 d^2} \right] \quad (113)\\
&\leq \varepsilon/3. && (114)
\end{aligned}$$

We will now use Theorem 2, which ensures that the average density matrix $\bar{\rho}$ from Algorithm 3, satisfies

$$\begin{aligned}
\mathrm{Tr}[(I \otimes \Pi)\bar{\rho}] &= \mathrm{Tr}[(I \otimes \Pi) |\psi_t\rangle \langle\psi_t|] && (115)\\
&\geq \frac{1}{\|c\|_1^2} \left[\mathrm{Tr}[\Pi H^t \rho_0 H^t] - \varepsilon \right] && (116)\\
&\geq \mathrm{Tr}[\Pi H^t \rho_0 H^t] - \varepsilon. && (117)
\end{aligned}$$

□

Thus, on average, the projection of the density matrix prepared by Algorithm 3 on to the space spanned by Π is at least as large as $\mathrm{Tr}[\Pi H^t \rho_0 H^t]$. Furthermore, Lemma 20 shows that the cost is $O(\sqrt{t \ln(1/\varepsilon)})$. Interestingly this procedure does not need the extra $O(\log t + \log\log(1/\varepsilon))$ ancilla qubits that *Standard LCU* does.

Let us now discuss why this is important in the context of spatial search by quantum walk. Notice that for any ergodic, reversible Markov chain, the discrete-time quantum walk unitary $U_D = U_P^{\dagger} S U_P$ (defined in Sec. 7.1) is a block encoding of the discriminant matrix D of P, i.e.

$$\left(\langle\bar{0}| \otimes I \right) U_P^{\dagger} S U_P \left(|\bar{0}\rangle \otimes I \right) = D.$$

Thus U_D is a $(1, \lceil \log_2 n \rceil, 0)$-block encoding of D, where $a = |X| = n$. Thus, *Standard LCU* prepares a quantum state that is $O(\varepsilon)$ - close to

$$|\psi_t\rangle = |\bar{0}\rangle D^t |\psi_0\rangle + |\Phi\rangle^{\perp},$$

in cost $O\left(\sqrt{t \log(1/\varepsilon)}\right)$, using $O(\log n + \log t + \log\log(1/\varepsilon))$ ancilla qubits. This is the essence of quantum fast-forwarding: the fact that we can implement t-steps of a random walk on D quadratically faster using quantum walks (up to normalization) [31].

On the other hand, suppose for a specific problem (such as quantum spatial search) it suffices to ensure that the projection of $D^t |\psi_0\rangle$ on some subspace of interest is high on

average. Then from Algorithm 3 we can prepare a quantum state whose projection, on average, is at least as high as the projection of $D^t |\psi_0\rangle$. For this, we can simply apply k steps of the quantum walk operator $V = R.U_D$ for some $k \in [0, \sqrt{2t \ln(24/\varepsilon)}]$ steps, sampled at random according to $\{c_k/\|c\|_1\}$. Indeed Algorithm 3, on average, prepares the density matrix $\bar{\rho}$ such that from Lemma 20,

$$\mathrm{Tr}[\Pi\bar{\rho}] \geq \mathrm{Tr}[\Pi D^t \rho_0 D^t] - \varepsilon.$$

Overall, quantum fast forwarding by *Ancilla-free LCU* requires only $O(\log n)$ ancilla qubits, which is the quantum walk space (edges of P). On the other hand, quantum fast-forwarding by *Standard LCU* requires an additional $O(\log t + \log\log(1/\varepsilon))$ ancilla qubits.

Now we are in a position to discuss the spatial search algorithm by discrete-time quantum walk which makes use of Algorithm 3. Now for the spatial search problem, we are interested in showing that on average, the projection of the state $D^t |\sqrt{\pi}\rangle$ in the marked subspace is high, for $t = O(HT)$. As a result, we can drop the ancilla register and simply apply the *Ancilla-free LCU* technique to implement the unitary V for a random number of steps, sampled according to the distribution of the LCU coefficients.

Method 1 – Spatial search by discrete-time quantum walk: Suppose there exists an ergodic, reversible Markov chain P with state space X and $|X| = n$. Let $M \subset X$ be a set of marked nodes. Classically, the spatial search algorithm boils down to applying the so-called, absorbing Markov chain P' (obtained from P by replacing the outgoing edges from M with self-loops) to the stationary distribution of P (say π). At every step, one checks to see if the vertex obtained is marked. The expected number of steps needed to find some $x \in M$ is known as the hitting time, denoted by HT. Consider the interpolated Markov chain $P(s) = (1-s)P + sP'$, where $s \in [0,1]$ and $D(P(s))$ be the corresponding discriminant matrix. Then $U_{D(s)}$ is a $(1, a, 0)$-block encoding of $D(s)$, where $a = \lceil \log_2(n) \rceil$.

First, let us look at the quantum spatial search algorithm of Ref. [32]. Starting from the state $|\sqrt{\pi}\rangle = \sum_x \sqrt{\pi_x} |x\rangle$ (which corresponds to a quantum encoding of the stationary distribution of P), the first step involves preparing $|\bar{0}\rangle |\sqrt{\pi_U}\rangle = |\bar{0}\rangle \sum_{y \in X \setminus M} \sqrt{\pi_y} |y\rangle$, which is simply the support of $|\sqrt{\pi}\rangle$ in the unmarked subspace. This is quite easy: if Π_M is the projector on to the marked subspace, simply measure $|\sqrt{\pi}\rangle$ in the basis $\{\Pi_M, I - \Pi_M\}$. Then $|\sqrt{\pi_U}\rangle$ is prepared if the outcome of the measurement is an unmarked vertex. Following this, the algorithm of [32] prepares the state $D(s)^t |\sqrt{\pi_U}\rangle$ using quantum fast-forwarding [31] (via *Standard LCU*) for some randomly chosen values of s, and $t \in \Theta(HT)$. Then the central result of Ref. [32] was to prove, using an involved combinatorial lemma, that the projection of this state on to the marked subspace is $\widetilde{\Omega}(1)$ on average for these choices of s and t. More precisely, the authors prove that if we choose parameters $s \in \{1 - 1/r : r = 1, 2, \cdots, 2^{\lceil \log T \rceil}\}$ and $T \in \Theta(HT)$ uniformly at random,

$$\mathbb{E}\left[\left\|\Pi_M D(s)^T |\sqrt{\pi_U}\rangle\right\|^2\right] \in \Omega\left(1/\log T\right).$$

This is optimal as it provides a generic quadratic speedup over classical random walks (up to a log factor) for any reversible P and marked subspace M of any cardinality, unlike prior works. However, the algorithm uses quantum fast forwarding by *Standard LCU*, which requires $O(\log HT)$ ancilla qubits to implement.

Note that in this case, it suffices to ensure that the projection of $D(s)^t |\sqrt{\pi_U}\rangle$ in the marked subspace is large on average for the aforementioned choices of s and t. Then from Lemma 20, we can replace the quantum fast forwarding by *Standard LCU* of Ref. [32],

with *Ancilla-free LCU*: Apply k steps of the quantum walk operator $V(s) = R.U_{D(s)}$, where $k \in [0, O(\sqrt{HT})]$ is chosen according to $\{c_k/\|c\|_1\}$, as stated in Algorithm 3. This prepares, on average, the density matrix $\bar{\rho}$ whose projection on to the marked subspace is at least $\Omega\left(1/\log T\right)$.

The quantum spatial search algorithm, after incorporating Algorithm 3 as a subroutine, is stated via Algorithm 4. The key difference as compared to the algorithm of [32] is that

Algorithm 4: `QSpatial Search - 1` Spatial search by DTQW

1. Pick t uniformly at random from $[0, T]$, where $T = \Theta\left(HT\right)$ and set $d = \lceil \sqrt{T \log(T)} \rceil$.
2. Set $s = 1 - 1/r$, where r is picked uniformly at random from $R = \{2^0, 2^1, \cdots, 2^{\lceil \log T \rceil}\}$.
3. Construct $U_{D(s)}$, which is a $(1, \lceil \log_2(|X|) \rceil, 0)$-block encoding of $D(s)$.
4. Prepare the quantum state $|\bar{0}\rangle |\sqrt{\pi}\rangle = |\bar{0}\rangle \sum_{y \in X} \sqrt{\pi_y} |y\rangle$.
5. Measure in the basis $\{\Pi_M, I - \Pi_M\}$ in the second register. If the output is marked, measure in the node basis to output some $x \in M$. Otherwise, we are in the state $|\bar{0}\rangle |\sqrt{\pi_U}\rangle = |\bar{0}\rangle \sum_{y \in X \setminus M} \sqrt{\pi_y} |y\rangle$.
6. Call `POW-HAM`$(t, d, V = R.U_{D(s)}, |\sqrt{\pi_U}\rangle)$.
7. Measure the first register in the basis of the nodes of P.

our procedure does not require quantum fast-forwarding by *Standard LCU*: it suffices to call Algorithm 3 instead. This ensures that our method requires $O(\log HT)$ fewer ancilla qubits.

From Lemma 20, we obtain that, Algorithm 4 outputs a density matrix, which on average, has a projection on to the marked subspace, lower bounded as

$$\mathrm{Tr}[(I \otimes \Pi_M)\bar{\rho}] \geq \left\| \Pi_M D(s)^T |\sqrt{\pi_U}\rangle \right\|^2 - \varepsilon, \tag{118}$$

for a small enough $\varepsilon \in \Theta(1/\log(T))$. Then, the combinatorial lemma of Ref. [32] ensures that expected value of the RHS of the aforementioned equation is $\widetilde{\Omega}(1)$. We refer the readers to [32] for the proof of this lemma.

What Equation (118) tells us is that if we run Algorithm 4, and measure the second register in the vertex basis, the probability of finding a marked element would be at least $\Omega(1/\log^2 T)$, on average, after $O(\sqrt{T \log T})$ DTQW steps (where $T = \Theta(HT)$). Thus, just as Ref. [32], this algorithm also yields a quadratic improvement over its classical counterpart (up to a log factor). However, it requires $O(\log HT))$ fewer ancilla qubits.

We shall use similar ideas to develop an alternative quantum algorithm for spatial search by discrete-time quantum walk.

7.3 Applying Ancilla-free LCU: Optimal quantum spatial search by fast-forwarding continuous-time random walks

We develop a quantum algorithm for fast-forwarding the dynamics of a continuous-time random walk. Given the discriminant matrix D, a continuous-time random walk is defined

by the operator e^{D-I}, where $Q = D - I$ is the continuous-time random walk kernel. We first show that a block encoding of $e^{t(D-I)}$ can be implemented in cost $O(\sqrt{t}\log 1/\varepsilon)$, using *Standard LCU*. This requires $O(\log n + \log t + \log\log(1/\varepsilon))$ ancilla qubits. Next, we demonstrate that the fast forwarding can be leveraged to develop an optimal quantum spatial search algorithm: requiring $\widetilde{O}(\sqrt{HT})$ steps on average to find marked vertices on any reversible P with M being the set of marked nodes. As this relies on the *Standard LCU* technique, the overall algorithm requires $O(\log n + \log HT)$ ancilla qubits, analogous to [32].

Finally, similar to the previous section, for optimal quantum spatial search, it suffices to ensure that the projection of $e^{t(D-I)}\,|\sqrt{\pi_U}\rangle$ in the marked subspace is high, on average after $t = O(HT)$ steps of the continuous-time random walk. Thus, we can bypass quantum fast forwarding by *Standard LCU*, and use *Ancilla-free LCU*, instead. This only requires $O(\log n)$ ancilla qubits required to implement the quantum walk on the edges of P, saving on $O(\log HT)$ ancilla qubits overall. As in the previous section, we will work with general Hamiltonians and discuss quantum walks as a particular case

Now the polynomial approximation to x^t can be used to obtain the following low degree polynomial that approximates $e^{-t(1-x)}$ [44]:

$$q_{t,d,d'}(x) = e^{-t}\sum_{j=0}^{d}\frac{t^j}{j!}p_{j,d'}(x).$$

This has degree

$$d' = \lceil\sqrt{2d\ln(4/\varepsilon)}\rceil \in O\left(\sqrt{t}\log(1/\varepsilon)\right),$$

for $d = \lceil\max\{te^2, \ln(2/\varepsilon)\}\rceil$. Indeed, it can be shown that

$$\sup_{x\in[-1,1]}\left|e^{-t(1-x)} - q_{t,d,d'}(x)\right| \leq \varepsilon. \tag{119}$$

Thus, given a block encoding of any Hermitian matrix H, with unit spectral norm, the operator $e^{-t(I-H)}$ can be implemented as a linear combination of unitaries. This is because the d'-degree polynomial $q_{t,d,d'}(x)$ approximates $e^{-t(I-H)}$ and is a linear combination of the d'-degree polynomial $p_{j,d'}(x)$. So overall, by LCU, we can implement the polynomial $q_{t,d,d'}(x)$, approximating $e^{-t(1-x)}$. We formally show this via the following lemma:

Lemma 21. *Suppose $\varepsilon \in (0,1)$ and we have access to U_H, which is a $(1,a,\delta)$-block encoding of a Hamiltonian H such that $\|H\| = 1$ and $U_H^2 = I$. Furthermore, let $d = \lceil\max\{te^2, \ln(8/\varepsilon)\}\rceil$ and $d' = \sqrt{2d\ln(16/\varepsilon)}$. Then, provided*

$$\delta \leq \frac{\varepsilon^2}{128d\ \ln(16/\varepsilon)},$$

for any $t \in \mathbb{N}$, we can implement a $(1, O(a + \log t + \log\log(1/\varepsilon)), \varepsilon)$-block encoding of $e^{-t(I-H)}$ in cost $O\left(\sqrt{t}\log(1/\varepsilon)\right)$.

Proof. Let U_H be a $(1,a,0)$-block encoding of H'. By definition, $\|H - H'\| \leq \delta$. For the polynomial $q_{t,d,d'}(x)$, we choose $d = \lceil\max\{te^2, \ln(8/\varepsilon)\}\rceil$ and $d' = \sqrt{2d\ln(16/\varepsilon)}$. This ensures $\left\|e^{-t(1-x)} - q_{t,d,d'}(x)\right\| \leq \varepsilon/4$.

As before, let $\widetilde{W}$ be the unitary that implements the LCU. Then,

$$\left(\langle\bar{0}|\otimes I\right)\widetilde{W}\left(|\bar{0}\rangle\otimes I\right) = \frac{q_{t,d,d'}(H')}{\|c\|_1}. \tag{120}$$

For the choice of d, d' we have

$$\|c\|_1 = e^{-t} \sum_{j=0}^{d} \frac{t^j}{j!} p_{j,d'}(H) \tag{121}$$

$$\geq e^{-t} \sum_{j=0}^{d} \frac{t^j}{j!} (1 - \varepsilon/8) \qquad [\text{ As } d' = \lceil \sqrt{2d \ln(16/\varepsilon)} \rceil \text{ }] \tag{122}$$

$$\geq \left(1 - e^{-t} \sum_{j=d+1}^{\infty} \frac{t^j}{j!}\right) (1 - \varepsilon/8) \tag{123}$$

$$\geq (1 - \varepsilon/8)(1 - \varepsilon/8) \tag{124}$$

$$\geq 1 - \varepsilon/4. \tag{125}$$

In order to go from the third to the fourth line we have used the fact that by Stirling's approximation:

$$e^{-t} \sum_{j=d+1}^{\infty} \frac{t^j}{j!} \leq e^{-t} \sum_{j=d+1}^{\infty} \left(\frac{te}{j}\right)^j \tag{126}$$

$$\leq e^{-t} \sum_{j=d+1}^{\infty} e^{-j} \left(\frac{te^2}{j}\right)^j \tag{127}$$

$$\leq e^{-t} \sum_{j=d+1}^{\infty} e^{-j} \qquad [\text{ As } te^2/j \leq 1,\ \forall j \geq d+1 \text{ }] \tag{128}$$

$$\leq e^{-t-d} \leq \varepsilon/8, \tag{129}$$

for $d = \lceil \ln(8/\varepsilon) \rceil$.

This implies $\|c\|_1 \in [1 - \varepsilon/4, 1]$. Now, we will show that $\widetilde{W}$ indeed implements a block encoding of $e^{-t(I-H)}$.

$$\left\| e^{-t(I-H)} - \frac{q_{t,d,d'}(H')}{\|c\|_1} \right\| \leq \left\| e^{-t(I-H)} - q_{t,d,d}(H) \right\| + \left\| e^{-t(I-H)} - q_{t,d,d}(H) \right\| + (1 - \|c\|_1) \tag{130}$$

$$\leq \varepsilon/4 + \varepsilon/4 + 4d'\sqrt{\delta} \qquad [\text{From Lemma 5}] \tag{131}$$

$$\leq \varepsilon/2 + \varepsilon/2 \qquad \left[\text{ As } \delta \leq \frac{\varepsilon^2}{64 d'^2}\right] \tag{132}$$

□

It is easy to see that this leads to the fast-forwarding of continuous-time random walks. The unitary $V = U_P^\dagger S U_P$ is a $(1, \lceil \log n \rceil, 0)$-block encoding of the random walk discriminant matrix D. Then by using Lemma 21, given an initial state $|\psi_0\rangle$, we can prepare a quantum state that is $O\left(\varepsilon \cdot \left\| e^{-(I-D)t} |\psi_0\rangle \right\|\right)$-close to

$$|\psi_t\rangle = |0\rangle \frac{e^{-t(I-D)} |\psi_0\rangle}{\left\| e^{-t(I-D)} |\psi_0\rangle \right\|} + |\psi^\perp\rangle, \tag{133}$$

with success probability $\Theta\left(\left\| e^{-(I-D)t} |\psi_0\rangle \right\|^2\right)$, in cost $O\left(\sqrt{t} \log\left(\varepsilon^{-1} \cdot \left\| e^{-(I-D)t} |\psi_0\rangle \right\|^{-1}\right)\right)$.

Finally, by applying $O\left(\left\|e^{-(I-D)t}\left|\psi_0\right\rangle\right\|^{-1}\right)$ rounds of amplitude amplification, we prepare, with $\Omega(1)$ probability, a quantum state that is $O(\varepsilon)$-close to $|\psi_t\rangle$ in cost

$$T = O\left(\frac{\sqrt{t}}{\left\|e^{-(I-D)t}\left|\psi_0\right\rangle\right\|}\log\left(\frac{1}{\varepsilon\left\|e^{-(I-D)t}\left|\psi_0\right\rangle\right\|}\right)\right).$$

Other than the walk space of $O(\log n)$ qubits (edges of the Markov chain), fast forwarding continuous-time random walks by *Standard LCU* additionally requires $O(\log t + \log\log(1/\varepsilon))$ ancilla qubits. As before, for the spatial search algorithm, we can drop these ancilla qubits completely and implement *Ancilla-free LCU* instead.

Fast-forwarding by Ancilla-free LCU: For the spatial search problem, we are concerned about the average projection of $e^{-t(I-H)}\left|\psi_0\right\rangle$, on to the marked subspace. Thus, we can apply *Ancilla-free LCU* instead. The overall procedure is outlined in Algorithm 5. We will be implementing $q_{t,d,d'}(H)$ which is itself a linear combination of $p_{t,d'}(H)$. So, Algorithm 5 also calls Algorithm 3 as a subroutine.

Algorithm 5: `EXP-HAM`$(t, d', d, V, |\psi_0\rangle)$

Inputs: A unitary V, $t \in \mathbb{R}^+$, $d, d' \in \mathbb{N}$, and an initial state $|\psi_0\rangle$.

1. Pick some integer $\ell \in [0, d]$ according to $c_\ell / \|c\|_1$, where $c_\ell = \dfrac{e^{-t}t^\ell}{\ell!}$
2. Call `POW-HAM`$(\ell, d', V, |\psi_0\rangle)$.

If $\rho_0 = |\psi_0\rangle\langle\psi_0|$, for $d, d' \in \mathbb{N}$, Algorithm 5, on average, prepares the density matrix

$$\bar{\rho} = e^{-t}\sum_{j=0}^{d}\frac{t^j}{j!}\left[\sum_{j\in\text{Even},k=0}^{d'/2} 2^{1-j}\binom{j}{j+k/2}V^{2k}\rho_0 V^{-2k} + \sum_{j\in\text{Odd},k=0}^{(d'-1)/2} 2^{1-j}\binom{j}{(j+1)/2+k}V^{2k+1}\rho_0 V^{-(2k+1)}\right], \tag{134}$$

where $V = R \cdot U_H$ is the quantum walk operator. On average $O(d')$ queries are made to V. However, in order to ensure that $\bar{\rho}$ satisfies

$$\mathrm{Tr}[\Pi\bar{\rho}] \geq \mathrm{Tr}[\Pi e^{-t(I-H)}\rho_0 e^{-t(I-H)}] - \varepsilon,$$

the appropriate values of δ, d, d' need to be chosen. We determine these via the following lemma:

Lemma 22. *Suppose $\varepsilon \in (0,1)$ and we have access to U_H, which is a $(1, a, \delta)$-block encoding of a Hamiltonian H such that $\|H\| = 1$ and $U_H^2 = I$. Then, provided $d = \lceil\max\{te^2, \ln(12/\varepsilon)\}\rceil$, $d' = \lceil\sqrt{2t\ln(48/\varepsilon)}\rceil$ and,*

$$\delta \leq \frac{\varepsilon^2}{1152\ d\ln(48/\varepsilon)},$$

for any $t \in \mathbb{R}^+$, projector Π and initial state $\rho_0 = |\psi_0\rangle\langle\psi_0|$, then Algorithm 5 prepares, on average, the density matrix $\bar{\rho}$ such that

$$\mathrm{Tr}[\Pi\bar{\rho}] \geq \mathrm{Tr}[\Pi e^{-t(I-H)}\rho_0 e^{-t(I-H)}] - \varepsilon,$$

using $O\left(\sqrt{t}\log(1/\varepsilon)\right)$ queries to $V = R.U_H$.

Proof. Let H' be a $(1, a, 0)$ block encoding of U_H. Then, by definition $\|H - H'\| \leq \delta$. By choosing the degree of the polynomial $q_{t,d,d'}(H')$ to be $d' = \lceil\sqrt{2d\ln(48/\varepsilon)}\rceil$, where $d = \lceil\max\{te^2, \ln(12/\varepsilon)\}\rceil$, ensures that $\left\|e^{-t(1-x)} - q_{t,d,d'}(x)\right\| \leq \varepsilon/12$.

Now, from Lemma 21, the full LCU procedure would implement the state

$$|\psi_t\rangle = |\bar{0}\rangle \frac{q_{t,d,d'}(H')}{\|c\|_1}|\psi_0\rangle + |\Phi\rangle^{\perp}.$$

Now, from the choice of d', we ensure that $\|c\|_1 \geq 1 - \varepsilon/12$. Also,

$$\left\|e^{-t(I-H)} - q_{t,d,d'}(H')/\|c\|_1\right\| \leq \left\|e^{-t(I-H)} - q_{t,d,d'}(H')\right\| + (1 - \|c\|_1)\left\|q_{t,d,d'}(H')/\|c\|_1\right\| \tag{135}$$

$$\leq \varepsilon/12 + \left\|e^{-t(I-H)} - q_{t,d,d'}(H)\right\| + \left\|q_{t,d,d'}(H) - q_{t,d,d'}(H')\right\| \tag{136}$$

$$\leq \varepsilon/12 + \varepsilon/12 + 4d'\sqrt{\delta} \qquad [\text{ From Lemma 5 }] \tag{137}$$

$$\leq \varepsilon/6 + \varepsilon/6 \qquad \left[\text{ As } \delta \leq \frac{\varepsilon^2}{576d'^2}\right] \tag{138}$$

$$\leq \varepsilon/3. \tag{139}$$

We will now use Theorem 2, which ensures that Algorithm 5, on average, prepares $\bar{\rho}$ such that

$$\mathrm{Tr}[\Pi\bar{\rho}] = \mathrm{Tr}[(I \otimes \Pi)|\psi_t\rangle\langle\psi_t|] \geq \frac{1}{\|c\|_1^2}\left[\mathrm{Tr}[\Pi e^{-t(I-H)}\rho_0 e^{-t(I-H)}] - \varepsilon\right] \tag{140}$$

$$\geq \mathrm{Tr}[\Pi e^{-t(I-H)}\rho_0 e^{-t(I-H)}] - \varepsilon. \tag{141}$$

□

Thus, if it suffices to ensure that the projection on to the subspace of interest is high on average, we can replace *Standard LCU* with *Ancilla-free LCU*, thereby saving on the $O(\log t + \log\log(1/\varepsilon))$ ancilla qubits. Finally, we apply this lemma to develop a new quantum algorithm for spatial search by discrete-time quantum walk, one that relies on quantum fast forwarding of continuous-time random walks.

Method 2 – Spatial search by discrete-time quantum walk: The ability to fast-forward continuous-time random walks imply that we can design an alternate quantum spatial search algorithm by discrete-time quantum walks.

Suppose we consider an interpolated Markov chain $P(s) = (1 - s)P + sP'$, where $s \in [0, 1]$, and P' corresponds to the absorbing Markov chain. Suppose $D(s)$ denotes the Discriminant matrix of $P(s)$ and we have access to $U_{D(s)}$, which is a $(1, a, 0)$-block

Algorithm 6: `QSpatial Search - 2` Alternative Spatial search by DTQW

1. Pick t uniformly at random from $[0, T]$, where $T = \Theta(HT)$. Set $d = \lceil Te^2 \rceil$ and $d' = \Theta(\sqrt{T \log T})$.
2. Set $s = 1 - 1/r$, where r is picked uniformly at random from $R = \{2^0, 2^1, \cdots, 2^{\lceil \log T \rceil}\}$.
3. Construct $U_{D(s)}$, which is a $(1, \lceil \log_2(|X|) \rceil, 0)$-block encoding of $D(s)$.
4. Prepare the quantum state $|\bar{0}\rangle |\sqrt{\pi}\rangle = |\bar{0}\rangle \sum_{y \in X} \sqrt{\pi_y} |y\rangle$.
5. Measure in the basis $\{\Pi_M, I - \Pi_M\}$ in the second register. If the output is marked, measure in the node basis to output some $x \in M$.
   Otherwise, we are in the state $|\bar{0}\rangle |\sqrt{\pi_U}\rangle = |\bar{0}\rangle \sum_{y \in X \setminus M} \sqrt{\pi_y} |y\rangle$.
6. Call `EXP-HAM`$(t, d', d, V = R.U_{D(s)}, |\sqrt{\pi_U}\rangle)$.
7. Measure the resulting state in the node basis in the first register.

encoding of $D(s)$, such that $U_{D(s)}^2 = I$. Consider Algorithm 6, which is very similar to the first spatial search algorithm except that it calls Algorithm 5 as a subroutine.

The output of Algorithm 6, on average, is the density matrix $\bar{\rho}$ such that

$$\mathrm{Tr}[(I \otimes \Pi_M)\bar{\rho}] \geq \left\| \Pi_M e^{T(D(s)-I)} |\sqrt{\pi_U}\rangle \right\|^2 - \varepsilon,$$

for small enough $\varepsilon \in \Theta(1/\log^2(T))$. It remains to show that Algorithm 6 succeeds on average with probability $\tilde{\Omega}(1)$, for appropriate choices of s and t. Indeed, we show that for the same choices of s and t as in Algorithm 4, the average success probability is high. We demonstrate this via the following lemma.

Lemma 23. *Consider an ergodic, reversible Markov chain P and a set of marked nodes M. If we choose parameters $s \in \{1 - 1/r : r = 1, 2, \cdots, 2^{\lceil \log T \rceil}\}$ and $T \in \Theta(HT)$ uniformly at random, then the the following holds*

$$\mathbb{E}\left[\left\| \Pi_M e^{(D(s)-I)T} |\sqrt{\pi_U}\rangle \right\|^2 \right] \in \Omega\left(1/\log^2 T\right).$$

Proof sketch: We only provide a sketch of the proof here, while we refer the readers to the proof of Lemma S5 in the Supplemental Material of Ref. [33]. The full derivation of this lemma can be obtained from the proof therein. The key idea is to show that the quantity we intend to estimate, i.e. $\left\| \Pi_M e^{(D(s)-I)T} |\pi_U\rangle \right\|^2$ is related to the behaviour of the original Markov chain $P(s)$ (which applies to any reversible Markov chain).

– The first step is to show that $\left\| \Pi_M e^{(D(s)-I)t} |\pi_U\rangle \right\|^2$ is lower bounded by the probability of the following event occurring in a continuous-time Markov chain, for any $t \geq 0$: starting from a distribution over the unmarked elements, a continuous-time random walk $X(s)$ is at some marked vertex after time t and is at an unmarked vertex after time $t + t'$, where $t' > 0$. Let us call this event $\mathcal{E}_{X(s)}$.

– The next step is to then show that the probability of this event occurring on a continuous-time Markov chain is lower bounded by the same event (say $\mathcal{E}_{Y(s)}$) happening in a discrete-time Markov chain $Y(s)$.

– So, by these two steps we have related the quantity $\left\|\Pi_M e^{(D(s)-I)T}\left|\pi_U\right\rangle\right\|^2$ to the probability of a specific event occurring on a discrete-time Markov chain. At this stage, we can make use of the combinatorial lemma of Ambainis et al. [32], wherein the authors proved that for any reversible Markov chain P, the probability of the event $\mathcal{E}_{Y(s)}$ occurring is $\widetilde{\Omega}(1)$, on average which allows us to prove

$$\mathbb{E}\left[\left\|\Pi_M e^{(D(s)-I)T}\left|\pi_U\right\rangle\right\|^2\right] = \tilde{\Omega}(1).$$

□

Thus, we managed to develop an optimal quantum spatial search algorithm that relies on the fast forwarding of continuous-time random walks. Moreover, by taking advantage of *Ancilla-free LCU* we require $O(\log HT)$ fewer ancilla qubits, as compared to *Standard LCU*. The recently developed spatial search algorithm by continuous-time quantum walk also falls under the framework of *Ancilla-free LCU*, as it bypasses the need to prepare the (continuous-variable) ancilla register in the Gaussian state. We refer the readers to [33] for the details. Overall, the *Ancilla-free LCU* framework is applicable for quantum spatial search. It allows to retain a generic quadratic speedup over classical random walks, while using no ancilla qubits (other than the space of the quantum walk). In comparison, *Standard LCU* requires $O(\log HT)$ ancilla qubits which are used to implement multi-qubit controlled operations. More broadly, *Ancilla-free LCU* also helped establish a connection between discrete and continuous-time quantum walks with their classical counterparts. In addition, in the Appendix (Sec. E), using the frameworks of block encoding and QSVT, we show that one can obtain a discrete-time quantum walk from a continuous-time quantum walk (and vice versa). Together, it connects both continuous-time quantum and random walks with discrete-time quantum and random walks, which has been shown in Fig. 3.

8 Discussion

We considered the framework of Linear Combination of Unitaries, a quantum algorithmic paradigm that has been used to develop several quantum algorithms of practical interest. However, standard techniques to implement LCU require several ancilla qubits, a sequence of multi-qubit controlled operations, and hence are only implementable using fully fault-tolerant quantum computers, which are perhaps decades away. In this work, we significantly reduce the resources required to implement any LCU. Our motivation was to explore whether a broadly applicable framework such as LCU can be implemented on quantum devices that do not have the capabilities of a fully fault tolerant quantum computer. To this end, we provided three new approaches for LCU, considering the different intermediate-term hardware possibilities.

The *Single-Ancilla LCU* makes repeated use of a short-depth quantum circuit and only a single ancilla qubit, to estimate the expectation value

$$\mathrm{Tr}[O\rho] = \mathrm{Tr}[O\ f(H)\rho_0 f(H)^\dagger]/\mathrm{Tr}[f(H)\rho_0 f(H)^\dagger],$$

where $f(H)$ can be well-approximated by a linear combination of unitaries, O is any observable and ρ_0 is the initial state. The cost of each coherent run of the generic procedure

is always lower than *Standard LCU*. More precisely, we show that the average cost of each run of our procedure is $\langle\tau\rangle$ which is upper bounded by the cost of implementing the most expensive unitary U_j, $O(\tau_{\max})$. For *Standard LCU*, the cost of each coherent run is $\tau_Q + 2\tau_R$, with τ_R and τ_Q being the cost of implementing the *prepare* and *select* unitaries, respectively. We showed that $\tau_Q \geq \tau_{\max}$, and the cost of each run of our method is lower, i.e. $\langle\tau\rangle < O(\tau_R + \tau_Q)$. Interestingly, in the worst case, $\tau_Q = O(M\tau_{\max})$. This occurs when each of the M unitaries in the LCU description costs $\tau_{\max}$, indicating the possibility of a significant separation between the cost of each run of *Standard LCU* with our method.

However, for our applications to ground state property estimation and quantum linear systems, $\tau_Q = O(\tau_{\max})$, whenever *Standard LCU* estimates $\mathrm{Tr}[O\rho]$ without using amplitude amplification and estimation. In this regard, it would be interesting to find an application where τ_Q is significantly larger than $\tau_{\max}$. The cost of each coherent run of *Single-Ancilla LCU* would be significantly lower than *Standard LCU* in such a scenario. Another direction of future research would be to apply our algorithms for Hamiltonian simulation and ground state property estimation to specific Hamiltonians in quantum chemistry [54, 85, 86] and condensed matter physics [87], and benchmark their performance against other near/intermediate-term quantum algorithms such as those making use of the Hadamard test [11, 12, 13].

Analog LCU is a physically motivated, continuous-time analogue of the LCU framework. It requires coupling a primary system to a continuous-variable ancilla. We apply this framework to develop continuous-time quantum algorithms for ground state preparation and also for solving quantum linear systems. This framework can be seen as a way to exploit qubit-qumode interactions to perform meaningful computational tasks. Such hybrid systems are currently being engineered in a number of quantum technological platforms such as photonics, trapped-ions, Circuit (or Cavity) QED and superconducting systems [14, 15, 16, 17, 3, 18, 19]. In order to experimentally implement the quantum algorithms we discuss, it is crucial to undertake a detailed comparative analysis of the resource requirements for each of these platforms. In future, we plan to develop an experimental proposal in this regard. Our work could lead to further research into whether other, simpler interactions can be engineered on hybrid platforms [88]. This would help bring generic quantum algorithmic frameworks closer to realization.

The *Ancilla-free LCU* approach is useful when we are interested in the projection of $f(H)\,|\psi_0\rangle$ in some subspace of interest, and it suffices if the measurement is successful on average. We have shown that it is applicable to the framework of quantum walks, in particular, to quantum spatial search algorithms. This technique has been useful to connect discrete and continuous-time quantum walks, with their classical counterparts. Using this framework, we have also developed other results. We believe that this method is more widely applicable to quantum optimization and sampling algorithms such as quantum simulated annealing [89] and quantum Metropolis sampling [90, 91].

Overall, the new LCU techniques we develop are quite generic. Owing to the wide applicability of the LCU framework, our work can lead to the development of several new quantum algorithms beyond those considered here. One immediate direction of future research would be to investigate whether variants of other generic quantum algorithmic paradigms which require access to the block encoding of an operator (such as QSVT [45, 46]), can be modified so that they are implementable on intermediate-term quantum computers.

Acknowledgments

I thank András Gilyén, Jérémie Roland, Simon Apers, and Leonardo Novo for helpful discussions. I also thank Samson Wang and Mario Berta for providing feedback on an early version of the manuscript. I am grateful to Leonardo Novo and Hamed Mohammady for proofreading this manuscript. I acknowledge funding from the Science and Engineering Research Board, Department of Science and Technology (SERB-DST), Government of India under Grant No. SRG/2022/000354, and from the Ministry of Electronics and Information Technology (MeitY), Government of India, under Grant No. 4(3)/2024-ITEA. I also acknowledge support from Fujitsu Ltd, Japan and from IIIT Hyderabad via the Faculty Seed Grant. Finally, I would like to thank Tanima Karmakar for acting as a sounding board during the writing of this manuscript.

Note added: We refer the readers to the concurrent independent work of Wang, McArdle, and Berta [92], where a technique similar to the Single-Ancilla LCU method was applied to ground state preparation and quantum linear systems.

References

[1] Google Quantum AI. Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505–510, 2019. doi:10.1038/s41586-019-1666-5.

[2] Han-Sen Zhong, Hui Wang, Yu-Hao Deng, Ming-Cheng Chen, Li-Chao Peng, Yi-Han Luo, Jian Qin, Dian Wu, Xing Ding, Yi Hu, et al. Quantum computational advantage using photons. *Science*, 370(6523):1460–1463, 2020. doi:10.1126/science.abe8770.

[3] Philippe Campagne-Ibarcq, Alec Eickbusch, Steven Touzard, Evan Zalys-Geller, Nicholas E Frattini, Volodymyr V Sivak, Philip Reinhold, Shruti Puri, Shyam Shankar, Robert J Schoelkopf, et al. Quantum error correction of a qubit encoded in grid states of an oscillator. *Nature*, 584(7821):368–372, 2020. doi:10.1038/s41586-020-2603-3.

[4] Lars S Madsen, Fabian Laudenbach, Mohsen Falamarzi Askarani, Fabien Rortais, Trevor Vincent, Jacob FF Bulmer, Filippo M Miatto, Leonhard Neuhaus, Lukas G Helt, Matthew J Collins, et al. Quantum computational advantage with a programmable photonic processor. *Nature*, 606(7912):75–81, 2022. doi:10.1038/s41586-022-04725-x.

[5] Google Quantum AI. Suppressing quantum errors by scaling a surface code logical qubit. *Nature*, 614:676–681, 2023. doi:10.1038/s41586-022-05434-1.

[6] John Preskill. Quantum computing in the NISQ era and beyond. *Quantum*, 2:79, 2018. doi:10.22331/q-2018-08-06-79.

[7] Kishor Bharti, Alba Cervera-Lierta, Thi Ha Kyaw, Tobias Haug, Sumner Alperin-Lea, Abhinav Anand, Matthias Degroote, Hermanni Heimonen, Jakob S Kottmann, Tim Menke, et al. Noisy intermediate-scale quantum algorithms. *Reviews of Modern Physics*, 94(1):015004, 2022. doi:10.1103/RevModPhys.94.015004.

[8] Earl T Campbell. Early fault-tolerant simulations of the hubbard model. *Quantum Science and Technology*, 7(1):015007, 2021. doi:10.1088/2058-9565/ac3110.

[9] Paul K. Faehrmann, Mark Steudtner, Richard Kueng, Maria Kieferova, and Jens Eisert. Randomizing multi-product formulas for Hamiltonian simulation. *Quantum*, 6:806, 2022. doi:10.22331/q-2022-09-19-806.

[10] Yulong Dong, Lin Lin, and Yu Tong. Ground-state preparation and energy estimation on early fault-tolerant quantum computers via quantum eigenvalue transformation of unitary matrices. *PRX Quantum*, 3(4):040305, 2022. doi:10.1103/PRXQuantum.3.040305.

[11] Ruizhe Zhang, Guoming Wang, and Peter Johnson. Computing ground state properties with early fault-tolerant quantum computers. *Quantum*, 6:761, 2022. doi:10.22331/q-2022-07-11-761.

[12] Lin Lin and Yu Tong. Heisenberg-limited ground-state energy estimation for early fault-tolerant quantum computers. *PRX Quantum*, 3(1):010318, 2022. doi:10.1103/PRXQuantum.3.010318.

[13] Guoming Wang, Daniel Stilck-França, Ruizhe Zhang, Shuchen Zhu, and Peter D Johnson. Quantum algorithm for ground state energy estimation using circuit depth with exponentially improved dependence on precision. *Quantum*, 7:1167, 2023. doi:10.22331/q-2023-11-06-1167.

[14] Andreas Wallraff, David I Schuster, Alexandre Blais, Luigi Frunzio, R-S Huang, Johannes Majer, Sameer Kumar, Steven M Girvin, and Robert J Schoelkopf. Strong coupling of a single photon to a superconducting qubit using circuit quantum electrodynamics. *Nature*, 431(7005):162–167, 2004. doi:10.1038/nature02851.

[15] J-M Pirkkalainen, SU Cho, Jian Li, GS Paraoanu, PJ Hakonen, and MA Sillanpää. Hybrid circuit cavity quantum electrodynamics with a micromechanical resonator. *Nature*, 494(7436):211–215, 2013. doi:10.1038/nature11821.

[16] Gershon Kurizki, Patrice Bertet, Yuimaru Kubo, Klaus Mølmer, David Petrosyan, Peter Rabl, and Jörg Schmiedmayer. Quantum technologies with hybrid systems. *Proceedings of the National Academy of Sciences*, 112(13):3866–3873, 2015. doi:10.1073/pnas.1419326112.

[17] Ulrik L Andersen, Jonas S Neergaard-Nielsen, Peter Van Loock, and Akira Furusawa. Hybrid discrete-and continuous-variable quantum information. *Nature Physics*, 11(9):713–719, 2015. doi:10.1038/nphys3410.

[18] HCJ Gan, Gleb Maslennikov, Ko-Wei Tseng, Chihuan Nguyen, and Dzmitry Matsukevich. Hybrid quantum computing with conditional beam splitter gate in trapped ion system. *Physical review letters*, 124(17):170502, 2020. doi:10.1103/PhysRevLett.124.170502.

[19] Nicolas PD Sawaya, Tim Menke, Thi Ha Kyaw, Sonika Johri, Alán Aspuru-Guzik, and Gian Giacomo Guerreschi. Resource-efficient digital quantum simulation of d-level systems for photonic, vibrational, and spin-s hamiltonians. *npj Quantum Information*, 6(1):49, 2020. doi:10.1038/s41534-020-0278-0.

[20] Dietrich Leibfried, Rainer Blatt, Christopher Monroe, and David Wineland. Quantum dynamics of single trapped ions. *Reviews of Modern Physics*, 75(1):281, 2003. doi:10.1103/RevModPhys.75.281.

[21] Daniel Gottesman, Alexei Kitaev, and John Preskill. Encoding a qubit in an oscillator. *Physical Review A*, 64(1):012310, 2001. doi:10.1103/PhysRevA.64.012310.

[22] Andrew M Childs and Nathan Wiebe. Hamiltonian simulation using linear combinations of unitary operations. *Quantum Information & Computation*, 12(11-12):901–924, 2012. URL: https://dl.acm.org/doi/10.5555/2481569.2481570.

[23] Dominic W. Berry, Andrew M. Childs, Richard Cleve, Robin Kothari, and Rolando D. Somma. Exponential improvement in precision for simulating sparse hamiltonians. In *Proceedings of the Forty-Sixth Annual ACM Symposium on Theory of Computing*, STOC '14, page 283–292, New York, NY, USA, 2014. Association for Computing Machinery. doi:10.1145/2591796.2591854.

[24] Dominic W. Berry, Andrew M. Childs, and Robin Kothari. Hamiltonian simulation with nearly optimal dependence on all parameters. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 792–809, 2015. doi:10.1109/FOCS.2015.54.

[25] Dominic W Berry, Andrew M Childs, Richard Cleve, Robin Kothari, and Rolando D Somma. Simulating Hamiltonian dynamics with a truncated Taylor series. *Physical review letters*, 114(9):090502, 2015. doi:10.1103/PhysRevLett.114.090502.

[26] Andrew M. Childs, Robin Kothari, and Rolando D. Somma. Quantum algorithm for systems of linear equations with exponentially improved dependence on precision. *SIAM Journal on Computing*, 46(6):1920–1950, 2017. doi:10.1137/16M1087072.

[27] Shantanav Chakraborty, András Gilyén, and Stacey Jeffery. The Power of Block-Encoded Matrix Powers: Improved Regression Techniques via Faster Hamiltonian Simulation. In Christel Baier, Ioannis Chatzigiannakis, Paola Flocchini, and Stefano Leonardi, editors, *46th International Colloquium on Automata, Languages, and Programming (ICALP 2019)*, volume 132 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 33:1–33:14, Dagstuhl, Germany, 2019. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/LIPIcs.ICALP.2019.33.

[28] Dominic W Berry, Andrew M Childs, Aaron Ostrander, and Guoming Wang. Quantum algorithm for linear differential equations with exponentially improved dependence on precision. *Communications in Mathematical Physics*, 356:1057–1081, 2017. doi:10.1007/s00220-017-3002-y.

[29] Jin-Peng Liu, Herman Øie Kolden, Hari K Krovi, Nuno F Loureiro, Konstantina Trivisa, and Andrew M Childs. Efficient quantum algorithm for dissipative nonlinear differential equations. *Proceedings of the National Academy of Sciences*, 118(35):e2026805118, 2021. doi:10.1073/pnas.2026805118.

[30] Andrew M Childs, Jin-Peng Liu, and Aaron Ostrander. High-precision quantum algorithms for partial differential equations. *Quantum*, 5:574, 2021. doi:10.22331/q-2021-11-10-574.

[31] Simon Apers and Alain Sarlette. Quantum fast-forwarding: Markov chains and graph property testing. *Quantum Information & Computation*, 19(3-4):181–213, 2019. URL: https://dl.acm.org/doi/10.5555/3370245.3370246.

[32] Andris Ambainis, András Gilyén, Stacey Jeffery, and Martins Kokainis. Quadratic speedup for finding marked vertices by quantum walks. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2020, page 412–424, New York, NY, USA, 2020. Association for Computing Machinery. doi:10.1145/3357713.3384252.

[33] Simon Apers, Shantanav Chakraborty, Leonardo Novo, and Jérémie Roland. Quadratic speedup for spatial search by continuous-time quantum walk. *Physical Review Letters*, 129(16):160502, 2022. doi:10.1103/PhysRevLett.129.160502.

[34] Yimin Ge, Jordi Tura, and J Ignacio Cirac. Faster ground state preparation and high-precision ground energy estimation with fewer qubits. *Journal of Mathematical Physics*, 60(2):022202, 2019. doi:10.1063/1.5027484.

[35] Trevor Keen, Eugene Dumitrescu, and Yan Wang. Quantum algorithms for ground-state preparation and Green's function calculation. *arXiv preprint arXiv:2112.05731*, 2021. doi:10.48550/arXiv.2112.05731.

[36] Min-Quan He, Dan-Bo Zhang, and Z. D. Wang. Quantum Gaussian filter for exploring ground-state properties. *Phys. Rev. A*, 106:032420, 2022. doi:10.1103/PhysRevA.106.032420.

[37] Anirban Narayan Chowdhury and Rolando D Somma. Quantum algorithms for gibbs sampling and hitting-time estimation. *Quantum Information & Computation*, 17(1-2):41–64, 2017. URL: https://dl.acm.org/doi/10.5555/3179483.3179486.

[38] Joran Van Apeldoorn, András Gilyén, Sander Gribling, and Ronald de Wolf. Quantum SDP-solvers: Better upper and lower bounds. *Quantum*, 4:230, 2020. doi:doi.org/10.22331/q-2020-02-14-230.

[39] M. Szegedy. Quantum speed-up of markov chain based algorithms. In *45th Annual IEEE Symposium on Foundations of Computer Science*, pages 32–41, 2004. doi:10.1109/FOCS.2004.53.

[40] Frédéric Magniez, Ashwin Nayak, Jérémie Roland, and Miklos Santha. Search via quantum walk. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 575–584, 2007. doi:10.1137/090745854.

[41] Hari Krovi, Frédéric Magniez, Maris Ozols, and Jérémie Roland. Quantum walks can find a marked element on any graph. *Algorithmica*, 74(2):851–907, 2016. doi:10.1007/s00453-015-9979-8.

[42] Simon Apers, András Gilyén, and Stacey Jeffery. A Unified Framework of Quantum Walk Search. In Markus Bläser and Benjamin Monmege, editors, *38th International Symposium on Theoretical Aspects of Computer Science (STACS 2021)*, volume 187 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 6:1–6:13, Dagstuhl, Germany, 2021. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.STACS.2021.6.

[43] Guang Hao Low and Isaac L Chuang. Hamiltonian simulation by qubitization. *Quantum*, 3:163, 2019. doi:10.22331/q-2019-07-12-163.

[44] András Gilyén, Yuan Su, Guang Hao Low, and Nathan Wiebe. Quantum singular value transformation and beyond: exponential improvements for quantum matrix arithmetics. *arXiv preprint arXiv:1806.01838*, 2018. doi:10.48550/arXiv.1806.01838.

[45] András Gilyén, Yuan Su, Guang Hao Low, and Nathan Wiebe. Quantum singular value transformation and beyond: Exponential improvements for quantum matrix arithmetics. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2019, page 193–204, New York, NY, USA, 2019. Association for Computing Machinery. doi:10.1145/3313276.3316366.

[46] John M Martyn, Zane M Rossi, Andrew K Tan, and Isaac L Chuang. Grand unification of quantum algorithms. *PRX Quantum*, 2(4):040203, 2021. doi:10.1103/PRXQuantum.2.040203.

[47] Gilles Brassard, Peter Hoyer, Michele Mosca, and Alain Tapp. Quantum amplitude amplification and estimation. *Contemporary Mathematics*, 305:53–74, 2002. doi:10.1090/conm/305.

[48] Andrew M Childs, Yuan Su, Minh C Tran, Nathan Wiebe, and Shuchen Zhu. Theory of Trotter error with commutator scaling. *Physical Review X*, 11(1):011020, 2021. doi:10.1103/PhysRevX.11.011020.

[49] Andrew M Childs, Dmitri Maslov, Yunseong Nam, Neil J Ross, and Yuan Su. Toward the first quantum simulation with quantum speedup. *Proceedings of the National Academy of Sciences*, 115(38):9456–9461, 2018. doi:10.1073/pnas.1801723115.

[50] Qi Zhao, You Zhou, Alexander F Shaw, Tongyang Li, and Andrew M Childs. Hamiltonian Simulation with random inputs. *Physical Review Letters*, 129(27):270502, 2022. doi:10.1103/PhysRevLett.129.270502.

[51] Pedro CS Costa, Dong An, Yuval R Sanders, Yuan Su, Ryan Babbush, and Dominic W Berry. Optimal scaling quantum linear-systems solver via discrete adiabatic theorem. *PRX Quantum*, 3(4):040303, 2022. doi:10.1103/PRXQuantum.3.040303.

[52] Guang Hao Low and Isaac L Chuang. Hamiltonian simulation by uniform spectral amplification. *arXiv preprint arXiv:1707.05391*, 2017. doi:10.48550/arXiv.1707.05391.

[53] Subir Sachdev. *Quantum Phase Transitions*. Cambridge University Press, 2nd edition, 2011. doi:10.1017/CBO9780511973765.

[54] Sam McArdle, Suguru Endo, Alán Aspuru-Guzik, Simon C Benjamin, and Xiao Yuan. Quantum computational chemistry. *Reviews of Modern Physics*, 92(1):015003, 2020. doi:10.1103/RevModPhys.92.015003.

[55] David J Griffiths and Darrell F Schroeter. *Introduction to quantum mechanics*. Cambridge University Press, 2018. doi:doi.org/10.1017/9781316995433.

[56] Christian Weedbrook, Stefano Pirandola, Raúl García-Patrón, Nicolas J Cerf, Timothy C Ralph, Jeffrey H Shapiro, and Seth Lloyd. Gaussian quantum information. *Reviews of Modern Physics*, 84(2):621, 2012. doi:10.1103/RevModPhys.84.621.

[57] Andrew M. Childs and Jeffrey Goldstone. Spatial search by quantum walk. *Phys. Rev. A*, 70:022314, 2004. doi:10.1103/PhysRevA.70.022314.

[58] Andrew M Childs. On the relationship between continuous-and discrete-time quantum walk. *Communications in Mathematical Physics*, 294(2):581–603, 2010. doi:10.1007/s00220-009-0930-1.

[59] Oleksandr Kyriienko. Quantum inverse iteration algorithm for programmable quantum simulators. *npj Quantum Information*, 6(1):7, 2020. doi:10.1038/s41534-019-0239-7.

[60] Earl Campbell. Random compiler for fast Hamiltonian Simulation. *Physical review letters*, 123(7):070503, 2019. doi:10.1103/PhysRevLett.123.070503.

[61] Chi-Fang Chen, Hsin-Yuan Huang, Richard Kueng, and Joel A. Tropp. Concentration for random product formulas. *PRX Quantum*, 2:040305, Oct 2021. doi:10.1103/PRXQuantum.2.040305.

[62] Guang Hao Low and Isaac L Chuang. Optimal Hamiltonian Simulation by Quantum Signal Processing. *Physical Review Letters*, 118(1):010501, 2017. doi:10.1103/PhysRevLett.118.010501.

[63] Daniel S Abrams and Seth Lloyd. Quantum algorithm providing exponential speed increase for finding eigenvalues and eigenvectors. *Physical Review Letters*, 83(24):5162, 1999. doi:10.1103/PhysRevLett.83.5162.

[64] Lin Lin and Yu Tong. Near-optimal ground state preparation. *Quantum*, 4:372, 2020. doi:10.22331/q-2020-12-14-372.

[65] Aram W Harrow, Avinatan Hassidim, and Seth Lloyd. Quantum algorithm for linear systems of equations. *Physical Review Letters*, 103(15):150502, 2009. doi:10.1103/PhysRevLett.103.150502.

[66] Leonard Wossnig, Zhikuan Zhao, and Anupam Prakash. Quantum linear system algorithm for dense matrices. *Physical Review Letters*, 120(5):050502, 2018. doi:10.1103/PhysRevLett.120.050502.

[67] Shantanav Chakraborty, Aditya Morolia, and Anurudh Peduri. Quantum Regularized Least Squares. *Quantum*, 7:988, April 2023. doi:10.22331/q-2023-04-27-988.

[68] Yiğit Subaşı, Rolando D Somma, and Davide Orsucci. Quantum algorithms for systems of linear equations inspired by adiabatic quantum computing. *Physical Review Letters*, 122(6):060504, 2019. doi:10.1103/PhysRevLett.122.060504.

[69] Lin Lin and Yu Tong. Optimal polynomial based quantum eigenstate filtering with application to solving quantum linear systems. *Quantum*, 4:361, 2020. doi:10.22331/q-2020-11-11-361.

[70] Hsin-Yuan Huang, Kishor Bharti, and Patrick Rebentrost. Near-term quantum algorithms for linear systems of equations with regression loss functions. *New Journal of Physics*, 23(11):113021, 2021. doi:10.1088/1367-2630/ac325f.

[71] Andrew M Childs, Enrico Deotto, Edward Farhi, Jeffrey Goldstone, Sam Gutmann, and Andrew J Landahl. Quantum search by measurement. *Physical Review A*, 66(3):032314, 2002. doi:10.1103/PhysRevA.66.032314.

[72] Shantanav Chakraborty, Kyle Luh, and Jérémie Roland. Analog quantum algorithms for the mixing of markov chains. *Physical Review A*, 102(2):022423, 2020. doi:10.1103/PhysRevA.102.022423.

[73] Shantanav Chakraborty, Leonardo Novo, Andris Ambainis, and Yasser Omar. Spatial search by quantum walk is optimal for almost all graphs. *Physical review letters*, 116(10):100501, 2016. doi:10.1103/PhysRevLett.116.100501.

[74] Shantanav Chakraborty, Leonardo Novo, and Jérémie Roland. Optimality of spatial search via continuous-time quantum walks. *Physical Review A*, 102(3):032214, 2020. doi:10.1103/PhysRevA.102.032214.

[75] Shantanav Chakraborty, Leonardo Novo, and Jérémie Roland. Finding a marked node on any graph via continuous-time quantum walks. *Physical Review A*, 102(2):022227, 2020. doi:10.1103/PhysRevA.102.022227.

[76] Anirban N Chowdhury, Rolando D Somma, and Yiğit Subaşı. Computing partition functions in the one-clean-qubit model. *Physical Review A*, 103(3):032422, 2021. doi:10.1103/PhysRevA.103.032422.

[77] Mary Beth Ruskai. Inequalities for traces on von neumann algebras. *Communications in Mathematical Physics*, 26:280–289, 1972. doi:doi.org/10.1007/BF01645523.

[78] Patrick Rall. Quantum algorithms for estimating physical quantities using block encodings. *Physical Review A*, 102(2):022408, 2020. doi:10.1103/PhysRevA.102.022408.

[79] Dmitry Grinko, Julien Gacon, Christa Zoufal, and Stefan Woerner. Iterative quantum amplitude estimation. *npj Quantum Information*, 7(1):52, 2021. doi:10.1038/s41534-021-00379-1.

[80] Kianna Wan, Mario Berta, and Earl T. Campbell. Randomized quantum algorithm for statistical phase estimation. *Physical Review Letters*, 129:030503, Jul 2022. doi:10.1103/PhysRevLett.129.030503.

[81] Jeongwan Haah, Matthew B Hastings, Robin Kothari, and Guang Hao Low. Quantum algorithm for simulating real time evolution of lattice hamiltonians. *SIAM Journal on Computing*, 52(6):FOCS18–250–FOCS18–284, 2021. doi:10.1137/18M1231511.

[82] Julia Kempe, Alexei Kitaev, and Oded Regev. The complexity of the local hamiltonian problem. *SIAM Journal on Computing*, 35(5):1070–1097, 2006. doi:10.1137/S00975397044452.

[83] Andris Ambainis. Variable time amplitude amplification and quantum algorithms for linear algebra problems. In Christoph Dürr and Thomas Wilke, editors, *29th International Symposium on Theoretical Aspects of Computer Science (STACS 2012)*, volume 14 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 636–647, Dagstuhl, Germany, 2012. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.STACS.2012.636.

[84] Sushant Sachdeva, Nisheeth K Vishnoi, et al. Faster algorithms via approximation theory. *Foundations and Trends in Theoretical Computer Science*, 9(2):125–210, 2014. doi:10.1561/0400000065.

[85] Yudong Cao, Jonathan Romero, Jonathan P. Olson, Matthias Degroote, Peter D. Johnson, Mária Kieferová, Ian D. Kivlichan, Tim Menke, Borja Peropadre, Nicolas P. D. Sawaya, Sukin Sim, Libor Veis, and Alán Aspuru-Guzik. Quantum chemistry in the age of quantum computing. *Chemical Reviews*, 119(19):10856–10915, 2019. PMID: 31469277. doi:10.1021/acs.chemrev.8b00803.

[86] Yuan Su, Dominic W. Berry, Nathan Wiebe, Nicholas Rubin, and Ryan Babbush. Fault-tolerant quantum simulations of chemistry in first quantization. *PRX Quantum*, 2:040332, Nov 2021. doi:10.1103/PRXQuantum.2.040332.

[87] Ryan Babbush, Nathan Wiebe, Jarrod McClean, James McClain, Hartmut Neven, and Garnet Kin-Lic Chan. Low-depth quantum simulation of materials. *Phys. Rev. X*, 8:011044, Mar 2018. doi:10.1103/PhysRevX.8.011044.

[88] Javier Argüello-Luengo, Alejandro González-Tudela, Tao Shi, Peter Zoller, and J Ignacio Cirac. Analogue quantum chemistry simulation. *Nature*, 574(7777):215–218, 2019. doi:10.1038/s41586-019-1614-4.

[89] Rolando D Somma, Sergio Boixo, Howard Barnum, and Emanuel Knill. Quantum simulations of classical annealing processes. *Physical Review Letters*, 101(13):130504, 2008. doi:10.1103/PhysRevLett.101.130504.

[90] Kristan Temme, Tobias J Osborne, Karl G Vollbrecht, David Poulin, and Frank Verstraete. Quantum metropolis sampling. *Nature*, 471(7336):87–90, 2011. doi:10.1038/nature09770.

[91] Man-Hong Yung and Alán Aspuru-Guzik. A quantum–quantum metropolis algorithm. *Proceedings of the National Academy of Sciences*, 109(3):754–759, 2012. doi:10.1073/pnas.1111758109.

[92] Samson Wang, Sam McArdle, and Mario Berta. Qubit-efficient randomized quantum algorithms for linear algebra. *PRX Quantum*, 5:020324, 2024. doi:10.1103/PRXQuantum.5.020324.

[93] Nai-Hui Chia, András Gilyén, Han-Hsuan Lin, Seth Lloyd, Ewin Tang, and Chunhao Wang. Quantum-Inspired Algorithms for Solving Low-Rank Linear Equation Systems with Logarithmic Dependence on the Dimension. In Yixin Cao, Siu-Wing Cheng, and Minming Li, editors, *31st International Symposium on Algorithms and Computation (ISAAC 2020)*, volume 181 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 47:1–47:17, Dagstuhl, Germany, 2020. Schloss Dagstuhl–Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.ISAAC.2020.47.

[94] Ewin Tang. Quantum principal component analysis only achieves an exponential speedup because of its state preparation assumptions. *Physical Review Letters*, 127:060503, 2021. doi:10.1103/PhysRevLett.127.060503.

[95] András Gilyén, Zhao Song, and Ewin Tang. An improved quantum-inspired algorithm for linear regression. *Quantum*, 6:754, 2022. doi:10.22331/q-2022-06-30-754.

[96] Changpeng Shao and Ashley Montanaro. Faster quantum-inspired algorithms for solving linear systems. *ACM Transactions on Quantum Computing*, 3(4):1–23, 2022. doi:10.1145/3520141.

Appendix

Here, we provide detailed derivations of the unproven theorems/lemmas in the main manuscript, as well as develop some additional general results from the LCU techniques that have been introduced. In Sec. A, we provide a proof of a slightly more general version of Theorem 9. We provide detailed derivations of the unproven results of our Hamiltonian Simulation procedure (Sec. 4 of the main manuscript) in Sec. B. Recall that in the main manuscript, we provided randomized quantum algorithms for ground state property estimation, as well as quantum linear systems. However, we assumed access to a Hamiltonian evolution oracle. In Sec. C, we analyze the performance of these algorithms while considering particular Hamiltonian simulation procedures. The goal is to provide end-to-end complexities (in terms of the gate depth required) while still requiring only a single ancilla qubit (and no multi-qubit controlled operation). We provide an optimal circuit model quantum algorithm for ground state preparation in Sec. D which makes use of QSVT to implement the polynomial e^{-tx^2}. Finally, in Sec. E, we use block encoding and QSVT to obtain a relationship between discrete-time and continuous-time quantum walks.

A Proof of Theorem 9

Here we prove a general version of the statement of Theorem 9.

Theorem A1 (Robustness of normalization factors)**.** *Let $\varepsilon \in (0,1)$, ρ_0 be some initial state and P be an operator. Furthermore, let $\ell_* \in \mathbb{R}^+$ satisfies $\ell^2 = \mathrm{Tr}[P\rho_0 P^\dagger] \geq \ell_*$, and O be some observable with $\|O\| \geq 1$. Suppose for positive integers $a, b > 1$, we obtain an estimate $\tilde{\ell}$ such that*

$$\left|\tilde{\ell} - \ell^2\right| \leq \frac{\varepsilon \ell_*}{a\|O\|}, \tag{A1}$$

and some parameter μ such that,

$$\left|\mu - \mathrm{Tr}[O\ P\rho_0 P^\dagger]\right| \leq \frac{\varepsilon \ell_*}{b}, \tag{A2}$$

then,

$$\left|\frac{\mu}{\tilde{\ell}} - \frac{\mathrm{Tr}[O\ P\rho_0 P^\dagger]}{\ell^2}\right| \leq \frac{\varepsilon}{a-1} + \frac{\varepsilon}{b(a-1)} + \frac{\varepsilon}{b}.$$

Proof. By the triangle inequality, we have

$$\left|\frac{\mu}{\tilde{\ell}} - \frac{\mathrm{Tr}[OP\rho_0P^\dagger]}{\ell^2}\right| \leq \left|\frac{\mu}{\ell^2} - \frac{\mu}{\tilde{\ell}}\right| + \left|\frac{\mu}{\ell^2} - \frac{\mathrm{Tr}[OP\rho_0P^\dagger]}{\ell^2}\right| \tag{A3}$$

$$\leq \left|\frac{1}{\ell^2} - \frac{1}{\tilde{\ell}}\right| |\mu| + \frac{\varepsilon\ell_*}{b\ell^2} \qquad [\text{ Using Eq. (A2) }] \tag{A4}$$

$$\leq \left|\frac{1}{\ell^2} - \frac{1}{\tilde{\ell}}\right| |\mu| + \frac{\varepsilon}{b} \qquad [\text{ As } \ell^2 \geq \ell_* \text{ }] \tag{A5}$$

$$\leq \left|\frac{1}{\ell^2} - \frac{1}{\tilde{\ell}}\right| \left(\frac{\varepsilon\ell_*}{b} + \mathrm{Tr}[OP\rho_0P^\dagger]\right) + \frac{\varepsilon}{b} \qquad [\text{ Using Eq. (A2) }] \tag{A6}$$

$$\leq \frac{1}{\ell^2}\left|\frac{\tilde{\ell} - \ell^2}{\tilde{\ell}}\right| \left(\frac{\varepsilon\ell_*}{b} + \|O\| \left\|P\rho_0P^\dagger\right\|_1\right) + \varepsilon/b \qquad [\text{ Using Lemma 6 with } p = \infty, q = 1 \text{ }] \tag{A7}$$

$$\leq \frac{1}{\ell^2}\left|\frac{\tilde{\ell} - \ell^2}{\tilde{\ell}}\right| \left(\frac{\varepsilon\ell_*}{b} + \|O\| \, \ell^2\right) + \varepsilon/b \qquad \left[\text{ As } \left\|P\rho_0P^\dagger\right\|_1 = \mathrm{Tr}[P\rho_0P^\dagger] = \ell^2 \right] \tag{A8}$$

$$\leq \left|\frac{\tilde{\ell} - \ell^2}{\tilde{\ell}}\right| \left(\frac{\varepsilon}{b} + \|O\|\right) + \varepsilon/b \qquad [\text{ As } \ell^2 \geq \ell_*] \tag{A9}$$

$$\leq \left|\frac{1}{\tilde{\ell}}\right| \frac{\varepsilon\ell_*}{a\|O\|} \left(\|O\| + \frac{\varepsilon}{b}\right) + \varepsilon/b \qquad [\text{ Using Eq. (A1) }] \tag{A10}$$

$$\leq \left|\frac{1}{\ell^2 - \frac{\varepsilon\ell_*}{a\,\|O\|}}\right| \frac{\varepsilon\ell_*}{a\,\|O\|} \left(\|O\| + \frac{\varepsilon}{b}\right) + \varepsilon/b \tag{A11}$$

$$\leq \left(\sum_{k=0}^{\infty} \left(\frac{\varepsilon\ell_*}{a\ell^2\|O\|}\right)^k\right) \frac{\varepsilon\ell_*}{a\,\ell^2\|O\|} \left(\|O\| + \frac{\varepsilon}{b}\right) + \varepsilon/b \qquad [\text{Taylor series expansion of } 1/\tilde{\ell}.] \tag{A12}$$

$$\leq \frac{a}{a-1}\left(\frac{\varepsilon}{a} + \frac{\varepsilon^2}{ab\,\|O\|}\right) + \varepsilon/b \qquad \left[\text{ As } \sum_{k=0}^{\infty} \left(\frac{\varepsilon\ell_*}{a\ell^2\|O\|}\right)^k \leq \sum_{k=0}^{\infty} \left(\frac{1}{a}\right)^k = a/(a-1) \right] \tag{A13}$$

$$\leq \frac{\varepsilon}{a-1} + \frac{\varepsilon}{(a-1)b} + \frac{\varepsilon}{b} \qquad \left[\text{ As } \frac{\varepsilon^2}{(a-1)b\,\|O\|} \leq \frac{\varepsilon}{(a-1)b} \right]. \tag{A14}$$

□

Theorem 9 is a particular case of this theorem, where we substitute $a = b = 3$.

B Hamiltonian simulation: Detailed proofs

In this section, we will proof the results of Sec. 4 in detail. Recall that, we considered a Hamiltonian

$$H = \sum_{k=1}^{L} \lambda_k P_k,$$

where P_k are strings of Pauli operators, such that $\beta = \sum_k |\lambda_k|$. First, set $\tilde{H} = H/\beta$ and $\tilde{t} = \beta t$. This give us,

$$\tilde{H} = H/\beta = \sum_{k=1}^{L} p_k P_k, \tag{A15}$$

where $\sum_k |p_k| = 1$. Also,

$$e^{-iHt} = \left(e^{-iHt/r}\right)^r = \left(e^{-i\tilde{H}\tilde{t}/r}\right)^r,$$

where r (to be selected later) is a parameter such that $r > t$.

First note that by truncating $S_r = e^{-i\tilde{H}\tilde{t}/r}$ to K terms, we obtain

$$\tilde{S}_r = \sum_{k=0}^{K} \frac{(-i\tilde{t}\tilde{H}/r)^k}{k!}.$$

Then by choosing some

$$K = O\left(\frac{\log(r/\gamma)}{\log\log(r/\gamma)}\right),$$

we ensure that $\left\|S_r - \tilde{S}_r\right\| \leq \gamma/r$.

We obtain the LCU decomposition of $\tilde{S}_r$, similar in spirit to Ref. [80]. This gives us,

$$\tilde{S}_r = \sum_{k=0}^{K} \frac{(-i\tilde{t}\tilde{H}/r)^k}{k!} \tag{A16}$$

$$= \sum_{k=0,\ k\in\text{even}}^{K} \frac{1}{k!}(-i\tilde{t}\tilde{H}/r)^k \left(I - \frac{i\tilde{t}\tilde{H}/r}{k+1}\right) \tag{A17}$$

$$= \sum_{k=0,\ k\in\text{even}}^{K} \frac{1}{k!}\left(-i\tilde{t}/r \sum_{\ell=1}^{L} p_\ell P_\ell\right)^k \left(I - \frac{i\tilde{t}/r}{k+1}\left(\sum_{m=1}^{L} p_m P_m\right)\right) \tag{A18}$$

$$= \sum_{k=0,\ k\in\text{even}}^{K} \frac{(-i\tilde{t}/r)^k}{k!} \sum_{\ell_1,\ell_2,\cdots\ell_k=1}^{L} p_{\ell_1} p_{\ell_2} \cdots p_{\ell_k} P_{\ell_1} P_{\ell_2} \cdots P_{\ell_k} \sum_{m=1}^{L} p_m \left(I - \frac{i\tilde{t}P_m/r}{k+1}\right) \tag{A19}$$

$$= \sum_{k=0,\ k\in\text{even}}^{K} \frac{(-i\tilde{t}/r)^k}{k!} \sqrt{1 + \left(\frac{\tilde{t}/r}{k+1}\right)^2} \sum_{\ell_1,\ell_2,\cdots\ell_k,m=1}^{L} p_{\ell_1} p_{\ell_2} \cdots p_{\ell_k} p_m P_{\ell_1} P_{\ell_2} \cdots P_{\ell_k} e^{-i\theta_m P_m}, \tag{A20}$$

where $e^{-i\theta_m P_m}$ is a Pauli rotation operator, defined as follows:

$$e^{-i\theta_m P_m} = \frac{1}{\sqrt{1 + \left(\frac{\tilde{t}/r}{k+1}\right)^2}} \left(I - \frac{i\tilde{t}P_m/r}{k+1}\right), \tag{A21}$$

such that

$$\theta_m = \arccos\left(\left[1 + \left(\frac{\tilde{t}/r}{k+1}\right)^2\right]^{-1/2}\right). \tag{A22}$$

Thus, $\tilde{S}_r = \sum_{j\in M} \alpha_j U_j$, where the index set M can be defined as

$$M = \{(k, \ell_1, \ell_2, \cdots \ell_k, m) : 0 \le k \le K; \ell_1, \ell_2, \cdots \ell_k, m \in \{1, 2, \cdots, L\}\}.$$

Also,

$$\alpha_j = \frac{(\tilde{t}/r)^k}{k!} \sqrt{1 + \left(\frac{\tilde{t}/r}{k+1}\right)^2}\, p_{\ell_1} p_{\ell_2} \cdots p_{\ell_k} p_m, \tag{A23}$$

while

$$U_j = (-i)^k P_{\ell_1} P_{\ell_2} \cdots P_{\ell_k} e^{-i\theta_m P_m}.$$

Now, the sum of coefficients

$$\sum_{j\in M} |\alpha_j| = \sum_{k=0,\ k\in\text{even}}^{K} \frac{(\tilde{t}/r)^k}{k!} \sqrt{1 + \left(\frac{\tilde{t}/r}{k+1}\right)^2} \sum_{\ell_1,\ell_2,\cdots\ell_k,m=1}^{L} p_{\ell_1} p_{\ell_2} \cdots p_{\ell_k} p_m \tag{A24}$$

$$= \sum_{k=0,\ k\in\text{even}}^{K} \frac{(\tilde{t}/r)^k}{k!} \sqrt{1 + \left(\frac{\tilde{t}/r}{k+1}\right)^2} \tag{A25}$$

$$\le \sum_{k=0,\ k\in\text{even}}^{\infty} \frac{(\tilde{t}/r)^k}{k!} \sqrt{1 + \left(\frac{\tilde{t}/r}{k+1}\right)^2} = \sum_{k=0}^{\infty} \frac{(\tilde{t}/r)^{2k}}{(2k)!} \sqrt{1 + \left(\frac{\tilde{t}/r}{2k+1}\right)^2} \tag{A26}$$

$$\le \sum_{k=0}^{\infty} \frac{(\tilde{t}/r)^{2k}}{k!} = e^{\tilde{t}^2/r^2}. \tag{A27}$$

Finally, in order to write down S as an LCU, we write $S = \tilde{S}_r^r$. That is,

$$S = \left(\sum_{j\in M} \alpha_j U_j\right)^r = \sum_{j_1, j_2, \cdots j_r \in M} \alpha_1 \alpha_2 \cdots \alpha_r\ U_{j_1} U_{j_2} \cdots U_{j_r} = \sum_m c_m W_m, \tag{A28}$$

where $\|c\|_1 = \sum_m |c_m| = (\sum_{j\in M} |\alpha_j|)^r \le e^{\tilde{t}^2/r}$. We choose $r = \tilde{t}^2 = \beta^2 t^2$, which ensures $\|c\|_1 = O(1)$. Moreover, for this choice of r and

$$\gamma \le \frac{\varepsilon}{6\|O\|},$$

by truncating the Taylor series of $e^{itH/r}$ at some

$$K = O\left(\frac{\log(\beta t \|O\|/\varepsilon)}{\log\log(\beta t \|O\|/\varepsilon)}\right),$$

we have

$$\left\| e^{-itH} - S \right\| \le \frac{\varepsilon}{6\|O\|}.$$

So from Theorem 8, if we can sample V_1, V_2 from the LCU decomposition of S, we will be able to output an ε-accurate estimate of $\mathrm{Tr}[O\ e^{-itH}\rho_0 e^{itH}]$, using Algorithm 1. We discuss this sampling strategy in a bit more detail as compared to the main manuscript here.

Sampling $\mathbf{V_1}$ and $\mathbf{V_2}$: We first pick an even integer $k \in [0, K]$ according to $\alpha_j / \sum_j \alpha_j$ and select $k+1$ unitaries, $P_{\ell_1}, P_{\ell_2}, \cdots P_{\ell_k}$, and P_m (as in Eq. (54)), where each P_{ℓ_i} is sampled according to the distribution $\{p_{\ell_i}\}_{\ell_i=1}^L$ and P_m is sampled from $\{p_m\}_{m=1}^L$. From this sampling procedure, we can obtain a product of the unitaries $W_1 = (-i)^k P_{\ell_1} P_{\ell_2} \cdots P_{\ell_k} P_m$, of

k Pauli operators and a Pauli rotation. Finally, we repeat this procedure r times, which essentially results in the final unitary

$$W = W_r W_{r-1} \cdots W_1.$$

Thus, this sampling procedure outputs some unitary W such that $\mathbb{E}[W] = S/\|c\|_1$.

This allows us to use Algorithm 1 and Theorem 8. Clearly, each run of our procedure has gate depth at most $O(Kr)$, which leads to the gate depth per coherent run, and the overall gate depth as stated in Theorem 12.

C Single Ancilla LCU: from the Hamiltonian Evolution model to gate depth

In the main manuscript, we analyzed the complexity of our Hamiltonian simulation algorithm by *Single-Ancilla LCU* (Sec. 4) in terms of the gate depth per coherent run, as well as the overall gate depth. On the other hand, for both ground state property estimation (Sec. 5.2) and estimating expectation values of observables with respect to the solution of quantum linear systems (Sec. 6.2), we assumed that the Hamiltonian H can be accessed through the Hamiltonian evolution oracle $U_\tau = \exp[-iH\tau]$. We measured the performance of our algorithm in terms of (a) the maximal time evolution of H in one coherent run (denoted as $\tau_{\max}$), and (b) the number of classical repetitions T, where $O(\tau_{\max} \cdot T)$ is the total evolution time. As argued in the main article, both (a) and (b) are different from the actual circuit depth required to implement these algorithms.

In this section, we obtain the gate depth required to run both these algorithms using the *Single-Ancilla LCU* method. To this end, we consider specific Hamiltonian simulation algorithms to implement U_t to some desired precision. Our goal is to keep the number of ancilla qubits to just one, and avoid the use of multi-qubit controlled operations. As we shall see next, this limits the simulation algorithms we can use. For both algorithms, we assume the following:

(i) The Hamiltonian is a linear combination of unitaries (e.g. strings of Paulis), i.e. $H = \sum_{k=1}^{L} \lambda_k P_k$. The total weight of the coefficients $\beta = \sum_k |\lambda_k|$.

(ii) The observable O we intend to measure is itself a linear combination of easy to implement unitary observables, i.e. $O = \sum_{j=1}^{L_O} h_j O_j$, such that $\|h\|_1 = \sum_j |h_j|$ and for each j, $\|O_j\| = 1$.

Note that any block encoding of H requires $O(\log L)$ ancilla qubits, and has a sub-normalization factor of β, while a block encoding of O requires $O(\log L_O)$ ancilla qubits, with $\|h\|_1$ being the sub-normalization factor. Additionally this requires a gate depth of $O(L)$ and $O(L_O)$, respectively. Thus, constructing the block encoding of both H and O needs multi-qubit controlled operations and adds to the overall gate depth, which are undesirable for early fault-tolerant quantum computers. Moreover, our goal of using a solitary ancilla qubit (and hence, no multi-qubit control) for our algorithms implies that we cannot use any Hamiltonian simulation algorithm that uses the block encoding of H. This rules out Hamiltonian simulation by qubitization, the state-of-the-art method [43]. In fact, given this restriction we can only use Trotter-based methods and the Hamiltonian simulation algorithm based on *Single-Ancilla LCU* (Sec. 4).

We indeed show that whenever H and O are linear combinations of unitaries, we can implement *Single-Ancilla LCU* using only a single ancilla qubit and no multi-qubit controlled unitaries. In this regard, we first adapt the generic *Single-Ancilla LCU* scheme to allow for (a) the measurement of any $O = \sum_j h_j O_j$, and (b) the application of imprecise unitaries. The reason for analyzing (b) is that whenever $f(H) \approx \sum_j c_j e^{-ijH}$, the Hamiltonian simulation algorithm needs to be implemented to some desired precision, which is what we shall estimate. Equipped with (a) and (b), we can directly calculate the gate depth of both ground state property estimation, and quantum linear systems, by invoking particular Hamiltonian simulation algorithms that fit our goals. We begin by incorporating (a) and (b) into Theorem 10.

C.1 Single Ancilla LCU: general observables and imperfect unitaries

In order to measure any $O = \sum_{j=1}^{L_O} h_j O_j$ within the framework of *Single-Ancilla LCU*, we do the following: instead of measuring O directly, we simply sample an O_j according to $\{h_j/\|h\|_1\}_j$, and implement a POVM measurement of $X \otimes O_j$ in Step 3 of Algorithm 1. Since $\|O_j\| = 1$, the POVM measurement yields an outcome in $[-1, +1]$. Note that this strategy ensures that the expected outcome of the j^{th} iteration of Algorithm 1 is

$$\mathbb{E}[\mu_j] = \frac{1}{\|c\|_1^2 \|h\|_1} \mathrm{Tr}[O\ g(H)\rho_0 g(H)^\dagger],$$

where $g(H) = \sum_j c_j U_j$ is the LCU that approximates the function $f(H)$ we wish to apply. So, μ and $\tilde{\ell}$ can be obtained as in Algorithm 2, except now $\mu = \|c\|_1^2 \|h\|_1 \sum_j \mu_j / T$, and $\|O\| \leq \|h\|_1$. If we consider that the cost of implementing any $O_j = \Theta(1)$, then the cost of each coherent run is still upper bounded by $O(2\tau_{\max} + \tau_{\rho_0})$. Furthermore, following the arguments of the proofs of Theorem 8 and Theorem 10, it is easy to show that $\mu/\tilde{\ell}$ is an ε-accurate estimate of $\mathrm{Tr}[O\rho]$, with a constant success probability, for

$$T = O\left(\frac{\|c\|_1^4 \|h\|_1^2}{\varepsilon^2 \ell_*^2}\right).$$

In order to take into account the implementation of imperfect unitaries, as in Theorem 8, consider that $f(H)$ is the function we wish to apply, and $g(H) = \sum_j c_j U_j$, is the LCU it approximates. However now, instead of $g(H)$, we implement some $h(H)$ such that $h(H) = \sum_j c_j \tilde{U}_j$. Then if

$$\|f(H) - g(H)\| \leq \frac{\varepsilon}{9\|O\|\|f(H)\|}, \tag{A29}$$

and,

$$\|g(H) - h(H)\| \leq \frac{\varepsilon}{9\|O\|\|f(H)\|}, \tag{A30}$$

it suffices if in Theorem 8, μ outputs an $\varepsilon/3$-accurate estimate of $\mathrm{Tr}[O\ h(H)\rho_0 h(H)^\dagger]$.

This ensures,

$$\left|\mu - \mathrm{Tr}[O\ f(H)\rho_0 f(H)^\dagger]\right| \leq \left|\mu - \mathrm{Tr}[O\ h(H)\rho_0 h(H)^\dagger]\right| + \left|\mathrm{Tr}[O\ f(H)\rho_0 f(H)^\dagger] - \mathrm{Tr}[O\ h(H)\rho_0 h(H)^\dagger]\right| \tag{A31}$$

$$\leq \varepsilon/3 + 3\|f(H)\|\|O\|\|f(H) - h(H)\| \qquad [\text{ Using Theorem 7 }] \tag{A32}$$

$$\leq \varepsilon/3 + 3\|f(H)\|\|O\| \left[\|f(H) - g(H)\| + \|g(H) - h(H)\|\right] \tag{A33}$$

$$\leq \varepsilon. \tag{A34}$$

The equivalent statement of Theorem 10 remains the same with the upper bound on the precision adjusted appropriately. Here, we combine both these results on performing *Single-Ancilla LCU* with (a) Imperfect unitaries, and (b) observables that are LCU, and state our findings formally via the following theorem:

Theorem A2. *Let $\varepsilon, \delta \in (0,1)$, O be some observable such that $\sum_{j=1}^{L_O} h_j O_j$, with $\|h\|_1 = \sum_\alpha |h_\alpha|$, and for any j, $\|O_j\| = 1$. Also, let ρ_0 be some initial state, prepared in cost τ_{ρ_0}. Suppose $H \in \mathbb{C}^{N\times N}$ be a Hermitian matrix such that for some function $f : [-1,1] \mapsto \mathbb{R}$ and unitaries $\{U_j\}_j$,*

$$\left\| f(H) - \sum_j c_j U_j \right\| \leq \frac{\varepsilon \ell_*}{27\|h\|_1 \|f(H)\|},$$

and $\ell^2 = \mathrm{Tr}[f(H)\rho_0 f(H)^\dagger] \geq \ell_*$. Moreover, suppose that each U_j can only be imperfectly implemented: $\tilde{U}_j$ approximates U_j such that

$$\max_j \left\| U_j - \tilde{U}_j \right\| \leq \frac{\varepsilon \ell_*}{27\|h\|_1 \|f(H)\| \|c\|_1}.$$

Furthermore, suppose that the maximum cost of implementing any $\tilde{U}_j$ is at most $\tau_{\max}$. Then there exists an algorithm that outputs μ and $\tilde{\ell}$ such that

$$\left|\mu/\tilde{\ell} - \mathrm{Tr}[O\rho]\right| \leq \varepsilon,$$

with probability $(1-\delta)^2$, using

$$T = O\left(\frac{\|c\|_1^4 \|h\|_1^2}{\varepsilon^2 \ell_*^2} \ln(1/\delta)\right)$$

classical repetitions, where the cost of each such run is at most $O(2\tau_{\max} + \tau_{\rho_0})$.

Proof. Let $g(H) = \sum_j c_j U_j$ and $h(H) = \sum_j c_j \tilde{U}_j$. Then from the upper bound of $\max_j \left\| U_j - \tilde{U}_j \right\|$ mentioned in the statement of the theorem, we have

$$\|g(H) - h(H)\| = \left\| \sum_j c_j \left(U_j - \tilde{U}_j \right) \right\|$$
$$\leq \frac{\varepsilon \ell_*}{27\|h\|_1 \|f(H)\|}.$$

Now, from Theorem 9, we need to obtain an estimate $\tilde{\ell}$ of ℓ^2 such that

$$\left|\tilde{\ell} - \ell^2\right| \leq \frac{\varepsilon \ell_*}{3\|h\|_1},$$

and also output a μ such that

$$\left|\mu - \mathrm{Tr}[O\ f(H)\rho_0 f(H)^\dagger]\right| \leq \varepsilon\ell_*/3.$$

The upper bounds on $\|f(H) - g(H)\|$ and $\|g(H) - h(H)\|$ can be obtained by substituting ε with $\varepsilon\ell_*/3$ in Eq. (A29) and Eq. (A30), respectively (additionally substituting $\|O\|$ with $\|h\|_1$). Both μ and $\tilde{\ell}$ can be obtained by running Algorithm 2, except in each iteration the observable measured is some O_j sampled according to $h_j/\|h\|_1$. □

If $g(H)$ is a linear combination of time evolution operators, i.e. $g(H) = \sum_j c_j e^{-iHt_j}$, Theorem Theorem A2 gives us the precision with which Hamiltonian simulation needs to be performed. This allows us to analyze the complexity of our algorithms for ground state property estimation (Sec. 5.2) and quantum linear systems (Sec. 6.2) in terms of their gate depth. Clearly, if the maximal time evolution of the underlying algorithm was $\tau_{\max}$, now we can run a Hamiltonian simulation algorithm to implement e^{-itH} for $t = \tau_{\max}$ and precision $O(\varepsilon\ell_*\|c\|_1^{-1}\|f(H)\|^{-1}\|h\|_1^{-1})$.

C.2 Ground state property estimation and quantum linear systems

In this section, we assume $H = \sum_{k=1}^{L} c_k P_k$ with $\beta = \sum_k |c_k|$, and the observable $O = \sum_{j=1}^{L_O} h_j O_j$, with $\|O_j\| = 1$. The assumptions on H for ground state property estimation are quite natural, as this is precisely the form of most physical Hamiltonians. For quantum linear systems too, we assume that the matrix to be inverted can be written down as linear combination of Paulis, which is non-standard. Indeed generally, it is assumed that H can be accessed via a block encoding (implicitly implying that H is sparse or it is stored in some quantum accessible data structure [27]). However, (i) this requires an additional overhead which is undesirable in the intermediate-term and (ii) assumes access to a quantum RAM. It is then reasonable to assume that in the early fault-tolerant era, quantum linear systems algorithm will be employed to solve physically relevant problems, where the underlying data matrix directly corresponds to some physical Hamiltonian, which would avoid the need to handle classical data. This motivates using H which can be expressed as a linear combination of Paulis, also for solving quantum linear systems.

For both of our quantum algorithms, we need to implement a linear combination of Hamiltonian evolution operators, i.e.

$$f(H) \approx \sum_j c_j e^{-it_j H}.$$

In order to estimate the gate depth, we need to choose a specific Hamiltonian simulation procedure to implement e^{-iHt} to the desired precision. The goal of running our algorithms with a solitary ancilla qubit and no multi-qubit controlled gates imply that we can make use of only Trotter-based methods and the *Single-Ancilla LCU* Hamiltonian simulation algorithm. Both these methods require have a super linear dependence on t in terms of the gate depth, which is suboptimal (See Table 2). However, state-of-the-art methods which make use of a block encoding of H, have a gate depth per coherent run which depends on L. Thus, there exist regimes where our method requires a shorter gate depth per iteration.

Problem	Hamiltonian Simulation procedure used	Ancilla	Gate depth per coherent run	Classical repetitions
Ground state property estimation	Single-Ancilla LCU (Sec. 4)	1	$\widetilde{O}\left(\frac{\beta^2}{\Delta^2}\right)$	$O\left(\frac{\|h\|_1^2}{\eta^4\varepsilon^2}\right)$
	$2k$-order Trotter	1	$\widetilde{O}\left(L\left(\frac{\beta}{\Delta}\right)^{1+\frac{1}{2k}}\left(\frac{\|h\|_1}{\varepsilon\eta^2}\right)^{\frac{1}{2k}}\right)$	$O\left(\frac{\|h\|_1^2}{\eta^4\varepsilon^2}\right)$
Quantum Linear Systems	Single-Ancilla LCU (Sec. 4)	1	$\widetilde{O}\left(\beta^2\kappa^2\right)$	$\widetilde{O}\left(\frac{\kappa^4\|h\|_1^2}{\varepsilon^2}\right)$
	$2k$-order Trotter	1	$\widetilde{O}\left(L\kappa^{1+\frac{3}{2k}}(\beta)^{1+\frac{1}{2k}}\left(\frac{\|h\|_1}{\varepsilon}\right)^{\frac{1}{2k}}\right)$	$\widetilde{O}\left(\frac{\kappa^4\|h\|_1^2}{\varepsilon^2}\right)$

Table A1: Summary of the complexity of quantum algorithms by *Single-Ancilla LCU* for ground state property estimation and quantum linear systems. For both the algorithms we assume that $H = \sum_{k=1}^{L} \lambda_k P_k$, where P_k is a string of Pauli operators. We define $\beta = \sum_k |\lambda_k|$. Also, for both cases we assume that the observable $O = \sum_{j=1}^{L_O} h_j O_j$, with total weight $\|h\|_1 = \sum_j |h_j|$ and each $\|O_j\| = 1$. The complexity is measured in terms of the gate depth per coherent run of the algorithm, with the overall gate depth being the product of this quantity with the total number of classical repetitions (product of the complexity of the last two columns). For ground state property estimation, we assume that H has a spectral gap Δ, and we have knowledge of its ground energy to some precision. Furthermore, we also assume that we have an initial state $|\psi_0\rangle$ with an overlap of at least η with $|v_0\rangle$, the unknown ground state of H. Our algorithm estimates $\langle v_0|O|v_0\rangle$ to additive accuracy ε. On the other hand, for quantum linear systems we assume that H has eigenvalues in the range $[-1, -1/\kappa] \cup [1/\kappa, 1]$ and access to an initial state $|b\rangle$. The algorithm outputs an ε-accurate estimate of $\langle x|O|x\rangle$, where $|x\rangle = H^{-1}|b\rangle / \|H^{-1}|b\rangle\|$.

We consider two Hamiltonian simulation procedures: (a) Hamiltonian simulation by *Single-Ancilla LCU*, and (b) $2k$-order Trotter. For these methods, given any H and O defined as above, in order to implement the LCU $g(H) = \sum_j c_j e^{-it_j H}$ requires invoking Hamiltonian simulation for a maximal time $\max_j t_j$ and precision of at most $\varepsilon_H = O(\varepsilon \ell_* \|c\|_1^{-1} \|h\|_1^{-1} \|f(H)\|^{-1})$. This can be integrated into the framework of *Single-Ancilla LCU* in the following way:

(i) **Single-Ancilla LCU:** For each coherent run, the crucial task is to sample V_1 and V_2. We first sample j according to $\{c_j/\|c\|_1\}_j$, fix $r = \beta^2 t_j^2$ and some

$$K = O((\log(\beta t_j \|h\|_1 / \varepsilon_H) / \log\log(\beta t_j \|h\|_1 / \varepsilon_H)).$$

Now we can sample a sequence of Pauli matrices and a single Pauli rotation, repeating the sampling procedure r times (as described in Sec. 4). As the ℓ_1 -norm of the LCU coefficients due to Hamiltonian simulation is $O(1)$, the ℓ_1-norm of the overall LCU is still $O(\|c\|_1)$. Moreover, this requires only a single ancilla qubit.

Then if $\tau_{\max}$ is the maximal evolution time, the gate depth of each coherent run is at most

$$O\left(\beta^2 \tau_{\max}^2 \frac{\log\left(\beta \tau_{\max} \|h\|_1 \|c\|_1 \|f(H)\| \ell_*^{-1} / \varepsilon\right)}{\log\log\left(\beta \tau_{\max} \|h\|_1 \|c\|_1 \|f(H)\| \ell_*^{-1} / \varepsilon\right)}\right) = \widetilde{O}(\beta^2 \tau_{\max}^2). \tag{A35}$$

The number of classical repetitions remain the same as in Theorem A2.

(ii) $2k$**-order Trotter**: In this case, we simply sample j according to $\{c_j/\|c\|_1\}_j$ and implement $e^{-it_j H}$ to precision ε_H, using $2k$-order Trotter. This boils down to implementing a product of Pauli rotations controlled by the single ancilla qubit. The gate depth for each coherent run is upper bounded by

$$O\left(L(\beta\tau_{\max})^{1+\frac{1}{2k}}\left(\frac{\|h\|_1\|c\|_1\|f(H)\|}{\varepsilon\ell_*}\right)^{\frac{1}{2k}}\right), \tag{A36}$$

where L is the total number of terms in the Hamiltonian H. The number of classical repetitions remain the same as in Theorem A2.

Finally, we can directly use Eq. (A35) and Eq. (A36) into our algorithms after substituting the appropriate values of $\tau_{\max}, \|f(H)\|, \ell_*$ and $\|c\|_1$. From Sec. 5.2, we have that for ground state property estimation (Sec. 5.2), $\tau_{\max} = O(\Delta^{-1}\log(\ ||h||_1\eta^{-1}\varepsilon^{-1}))$, $\|f(H)\| = 1$, $\ell_* \geq \eta^2$ and $\|c\|_1 = O(1)$. On the other hand, from Sec. 6.2, for quantum linear systems we have, $\tau_{\max} = O(\kappa\log(\kappa\|h\|_1/\varepsilon))$, $\|f(H)\| \leq \kappa$, $\ell_* = 1$ and $\|c\|_1 = \widetilde{O}(\kappa)$. The gate depth for each coherent run and the total number of classical repetitions are summarized in Table A1. The overall gate depth is the product of the gate depth per coherent run and the number of classical repetitions.

C.2.1 Comparison with other methods

Let us now compare our algorithms with other methods. As before (Table 3 and Table 4), we compare with Standard LCU, QSVT, as well as early fault-tolerant quantum algorithms. For Standard LCU, we need to implement multi-qubit controlled Hamiltonian simulation. For this, we assume that *Standard LCU* uses the state-of-the-art algorithm (Hamiltonian simulation by qubitization [43]). This requires a block encoding access to H. When H is a linear combination of strings of Pauli operators as defined previously, it is possible to construct a $(\beta, O(\log L), 0)$-block encoding of H, in a gate depth of $O(L)$. The QSVT-based quantum algorithms that we consider also require access to this block encoding. So, this cost will be incorporated into the complexity of the algorithms we analyze next.

Ground state property estimation: We compare the complexities of our method with other approaches in Table A2. Let us begin by looking at the complexity of the three ways in which the *Standard LCU* procedure can estimate $\langle v_0|O|v_0\rangle$:

- By using Standard LCU to implement controlled Hamiltonian simulation followed by amplitude amplification, the state $|v_0\rangle$ can be prepared. If the Hamiltonian simulation makes use of qubitization as in [43], the overall method requires ancilla qubits to implement (a) the block encoding of H, and (b) the linear combination of Hamiltonian simulation. Overall, $O(\log L + \log(\log(\|h\|_1\,\eta^{-1}\varepsilon^{-1})/\Delta)$ ancilla qubits are needed, along with multi-qubit controlled operations. The gate depth of the circuit that prepares $|v_0\rangle$ is $\widetilde{O}(\beta L\Delta^{-1}\eta^{-1})$. Following this, $\langle v_0|O|v_0\rangle$ can be measured by simply sampling O_j according to $h_j/\|h\|_1$ in each run, and measuring O_j. This requires $O(\ \|h\|_1^2/\varepsilon^2)$ classical repetitions. Thus, as compared to our method

Algorithm	Variant	Ancilla	Gate depth per coherent run	Classical repetitions
Standard LCU [34] (with Hamiltonian simulation by qubitization [43])	QAA	$O\left(\log L+\log\left(\log\left(\frac{\|O\|}{\eta\varepsilon}\right)/\Delta\right)\right)$	$\tilde{O}\left(\frac{\beta L}{\Delta\eta}\right)$	$O\left(\frac{\|h\|_1^2}{\varepsilon^2}\right)$
	QAE	$O\left(\log L_O+\log L+\log\left(\log\left(\frac{\|O\|}{\eta\varepsilon}\right)/\Delta\right)\right)$	$\tilde{O}\left(\frac{\beta L\|h\|_1}{\Delta\eta\varepsilon}+\frac{\|h\|_1 L_O}{\varepsilon}\right)$	$O(1)$
	Without QAA or QAE	$O\left(\log L+\log\left(\log\left(\frac{\|O\|}{\eta\varepsilon}\right)/\Delta\right)\right)$	$\tilde{O}\left(\frac{\beta L}{\Delta}\right)$	$O\left(\frac{\|h\|_1^2}{\varepsilon^2\eta^2}\right)$
QSVT [64]	QAA	$O(\log L)$	$\tilde{O}\left(\frac{\beta L}{\Delta\eta}\right)$	$O\left(\frac{\|h\|_1^2}{\varepsilon^2}\right)$
	QAE	$O\left(\log L_O+\log L\right)$	$\tilde{O}\left(\frac{\beta L\|h\|_1}{\Delta\eta\varepsilon}+\frac{\|h\|_1 L_O}{\varepsilon}\right)$	$O(1)$
	Without QAA or QAE	$O\left(\log L\right)$	$\tilde{O}\left(\frac{\beta L}{\Delta}\right)$	$O\left(\frac{\|h\|_1^2}{\varepsilon^2\eta^2}\right)$
This work	Ham. Sim. by Single-Ancilla LCU	1	$\tilde{O}\left(\frac{\beta^2}{\Delta^2}\right)$	$O\left(\frac{\|h\|_1^2}{\varepsilon^2\eta^4}\right)$
	Ham. Sim. by $2k$-order Trotter	1	$\tilde{O}\left(L\left(\frac{\beta}{\Delta}\right)^{1+\frac{1}{2k}}\left(\frac{\|h\|_1}{\varepsilon\eta^2}\right)^{\frac{1}{2k}}\right)$	$O\left(\frac{\|h\|_1^2}{\eta^4\varepsilon^2}\right)$

Table A2: Comparison of the ground state property estimation algorithm by Single-Ancilla LCU with other methods in terms of gate depth. We assume that H is a linear combination of L terms (strings of Pauli operators), i.e $H=\sum_{k=1}^{L}\lambda_k P_k$ such that $\beta=\sum_k|\lambda_k|$. The algorithms in the table estimate $\langle v_0|O|v_0\rangle$ to ε-additive accuracy, where $|v_0\rangle$ is the unknown ground state of H, and O is an observable which is also a linear combination of (easy to implement) unitaries, i.e. $\sum_{j=0}^{L_O}h_jO_j$, with $\|O_j\|=1$ and $\|h\|_1=\sum_{j=1}^{L_O}|h_j|$. The overall gate depth is the product of the complexities in the last two columns. We assume (i) access to a quantum state with an overlap at least η with $|v_0\rangle$, and (ii) that the ground energy of H is known to precision ε_g (See Sec. 5). We assume that a $(\beta,\lceil\log L\rceil,0)$-block encoding of H is implementable in gate depth $O(L)$ (Standard LCU and QSVT-based algorithms require access to such a block encoding). Furthermore, coherent procedures to estimate the desired expectation value require access to a block encoding of O. For any O which is an LCU, we can implement a $(\|h\|_1,\lceil\log L_O\rceil,0)$-block encoding. The gate depth of this construction is $O(L_O)$.

(Table A1), the gate depth per coherent run has a better dependence on β and Δ, but also depends on L and $1/\eta$. Clearly, there are regimes where our method (using Hamiltonian simulation by *Single-Ancilla LCU*) has a shorter gate depth per coherent run as compared to this approach. For instance, for any Hamiltonian satisfying $L\geq\beta/\Delta$, our method has a shorter gate depth by a factor of at least $O(1/\eta)$. On the other hand, the overall gate depth of this approach has a better dependence on $1/\eta$ and $1/\Delta$. However, our method requires a solitary ancilla qubit and no multi-qubit controlled operations.

- It is possible to directly estimate $\langle v_0|O|v_0\rangle$ by using standard LCU with quantum amplitude estimation (QAE). For any observable which is a linear combination of L_O terms, we can obtain a $(\|h\|_1,\log L_O,0)$-block encoding of O in circuit depth $O(L_O)$. So, overall the total number of ancilla qubits increases to $O(\log L+\log L_O+\log(\log(\|h\|_1\eta^{-1}\varepsilon^{-1})/\Delta))$. On the other hand, the gate depth of the procedure is

$$\widetilde{O}\left(\frac{\beta L\|h\|_1}{\Delta\eta\varepsilon}+\frac{\|h\|_1 L_O}{\varepsilon}\right).$$

Only $O(1)$ classical repetitions are needed. As compared to our algorithm, the over-

all complexity is lower, at the cost of exponentially increasing the gate depth (in terms of $1/\varepsilon$). Moreover, the total number of ancilla qubits required also increases substantially for this strategy.

- Without using quantum amplitude amplification or estimation, $\langle v_0|O|v_0\rangle$ can be measured by simply using standard LCU followed by repeatedly measuring some O_j sampled according to $h_j/\|h\|_1$. For each coherent run, the gate depth is $\widetilde{O}\left(\beta L/\Delta\right)$, which has a better dependence on both β and Δ, as compared to our method primarily due to the advantage of using the state-of-the-art Hamiltonian simulation procedure. However, the dependence on L ensures that our method (using Hamiltonian simulation by *Single-Ancilla LCU*) has a shorter gate depth per coherent run, for any Hamiltonian where $L > \beta/\Delta$. This advantage can be observed in several quantum chemistry Hamiltonians where typically $\beta \ll L$ [60, 86], as well as Hamiltonians in condensed matter physics such as quantum Ising Hamiltonians with long-range interactions [53]. However, the number of classical repetitions needed is quadratically better dependence on $1/\eta$ as compared to our method, which also means that the overall gate depth of this approach is generally lower. As before, our method has a better scaling in terms of the number of ancilla qubits, and the fact that our method requires no multi-qubit controlled operation.

The complexity of ground state preparation by QSVT [64] also compares similarly to our method. All the three variants require more ancilla qubits and multi-qubit controlled operations as compared to our method (using Hamiltonian simulation by *Single-Ancilla LCU*), while the overall complexity is lower. Moreover, as compared to each of the three variants, there are regimes where the gate depth per coherent run of our method is lower, despite requiring fewer ancilla qubits and no multi-qubit controlled gates. The details can be found in Table A2.

The early fault-tolerant quantum algorithm of Dong et al. [10] for ground state preparation can be leveraged to estimate $\langle v_0|O|v_0\rangle$. However, there are issues if one wants to incorporate the *Single-Ancilla LCU* Hamiltonian simulation technique into their method. This is because, the algorithm requires querying $U = e^{-iH}$, which needs to be implemented without any subnormalization factor. However, our Hamiltonian simulation algorithm implements an LCU S such that $S/\|c\|_1 \approx e^{-iH}$, where $\|c\|_1 = O(1)$. Thus, several queries to U and $U^\dagger$ would exponentially blow up the simulation cost (d queries would lead to an effective overhead of $\|c\|_1^d$). Consequently, only Trotter based methods can be suitably incorporated into this algorithm, which has been already analyzed in the article. This also means that the gate depth per coherent run of our approach (using where we use Hamiltonian simulation by *Single-Ancilla LCU*) is sub-exponentially better (in terms of $1/\varepsilon$). Additionally the gate depth also depends on L, which can be significantly larger than β in many cases. However, the overall gate depth has a better dependence on our method in terms of $1/\eta$.

Quantum Linear Systems: The detailed complexities have been summarized in Table A3. As with ground state property estimation, we assume that the quantum algorithm by Childs et al. [26], relying on *Standard LCU*, makes use of the state-of-the-art Hamiltonian simulation procedure by Low and Chuang [43]. This reduces the overall gate depth as compared to our method but increases the number of ancilla qubits and multi-qubit controlled operations.

- Standard LCU, along with controlled Hamiltonian simulation followed by amplitude

Algorithm	Variant	Ancilla	Gate depth per coherent run	Classical repetitions
Standard LCU [26] (with Hamiltonian simulation by qubitization [43])	QAA	$O\left(\log L + \log\left(\frac{\kappa\|O\|}{\varepsilon}\right)\right)$	$\widetilde{O}\left(\beta L\kappa^2\right)$	$O\left(\frac{\|h\|_1^2}{\varepsilon^2}\right)$
	QAE	$O\left(\log L_O + \log L + \log\left(\frac{\kappa\|O\|}{\varepsilon}\right)\right)$	$\widetilde{O}\left(\frac{\beta L\|h\|_1\kappa^2}{\varepsilon} + \frac{L_O\|h\|_1}{\varepsilon}\right)$	$O(1)$
	Without QAA or QAE	$O\left(\log L + \log\left(\frac{\kappa\|O\|}{\varepsilon}\right)\right)$	$\widetilde{O}\left(\beta L\kappa\right)$	$\widetilde{O}\left(\frac{\kappa^2\|h\|_1^2}{\varepsilon^2}\right)$
QSVT [45]	QAA	$O(\log L)$	$\widetilde{O}\left(\beta L\kappa^2\right)$	$O\left(\frac{\|h\|_1^2}{\varepsilon^2}\right)$
	QAE	$O\left(\log L_O + \log L\right)$	$\widetilde{O}\left(\frac{\beta L\|h\|_1\kappa^2}{\varepsilon} + \frac{\|h\|_1 L_O}{\varepsilon}\right)$	$O(1)$
	Without QAA or QAE	$O\left(\log L\right)$	$\widetilde{O}\left(\beta L\kappa\right)$	$\widetilde{O}\left(\frac{\kappa^2\|h\|_1^2}{\varepsilon^2}\right)$
Discrete adiabatic theorem [51]	Classical repetitions	$O(\log L)$	$\widetilde{O}\left(\beta L\kappa\right)$	$O\left(\frac{\|h\|_1^2}{\varepsilon^2}\right)$
	QAE	$O\left(\log L + \log L_O\right)$	$\widetilde{O}\left(\frac{\beta L\|h\|_1\kappa}{\varepsilon} + \frac{\|h\|_1 L_O}{\varepsilon}\right)$	$O(1)$
This work	Ham. Sim. by Single-Ancilla LCU	1	$\widetilde{O}\left(\beta^2\kappa^2\right)$	$\widetilde{O}\left(\frac{\|h\|_1^2\kappa^4}{\varepsilon^2}\right)$
	Ham. Sim. by $2k$-order Trotter	1	$\widetilde{O}\left(L\kappa^{1+\frac{3}{2k}}(\beta)^{1+\frac{1}{2k}}\left(\frac{\|h\|_1}{\varepsilon}\right)^{\frac{1}{2k}}\right)$	$\widetilde{O}\left(\frac{\kappa^4\|h\|_1^2}{\varepsilon^2}\right)$

Table A3: Comparison of quantum algorithms to estimate expectation values of observables with respect to the solution of quantum linear systems. We compare the complexity of the quantum algorithm by *Single-Ancilla LCU* with other methods. We assume that H is a linear combination of L terms (strings of Pauli operators), i.e $H = \sum_{k=1}^{L}\lambda_k P_k$ such that $\beta = \sum_k |\lambda_k|$ and the eigenvalues of H lie in $[-1,-1/\kappa]\cup[1/\kappa,1]$. We also assume that an initial state $|b\rangle$ can be prepared efficiently. The algorithms in the table estimate $\langle x|O|x\rangle$ to ε-additive accuracy, where $|x\rangle = H^{-1}|b\rangle / \left\|H^{-1}|b\rangle\right\|$. O is an observable which is also a linear combination of (easy to implement) unitaries, i.e. $\sum_{j=0}^{L_O} h_j O_j$, with $\left\|O_j\right\| = 1$ and $\|h\|_1 = \sum_{j=1}^{L_O}|h_j|$. The overall circuit depth is the product of the complexities in the last two columns. We assume that a $(\beta, \lceil\log L\rceil, 0)$-block encoding of H is implementable in gate depth $O(L)$ (Standard LCU and QSVT-based algorithms require access to such a block encoding). Furthermore, coherent procedures to estimate the desired expectation value require access to a block encoding of O. For any O which is an LCU, we can implement a $(\|h\|_1, \lceil\log L_O\rceil, 0)$-block encoding. The circuit depth of this construction is L_O.

amplification, prepares the state $|x\rangle$. This requires ancilla qubits to implement (a) the block encoding of H, and (b) the linear combination of Hamiltonian simulation. Overall, $O(\log L + \log\log(\kappa\|h\|_1/\varepsilon))$ ancilla qubits are needed, along with multi-qubit controlled operations. Using Hamiltonian simulation by qubitization results in the gate depth per coherent scaling as $\widetilde{O}(\beta L\kappa^2)$. Following this, $\langle x|O|x\rangle$ can be measured by sampling O_j according to $h_j/\|h\|_1$ in each run, and measuring O_j. This requires $O(\|h\|_1^2/\varepsilon^2)$ classical repetitions. As compared to our method (Table A1), the gate depth per coherent run has a better dependence on β, primarily due to the use of a more advanced simulation algorithm. However, despite this, the dependence on L ensures that for any H with $L > \beta\kappa$, our method (using Hamiltonian simulation by *Single-Ancilla LCU*) has a shorter gate depth per coherent run. The overall gate depth of this approach has a better dependence on κ (by a factor of κ^4), but the dependence on L means that there are regimes of L, β and κ for which our method also has a shorter overall gate depth. Note that our method requires only a single

ancilla qubit and no multi-qubit controlled operations.

- The coherent estimation of $\langle x|O|x\rangle$ by using quantum amplitude estimation requires accessing a block encoding of O, as defined previously. So, the total number of ancilla qubits increases to $O(\log L + \log(\kappa \|h\|_1 / \varepsilon) + \log L_O)$. On the other hand, the gate depth of the procedure is

$$\widetilde{O}\left(\frac{\beta \|h\|_1 \kappa^2 L}{\varepsilon} + \frac{\|h\|_1 L_O}{\varepsilon}\right).$$

  Only $O(1)$ classical repetitions are needed. So this is also the overall gate depth. As compared to our algorithm, the overall complexity is lower, at the cost of exponentially increasing the gate depth per coherent run (in terms of $1/\varepsilon$). Moreover, the total number of ancilla qubits required is quite high.

- Estimating $\langle x|O|x\rangle$ without using quantum amplitude amplification or estimation a requires a gate depth per coherent run of $\widetilde{O}(\kappa\beta L)$, which has a better dependence on both β and κ, as compared to our method primarily due to the advantage of using the state-of-the-art Hamiltonian simulation procedure. However, our method requires a shorter gate depth per coherent run for H satisfying $L > \beta\kappa$. The number of classical repetitions needed, given by $O(\ \|h\|_1^2 \kappa^2/\varepsilon^2)$ has a quadratically better dependence on κ as compared to our method (which also means that the overall gate depth is lower in general). However, this procedure requires $O(\log L)$ ancilla qubits and multi-qubit controlled gates.

The quantum linear systems algorithm by QSVT [45] can also estimate $\langle x|O|x\rangle$ in three ways. As before, all three variants always require more ancilla qubits and multi-qubit controlled operations, as compared to our method. Despite this, there are regimes where our method (using Hamiltonian simulation by *Single-Ancilla LCU*) has a shorter gate depth per coherent run. The overall complexity of QSVT-based approaches is however, lower. The details can be found in Table A3.

The state-of-the-art quantum linear systems algorithm [51] requires a circuit depth per coherent run that is linear in κ while $O(\ ||h||_1^2/\varepsilon^2)$ classical repetitions are needed. However, this approach still requires access to a block encoding of H, which requires $O(\log L)$ ancilla qubits. This adds an overhead of $O(L)$ to the gate depth per coherent run. Consequently, there are regimes where our method has an advantage. By using QAE, the ancilla qubit overhead is higher. The gate depth per run is exponentially worse than our method in terms of $1/\varepsilon$, but is quadratically better in terms of β and κ. For both these approaches to estimate the desired expectation value, the overall gate depth has a better dependence on κ as compared to our method.

Overall, our algorithm provides a generic exponential speedup over the best known classical algorithms. Given recent dequantization algorithms [93, 94, 95, 96], the speedup is polynomial for any H that is low-rank.

D Ground state preparation using QSVT on fully fault-tolerant quantum computers

In this section, we provide a quantum algorithm for the GSP problem for fully fault-tolerant quantum computers. The key idea is to implement the function $f(H) = e^{-tH^2}$

in the circuit model. A straightforward approach would be to use the decomposition of $f(H)$ in Lemma 14 and implement a standard LCU procedure. However, a more efficient approach would be to implement a polynomial approximation of $f(H)$ using QSVT. We begin by providing a polynomial approximation of the Gaussian operator e^{-tx^2}.

Lemma A3 (Polynomial approximation to e^{-tx^2})**.** *Suppose $x \in [-1,1]$, $\varepsilon \in [0,1/2)$ and $t \in \mathbb{R}^+$. Furthermore, suppose $d = \lceil \max\{te^2/2, \ln(2/\varepsilon)\} \rceil$. Then, there exists a polynomial $\tilde{q}_{t,d,d'}(x)$ of degree*

$$d' = \lceil \sqrt{2d \ln(4/\varepsilon)} \rceil \in O\left(\sqrt{t} \log(1/\varepsilon)\right),$$

for which the following holds

$$\sup_{x \in [-1,1]} \left| e^{-tx^2} - \tilde{q}_{t,d,d'}(x) \right| \leq \varepsilon.$$

Proof. For $x \in [-1,1]$, we assign $z = 1 - 2x^2$. Note that as $x^2 \in [0,1]$, we have $z \in [-1,1]$. Moreover, following this substitution, we need a polynomial approximation to $e^{-\frac{t}{2}(1-z)}$. Now, we define the polynomial $\tilde{q}_{t,d,d'}(x) = q_{\frac{t}{2},d,d'}(1-2x^2)$. The degree of $\tilde{q}_{t,d,d'}(x)$ is the same as that of $q_{\frac{t}{2},d,d'}(1-2x^2)$, which is d'. So, we have to bound

$$\sup_{x \in [-1,1]} \left| e^{-tx^2} - \tilde{q}_{t,d,d'}(x) \right| = \sup_{z \in [-1,1]} \left| e^{-\frac{t}{2}(1-z)} - q_{\frac{t}{2},d,d'}(z) \right|.$$

Now from Eq. (119),

$$\sup_{z \in [-1,1]} \left| e^{-\frac{t}{2}(1-z)} - q_{\frac{t}{2},d,d'}(z) \right| \leq \varepsilon,$$

for $d = \lceil \max\{te^2/2, \ln(2/\varepsilon)\} \rceil$ and $d' = \lceil \sqrt{2d \ln(4/\varepsilon)} \rceil$. □

From Lemma A3, we can use QSVT to obtain a block encoding of e^{-tH^2}, given an approximate block encoding of H. Subsequently, we shall show that this results in a robust quantum algorithm for preparing the ground state of H under the assumptions we have considered.

Lemma A4. *Let H be a Hermitian matrix with eigenvalues in $[-1,1]$ and $\varepsilon \in (0,1/2)$. Furthermore, suppose $t \in \mathbb{R}^+$ and U_H is an $(1,a,\delta)$-block encoding of H, implementable in time T_H. Also, let $d = \lceil \max\{te^2/2, \ln(4/\varepsilon)\} \rceil$ and $d' = \lceil \sqrt{2d \ln(8/\varepsilon)} \rceil$. Then, provided*

$$\delta \leq \frac{\varepsilon^2}{128d \ \ln(8/\varepsilon)},$$

we can implement an $(1, a+1, \varepsilon)$-block encoding of e^{-tH^2} in cost

$$T = O\left(T_H \sqrt{t} \log(1/\varepsilon)\right).$$

Proof. Now suppose $\tilde{H} = \left(\langle 0|^{\otimes a} \otimes I\right) U_H \left(|0\rangle^{\otimes a} \otimes I\right)$. Then, from the definition of block encoding of operators, $\left\| H - \tilde{H} \right\| \leq \delta$. Also, from Lemma A3, we can use the polynomial of degree $d' = \lceil \sqrt{2d \log(8/\varepsilon)} \rceil$ to implement $(1, a+1, \varepsilon/2)$-block encoding of $\tilde{q}_{t,d,d'}(\tilde{H})$ in cost

$$T = d' \in O\left(T_H \sqrt{t} \log(1/\varepsilon)\right).$$

The number of ancilla qubits increased by one because of the QSVT procedure. So, we have

$$\left\| e^{-tH^2} - \tilde{q}_{d,d'}\left(\tilde{H}\right) \right\| \leq \left\| e^{-tH^2} - \tilde{q}_{d,d'}\left(\tilde{H}\right) \right\| + \left\| \tilde{q}_{d,d'}\left(\tilde{H}\right) - \tilde{q}_{d,d'}\left(H\right) \right\| \tag{A37}$$

$$\leq \varepsilon/2 + 4d'\sqrt{\delta} \tag{A38}$$

$$\leq \varepsilon/2 + \varepsilon/2 = \varepsilon. \qquad [\text{ Substituting the value of } \delta \text{ and } d']. \tag{A39}$$

□

Now that we have a procedure to implement a block encoding of e^{-tH^2}, given an approximate block encoding of H, we can use this to obtain a circuit model quantum algorithm for preparing the 0-eigenstate of H. As before, let us make some assumptions on the spectrum of H. We assume that we are given a Hamiltonian H of unit norm with ground energy, λ_0 and we intend to prepare a state that is close to its ground state, $|v_0\rangle$. We assume that the gap between the ground state and the rest of the spectrum is lower bounded by Δ. We also assume that we have knowledge of E_0 such that

$$|\lambda_0 - E_0| \leq O\left(\Delta/\sqrt{\log\frac{1}{\eta\varepsilon}}\right).$$

Lemma A5. *Let $\varepsilon \in (0, 1/2)$ and H be a Hamiltonian with unit spectral norm. Furthermore, suppose we are given U_H, which is a $(1, a, \delta)$-block encoding of H, implemented in time T_H. Let $|v_0\rangle$ be the ground state of H with eigenvalue λ_0 such that the value of λ_0 is known up to precision $\varepsilon_g \in \mathcal{O}\left(\Delta/\sqrt{\log\frac{1}{\eta\varepsilon}}\right)$, where Δ is a lower bound on the spectral gap of H.*

Additionally, let us assume access to a state preparation procedure B which prepares a state $|\psi_0\rangle$ in time T_{ψ_0} such that $|\langle\psi_0|v_0\rangle| \geq \eta$. Also, let

$$\delta \leq \frac{\varepsilon^2\eta^2}{512d\,\ln\left(\frac{16}{\eta\varepsilon}\right)},$$

where, $d = \lceil \max\{te^2/2, \ln(8/\varepsilon)\} \rceil$, and

$$t > \frac{1}{2\Delta^2}\log\left(\frac{4(1-\eta^2)}{\eta^4\varepsilon^2}\right).$$

Then there exists a quantum algorithm that prepares a quantum state that is $O(\varepsilon)$-close to $|v_0\rangle$ with $\Omega(1)$ probability in cost

$$T = O\left(\frac{T_H}{\eta\Delta}\log\left(\frac{1}{\eta\varepsilon}\right) + \frac{T_{\psi_0}}{\eta}\right). \tag{A40}$$

Proof. In lemma A4, we replace ε with $\varepsilon\eta/2$ to prepare an $(1, a+1, \varepsilon\eta/2)$-block encoding of e^{-tH^2}. Furthermore, we choose

$$t \geq \frac{1}{2\Delta^2}\log\left(\frac{4(1-\eta^2)}{\eta^4\varepsilon^2}\right) = O\left(\frac{1}{\Delta^2}\log\left(\frac{1}{\eta\varepsilon}\right)\right). \tag{A41}$$

To get an $\varepsilon\eta/2$-precision in the block encoding the degree of the polynomial $\tilde{q}_{t,d,d'}(H')$ is

$$d' = \left\lceil \sqrt{2d \ln\left(\frac{16}{\eta\varepsilon}\right)} \right\rceil,$$

where $d = \left\lceil \max\{te^2/2, \ln\left(\frac{8}{\varepsilon\eta}\right)\} \right\rceil$. This yields that the precision of block encoding of H needs to be at least δ-precise where,

$$\delta \leq \frac{\varepsilon^2\eta^2}{512d \ln\left(\frac{16}{\eta\varepsilon}\right)}.$$

Thus, with cost,

$$O\left(\frac{T_H}{\Delta} \log\left(\frac{1}{\eta\varepsilon}\right) + T_{\psi_0}\right),$$

we prepare a quantum state that is $O(\varepsilon\eta/2)$-close to

$$|\eta_t\rangle = |\bar{0}\rangle \frac{e^{-tH^2}}{\sqrt{\langle\psi_0|e^{-2tH^2}|\psi_0\rangle}} |\psi_0\rangle + |\Phi^{\perp}\rangle.$$

Post-selected on obtaining $|\bar{0}\rangle$ in the first register, we obtain a quantum state that is $O(\varepsilon\eta/2)$-close to

$$|\phi\rangle = \frac{e^{-tH^2}}{\sqrt{\langle\psi_0|e^{-2tH^2}|\psi_0\rangle}} |\psi_0\rangle, \tag{A42}$$

with amplitude $\sqrt{\langle\psi_0|e^{-2tH^2}|\psi_0\rangle} = \Omega(\eta)$. Now by choosing t as in Eq. (A41), we have

$$\| |v_0\rangle - |\phi\rangle \| \leq O(\varepsilon\eta/2).$$

By the triangle inequality, this implies that the quantum state prepared is $O(\varepsilon\eta)$-close to $|v_0\rangle$ with probability $\Omega(\eta)$. So by using $O(1/\eta)$-rounds of amplitude amplification, we obtain a quantum state that is $O(\varepsilon)$-close to $|v_0\rangle$ with probability $\Omega(1)$. The overall cost will be

$$T = O\left(\frac{T_H}{\eta\Delta} \log\left(\frac{1}{\eta\varepsilon}\right) + \frac{T_{\psi_0}}{\eta}\right).$$

□

E Relationship between discrete-time and continuous-time quantum walks

The *Ancilla-free LCU* framework helped us relate between discrete and continuous-time quantum walks and their classical counterparts. Here, using block encoding and QSVT, we establish a relationship between discrete-time quantum walks and continuous-time quantum walks, from both directions. In a seminal work, Childs [58] showed that, given any Hamiltonian H, one can implement e^{-iHt} using a discrete-time quantum walk. This generates a continuous-time quantum walk on the vertices of the underlying Markov chain P. However, such a continuous-time quantum walk time evolution operator cannot be

leveraged to fast-forward continuous-time random walk dynamics. Here we show that there exists a Hamiltonian H_P (defined on the edges of P) is able to achieve this. For this, we show that the block encoding of H_P can be efficiently constructed from a block encoding of H.

The problem of obtaining a discrete-time quantum walk, given access to a continuous-time quantum walk has not been addressed in the literature. In fact, there has been very little progress towards answering this question. In this section, we show that one can establish a relationship in this direction by minor modifications to existing theorems in Ref. [44].

In order to obtain a continuous-time quantum walk from a discrete-time quantum walk, we show that given U_H, a block encoding of any H such that $U_H^2 = I$, we can obtain a Hamiltonian H_P, such that H_P^2 is a block encoding of $I - H^2$. Thus, when $H = D$, simulating H_P allows us to obtain a continuous-time quantum walk, on the edges of P, from a block encoding of H.

For the other direction, that is, to obtain a discrete-time quantum walk from a continuous-time quantum walk, we will assume that we have access to some $U = e^{iH}$. This corresponds to a continuous-time quantum walk with respect to the Hamiltonian H. From this, using Corollary 71 of Ref. [44], we will show that we can obtain a block encoding of H. Finally, we show that any block encoding of H can lead to a discrete-time quantum walk on the edges of H. We discuss each of these approaches next.

From discrete-time quantum walks to continuous-time quantum walks: We first show that given any block encoding of H, we can obtain a Hamiltonian that block-encodes $I - H^2$.

Lemma A6. *Suppose $\varepsilon \in (0,1)$, $\Pi_0 = |\bar{0}\rangle\langle\bar{0}| \otimes I$ and $R = 2\Pi_0 - I \otimes I$. Let U_H be any $(1,a,0)$-block encoding of H such that $U_H^2 = I$. Then the Hamiltonian,*

$$H_P = i[U_H, \Pi_0] \tag{A43}$$

can be constructed from one query to the (controlled) discrete-time quantum walk unitary $V = R \cdot U_H$ and its conjugate transpose. Furthermore, H_P^2 is a $(1, a+1, \varepsilon)$ block encoding of $I - H^2$.

Proof. It is easy to see that $H_P = i(V - V^\dagger)/2$. So if $W_V = |0\rangle\langle 0| \otimes e^{i\pi/2}V + |1\rangle\langle 1| \otimes e^{-i\pi/2}V^\dagger$, then $Q = (H \otimes I)\, W_V\, (H \otimes I)\, (\sigma_x \otimes I)$ is a $(1, a+1, 0)$-block encoding of H_P. Q is implemented by versions of V and $V^\dagger$ and also single Hadamard gates.

It is easy to verify that H_P is a Hamiltonian (Hermitian operator) of unit norm. To prove that H_P^2 is a $(1,a,0)$ block encoding of $I - H^2$, observe

$$\left(\langle\bar{0}| \otimes I\right) H_P^2 \left(|\bar{0}\rangle \otimes I\right) = \left(\langle\bar{0}| \otimes I\right) [\Pi_0 + U\Pi_0 U - U\Pi_0 U\Pi_0 - \Pi_0 U\Pi_0 U] \left(|\bar{0}\rangle \otimes I\right) \tag{A44}$$

$$= I + H^2 - 2H^2 = I - H^2. \tag{A45}$$

□

From a $(1, a+1, \varepsilon)$ block encoding of H_P, using QSVT, we can implement a $(1, a+3, \varepsilon)$ block encoding of e^{-itH_P} using $\Theta(t + \log(1/\varepsilon))$ queries to the controlled versions of the DTQW unitary V and its conjugate transpose [43, 45].

This implies, from a block encoding of H, we can simulate a continuous-time quantum walk on the vertices of H (by implementing e^{-iHt}) as well as on the edges of H (by implementing $e^{-iH_P t}$), requiring in both cases, $\Theta\left(t + \log(1/\varepsilon)\right)$ queries to the corresponding discrete-time quantum walk unitary.

From continuous-time quantum walks to discrete-time quantum walks: In this section, we begin by assuming that we have access to a continuous-time quantum walk evolution operator $U = e^{iH}$. The goal would be to construct a discrete-time quantum walk, given access to U. For this, first we shall show that from U, we can obtain a block encoding of H with unit sub-normalization. A good starting point is Corollary 71 of Ref. [44], which shows that it is possible to obtain a block encoding of H given access to U, provided $\|H\| \leq 1/2$. We restate the same here.

Lemma A7 (Corollary 71 of [44])**.** *Given any $U = e^{iH}$, such that H is some Hamiltonian with $\|H\| \leq 1/2$. Let $\varepsilon \in (0, 1/2]$. Then we can implement a $(1, 2, \varepsilon)$-block encoding of H with cost $O(\log 1/\varepsilon)$.*

This Lemma already gives a block encoding of H, starting from U. However, one issue here is that Lemma A7 does not work when $\|H\| = 1$. This is because, the polynomial that implements this transformation , only approximates $\arcsin(x)$ in the domain $[-1 + \delta, 1 - \delta]$, for some $\delta > 0$. For discrete-time quantum walks it is important that the sub-normalization factor of the block-encoded matrix is one. To see this, observe that Lemma A7 is effectively a $(1/2, 2, \varepsilon)$-block encoding of $H/\|H\|$. Implementing t quantum walk steps would shrink this factor to 2^{-t} which is undesirable. Moreover, in the context of quantum fast forwarding, the polynomials $p_{t,d}(x)$ and $q_{t,d,d'}(x)$ approximate x^t and $e^{t(x-1)}$ (respectively) on the entire domain $\mathcal{I} \in [-1, 1]$. However, for block-encoded matrices with normalization $\alpha > 1$, we would need to approximate these functions in $[-1/\alpha, 1/\alpha]$. Using $p_{t,d}(x/\alpha)$ or $q_{t,d,d'}(x/\alpha)$ would lead to an exponential overhead of α^t in the cost.

One way to circumvent this problem is to instead consider access to the continuous-time evolution operator $U = e^{iH/2}$, where now $\|H\| = 1$. Using Lemma A7, we obtain a $(2, 2, \varepsilon/2)$-block encoding of H in cost $O(\log(1/\varepsilon))$. At this stage, we can make use of the procedure of uniform singular value amplification [Theorem 17 of Ref. [45]], which amplifies all the singular values (in our case the eigenvalues) of a block-encoded matrix. This allows us to obtain a $(1, 3, \varepsilon)$ block encoding of H as we prove next.

Theorem A8 (From continuous-time quantum walks to discrete-time quantum walks)**.** *Suppose $\varepsilon \in (0, 1)$ and H is a Hermitian operator. Suppose we have access to $U = e^{iH/2}$. Then there exists a procedure that implements a $(1, 3, \varepsilon)$ - block encoding of H in cost $O\left(\frac{1}{\varepsilon}\log(1/\varepsilon)\right)$.*

Proof. From U, we obtain U_H, which is a $(2, 2, \delta)$ - block encoding of H in cost $O(\log(1/\delta))$, using Lemma A7, for any $\delta \leq \varepsilon/2$. Then, we use the uniform spectral amplification theorem [Theorem 17 of [45]]. In Theorem 17 of [45], set $\gamma = 2(1 - \varepsilon)$. This gives us a $(1, 3, \varepsilon)$ - block encoding of H in cost $O(\frac{1}{\varepsilon}\log(1/\varepsilon))$. □

Thus given access to a continuous-time quantum walk $U = e^{iH/2}$, we can obtain U_H, which is a block encoding of H. Then, if $U_H^2 = I$, following [32], it is possible to show that if $R = (2\,|\bar{0}\rangle\langle\bar{0}| - I) \otimes I$ is a reflection operator, then $V = R.U_H$ can generate a

discrete-time quantum walk on the edges of H, as V^t is a block encoding of $T_t(H)$. Thus a discrete-time quantum walk can be implemented given access to the continuous-time quantum walk evolution operator $U = e^{-iH/2}$. However, the condition $U_H^2 = I$ need not be satisfied in general. However, it is easy to show that if U_H is any block encoding of H, the unitary $V = R.U_H^\dagger.R.U_H$ is a block encoding of the Chebyshev polynomial $T_2(H)$. Then V^t is a block encoding of $T_{2t}(H)$, similar to the standard discrete-time quantum walk operator.

Let us now discuss the issue of fast-forwarding this block encoding of H. For this we need to show that given any block encoding of H, we can have a quantum walk that fast forwards H^t. We shall prove this next. Ours is a slightly more general result as compared to Ref. [42] in that (a) it works for both even and odd t, and (b) it is robust: provides the precision with which U_H approximates the block encoding of H. The later estimate is crucial for highlighting the limitations of fast-forwarding discrete-time quantum walks when given access to the time evolution ocontinuous-time quantum walk .

Lemma A9. *Suppose $\varepsilon \in (0,1)$ and we have access to U_H, which is a $(1,a,\delta)$-block encoding of a Hamiltonian H such that $\|H\| = 1$. Then, provided*

$$\delta \leq \frac{\varepsilon^2}{128t\ \ln(8/\varepsilon)},$$

for any $t \in \mathbb{N}$, we can implement a $(1, O(a + \log t + \log\log(1/\varepsilon)), O(\varepsilon))$-block encoding of H^t in cost $O\left(\sqrt{t\log(1/\varepsilon)}\right)$.

Proof. We will implement a $(1, a+1, \varepsilon)$ block encoding of H^t by separating out the cases where t is even or odd. When t is even, we implement H^t by approximating it with the polynomial defined in Eq. (102). They are guaranteed to be ε-close following Lemma 19. The odd case also follows through via similar arguments.

Let U_H be a $(1,a,0)$-block encoding of H'. Then $\|H - H'\| \leq \delta$. Now the unitary $V = RU_H^\dagger RU_H$ is a $(1,a,0)$-block encoding of $T_2(H')$. We will use LCU to implement the polynomial $p_{t,d}(H')$ defined in Eq. (102). The degree of the polynomial is chosen to be $d = \lceil\sqrt{2t\ln(8/\varepsilon)}\rceil$, which ensures (from Lemma 19) that $\left\|x^t - p_{t,d}(x)\right\| \leq \varepsilon/4$. Consider the unitary Q such that

$$Q\left|\bar{0}\right\rangle = \frac{1}{\sqrt{\alpha}}\sum_{l=0}^{d/2}\sqrt{c_l}\left|l\right\rangle, \tag{A46}$$

where,

$$c_l = \begin{cases} 2^{1-t}\binom{t}{\frac{t}{2}+l}, & l > 0 \\ 2^{-t}\binom{t}{t/2}, & l = 0, \end{cases} \tag{A47}$$

and $\alpha = \|c\|_1$, where $c = (c_0, \cdots, c_{d/2})$. Also, define the controlled unitary

$$W = \sum_{j=0}^{d/2}\left|j\right\rangle\left\langle j\right| \otimes V^j,$$

where $V = RU_H^\dagger RU_H$. Then, it is easy to see, using LCU that the unitary $\widetilde{W} = (Q^\dagger \otimes I)W(Q \otimes I)$ is a $(\alpha, a + \lceil\log_2 d\rceil - 1, 0)$-block encoding of $p_{t,d}(H')$. That is,

$$\left(\left\langle\bar{0}\right| \otimes I\right)\widetilde{W}\left(\left|\bar{0}\right\rangle \otimes I\right) = \frac{p_{t,d}(H')}{\alpha}, \tag{A48}$$

where α is obtained by observing that for any $x \in [-1,1]$

$$\alpha = \left| x^t - \sum_{l=d/2+1}^{t} 2^{1-t} \binom{t}{t/2+l} \right| \tag{A49}$$

$$\geq 1 - \left| \sum_{l=d/2+1}^{t} 2^{1-t} \binom{t}{t/2+l} \right| \tag{A50}$$

$$\geq 1 - \varepsilon/4 \qquad [\text{ From Lemma 19 }]. \tag{A51}$$

Now by using triangle inequality we obtain,

$$\left\| H^t - p_{t,d}(H')/\alpha \right\| \leq \left\| H^t - p_{t,d}(H') \right\| + (1-\alpha) \left\| p_{t,d}(H')/\alpha \right\| \tag{A52}$$

$$\leq \varepsilon/4 + \left\| H^t - p_{t,d}(H) \right\| + \left\| p_{t,d}(H) - p_{t,d}(H') \right\| \tag{A53}$$

$$\leq \varepsilon/4 + \varepsilon/4 + 4d\sqrt{\delta} \qquad [\text{ From Lemma 5 }] \tag{A54}$$

$$\leq \varepsilon/2 + \varepsilon/2 \qquad \left[\text{As } \delta \leq \frac{\varepsilon^2}{64d^2} \right] \tag{A55}$$

$$\leq \varepsilon. \tag{A56}$$

Now for odd t, we will modify the quantum walk unitary slightly. Let $t = 2k+1$ for some $k \in \mathbb{N}$. Note that

$$\left| x^{2k+1} - x.p_{2k,d}(x) \right| \leq |x| \,. \left| x^{2k} - p_{2k,d}(x) \right| \leq \varepsilon. \tag{A57}$$

We already know that there exists a unitary $\widetilde{W}$ which is a $(\alpha, a + \lceil \log_2 d \rceil - 1, \varepsilon)$-block encoding of H^{2k}, which can be implemented with cost $O(\sqrt{2k \log(1/\varepsilon)})$. Note that U_H is a $(1, a, \delta)$-block encoding of $H = T_1(H)$. Furthermore, let us define $b = a + \lceil \log_2 d \rceil - 1$ to be the number of ancillary qubits required to implement the unitary $\tilde{W}$. We will show that the unitary $(I_b \otimes U_H)(I_a \otimes \widetilde{W})$ is a $(1, a+b, O(\varepsilon))$ block encoding of H^{2k+1}.

$$\left\| H^{2k+1} - (\langle \bar{0}|^{\otimes a+b} \otimes I)(I_b \otimes U_H)(I_a \otimes \widetilde{W})(|\bar{0}\rangle^{\otimes a+b} \otimes I) \right\| \tag{A58}$$

$$= \left\| H^{2k+1} - \underbrace{(\langle \bar{0}|^{\otimes a} \otimes I) U_H (|\bar{0}\rangle^{\otimes a} \otimes I)}_{=H'} \underbrace{(\langle \bar{0}|^{\otimes b} \otimes I) \widetilde{W} (|\bar{0}\rangle^{\otimes b} \otimes I)}_{=p_{2k,d}(H')/\alpha} \right\| \tag{A59}$$

$$\leq \left\| \left(H - H' \right) H^{2k} \right\| + \left\| H' \right\| \left\| H^{2k} - p_{2k,d}(H')/\alpha \right\| \tag{A60}$$

$$\leq \delta + \varepsilon = O(\varepsilon). \tag{A61}$$

□

Having proven this, the next question we ask is: Can the block encoding of H obtained from the time evolution operator $U = e^{-iH/2}$ (Theorem A8) be used to fast-forward discrete-time random walks? We argue here that some issues still remain. For instance, from Lemma A9, we can see that the precision δ required in the block encoding of H is $\widetilde{O}(\varepsilon^2/t)$. Theorem A8 implies that to implement a block encoding of H, from U would require $\widetilde{O}(t/\varepsilon^2)$ cost. Thus, the advantage of quantum fast-forwarding would be lost.

In order to avoid this, we need a polynomial of degree t that approximates the monomial $(2x)^t$ in the domain $\mathcal{D} := \left[-\frac{1}{2}(1-1/t), \frac{1}{2}(1-1/t)\right]$. The existence of such a polynomial $P(x)$ of degree $O(t\log(1/\varepsilon))$ can indeed be guaranteed by Corollary 66 of [44]. For this set $f(x) = (2x)^t, x_0 = 0$, $r = 1/2$ and $\delta = 1/t$ in the corollary. We have not been able to find an explicit construction of this polynomial and leave it open for future work. Expressing $P(x)$ as a linear combination of Chebyshev polynomials, we would obtain an ε-approximation of $(2x)^t$ in $\mathcal{D}$ (having $\sqrt{t}$ terms). Given access to $U = e^{iH/2}$, we obtain a $(2, 2, \varepsilon)$ - block encoding of H using Lemma A7. We can then directly apply QSVT directly to implement the polynomial $P(x)$ on this block-encoded Hamiltonian. This would allow us to fast-forward discrete-time quantum walks, starting from continuous-time quantum walks. We leave open the explicit construction of such a polynomial, for future work.