

State preparation boosters for early fault-tolerant quantum computation

Guoming Wang¹, Sukin Sim², and Peter D. Johnson²

¹Zapata Computing Canada Inc., 25 Adelaide St E, Suite 1500, Toronto, ON M5C 3A1, Canada

²Zapata Computing Inc., 100 Federal Street, 20th Floor, Boston, MA 02110, USA

Quantum computing is believed to be particularly useful for the simulation of chemistry and materials, among the various applications. In recent years, there have been significant advancements in the development of near-term quantum algorithms for quantum simulation, including VQE and many of its variants. However, for such algorithms to be useful, they need to overcome several critical barriers including the inability to prepare high-quality approximations of the ground state. Current challenges to state preparation, including barren plateaus and the high-dimensionality of the optimization landscape, make state preparation through ansatz optimization unreliable. In this work, we introduce the method of ground state boosting, which uses a limited-depth quantum circuit to reliably increase the overlap with the ground state. This circuit, which we call a booster, can be used to augment an ansatz from VQE or be used as a stand-alone state preparation method. The booster converts circuit depth into ground state overlap in a controllable manner. We numerically demonstrate the capabilities of boosters by simulating the performance of a particular type of booster, namely the Gaussian booster, for preparing the ground state of N_2 molecular system. Beyond ground state preparation as a direct objective, many quantum algorithms, such as quantum phase estimation, rely on high-quality state preparation as a subroutine. Therefore, we foresee ground state boosting and similar methods as becoming essential algorithmic components as the field transitions into using early fault-tolerant quantum computers.

Guoming Wang: guoming.wang@zapatacomputing.com

Sukin Sim: sukin.sim@gmail.com

Peter D. Johnson: peter@zapatacomputing.com

1 Introduction

Quantum computing holds the promise to solve previously intractable problems in quantum chemistry and materials [1, 2]. Essential to this promise is the ability of a quantum computer to efficiently encode and process wave functions that represent the quantum system of interest. This system is encoded on the quantum computer via a qubit representation of the Hamiltonian. An important step of many quantum algorithms for quantum chemistry and materials is to prepare on the quantum computer an approximation to the ground state of this Hamiltonian [1, 3]. The performance of many quantum algorithms relies on the performance of the ground state preparation step [4, 5, 6]. In the realm of “far-term” quantum algorithms, the runtime of quantum phase estimation and related techniques scales as $O(1/p^2)$, where p is the quality, or fidelity, of the input ground state approximation [5, 6]. Improved methods [7] reduce this runtime to $O(1/p)$ and recent work [14] improves this runtime scaling to $O(1/\sqrt{p})$ at the cost of a $O(1/\sqrt{p})$ factor increase in circuit depth. More “near-term” quantum algorithms, such as the variational quantum eigensolver (VQE), aim to estimate the ground state energy by directly preparing an approximation to the ground state; the quality of the energy estimate is limited by the quality of the ground state preparation [8, 9]. The benefit of near-term quantum algorithms like VQE is that they have the potential to prepare good approximations to the ground state, while using relatively little circuit depth. The relatively low circuit depth limits the accrual of error in the state preparation task. While such techniques may serve as a good “head start” in approximate ground state preparation, recent numerical studies suggest that the approximations achievable with these techniques are insufficient for outperforming state-of-the-art classical methods [10, 11, 12].

This challenge motivates the use of further quantum circuitry, beyond heuristic parameterized quantum circuits to improve, or “boost”, the ground state preparation subroutine in near-term quantum algorithms. Such techniques are likely needed to achieve quantum advantage for chemistry with early fault-tolerant quantum computers. While there do exist methods for improving the approximation of ground states [13, 14] and excited states [15], these methods are designed to be performant with an idealized quantum computer. With early fault-tolerant quantum computers, we should design algorithms which accommodate the build-up of error during the quantum computation.

This motivates the need for quantum algorithms which reliably increase the overlap with the ground state of a target Hamiltonian using limited circuit depth. One such method is developed in Ref. [16]. This approach, suited for few-body interacting Hamiltonians, gives a low-depth quantum circuit which ensures a fixed energy reduction. However, it is not clear how to further reduce the energy or increase the ground state overlap beyond this initial amount. If a particular energy reduction or overlap is required, the reduction ensured by this method might not be sufficient. Furthermore, the performance of the method is inversely related to the locality of the Hamiltonian and its degree of interaction, which limits its application to quantum chemistry Hamiltonians. Nevertheless, the method may provide a good input to additional state

preparation methods. Another such method is the quantum imaginary time evolution (QITE) algorithm, introduced in Ref. [17]. QITE aims to prepare an approximation to a low-temperature thermal state of a Hamiltonian. By choosing a sufficiently low temperature, the output state is a good approximation to the ground state. The advantage of the QITE algorithm compared to previous imaginary time evolution algorithms [18] is that the performance of the method does not rely on the quality of a proposed ansatz circuit. Rather, this iterative method constructs the best approximation to each imaginary time evolution by making local tomographic estimates of the current state. One disadvantage of the QITE algorithm is that its favorable runtime requires the system of interest to have local interactions; many systems of interest, including the electronic structure of molecules are most-conveniently represented with non-local interactions. Another disadvantage of the QITE algorithm is that it is unclear how the method will perform when subject to error in the quantum circuit. A worry is that error will accrue unfavorably through the iterative process. The reliability of the method in the presence of error for large-scale calculations is still an open question.

More recently, Amaro et al. [19] proposed a near-term ground state filtering method. Drawing inspiration from the ground state filtering of Ref. [5], instead of applying the filter operator to the initial state, they emulate the action of the filter operator by using a parameterized quantum circuit (PQC) and optimizing the parameters to minimize the Euclidean distance between the PQC-prepared state and the filter-applied state. They carry this out in an iterative fashion until termination criteria are reached. They investigate several filters (function types) and demonstrate the performance with QAOA. While this work does use similar concepts of filtering that we make use of, their method relies on a heuristic parameterized circuit optimization. Accordingly, this filter does not admit performance guarantees and is likely to suffer from the same reliability issues as VQE.

In this context our work addresses the following question: *using limited circuit depth, how can we increase the overlap with the ground state in a way that works efficiently for all Hamiltonians and gives reliable performance?* We develop the method of *state preparation boosting* for reliably improving the overlap with a target state while using quantum circuits of limited circuit depth. In contrast to the optimization of parameterized quantum circuits, state preparation boosters reliably convert precious quantum circuit depth into overlap with the target state. Moreover, ground state boosters can be appended to an already-optimized VQE circuit to reliably increase its overlap with the target state. Alternatively, one could append a state preparation booster to a parameterized quantum circuit throughout the VQE optimization.

In Section 2 we introduce the concept of state preparation boosters and introduce several methods for designing them. In Section 3 we analyze the performance of boosters for a small molecular system and show that the boosters reliably convert circuit depth into overlap with the ground state. We conclude with a discussion on future work for state preparation boosters and an outlook on their use in helping to achieve quantum advantage. Several important technical details are included in the appendix.

While preparing this paper, we came across a recent work [20] presenting the Gaus-

sian filter method similar to a particular type of booster investigated in this work. They present a hybrid quantum-classical method that estimates the ground state energy but does not prepare the ground state. Their fully quantum method, which prepares the ground state, is well-suited for qubit-qumode quantum processors, but the authors do not address how to prepare the ground state on a qubit-based quantum computer, which is the focus of our work. Furthermore, our work extends beyond Gaussian filters or boosters; we present a general framework that can be used to design low-depth boosters for preparing ground states.

After an earlier version of this paper was posted on arXiv, Dong et al. [21] proposed two algorithms for approximately preparing the ground state of a given Hamiltonian on early fault-tolerant quantum computers. They aim to achieve a fidelity close to one in the output state, and hence the costs of their algorithms are relatively high. Precisely, one of their algorithms has query depth $\tilde{O}(1/\Delta)$ and query complexity $\tilde{O}(1/(\Delta\gamma^2))$, and the other has query depth $\tilde{O}(1/(\Delta\gamma))$ and query complexity $\tilde{O}(1/(\Delta\gamma))$, where Δ and γ are the spectral gap and initial ground state overlap, respectively. Although these algorithms could be modified to have lower circuit depth at the cost of producing a worse approximation ¹, this idea is not explored in their paper.

In contrast, here we aim to improve heuristic ground state preparation methods like VQE by using limited amount of resources. Our circuit depth can be tuned based on our need and hardware constraints, and using deeper circuits leads to steady increase of the ground state overlap. Although our method can be also used to prepare an extremely-accurate approximation of the ground state, this requires quite deep circuits and is not the main focus of the paper. We leave it as future work to compare the resource costs of our scheme and the one of Ref. [21].

2 Constructing state preparation boosters

In this section, we describe the methodology of state preparation boosters for improving ground state preparation. Our basic idea is that, given any ansatz circuit for approximately preparing the ground state of a Hamiltonian H , we compose it with a shallow circuit that implements a function of H , so that the extended circuit yields a state that has larger overlap with the ground state of H . The appended operation, denoted by $f(H)$, suppresses the high-energy eigenstates of H and hence effectively boosts the low-energy ones for any input state. Thus we refer to these operations as *boosters*.

Formally, let $H = \sum_{j=1}^D \lambda_j |\lambda_j\rangle \langle \lambda_j|$ be a Hamiltonian with eigenvalues λ_j 's and

¹Ref. [21] uses a technique called quantum eigenvalue transformation of unitary matrices (QET-U) which enables the implementation of certain polynomials of $\cos(H/2)$, where H is the Hamiltonian of interest. For ground state preparation, the polynomial approximates a threshold function so that the QET-U circuit approximately implements a projective measurement onto the low-energy and high-energy subspaces of H . The depths of the resulting circuits are proportional to the degree of this polynomial. One might use a different polynomial with lower degree to construct a shallower circuit which generates a worse approximation of the ground state.

eigenstates $|\lambda_j\rangle$'s, where $a \leq \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_D \leq b$ for some $a, b \in \mathbb{R}$. Suppose $|\psi\rangle = \sum_{j=1}^D \mu_j |\lambda_j\rangle$ is the state produced by an ansatz circuit such that $\mu_1 \neq 0$. We will find a function $f : \mathbb{R} \rightarrow \mathbb{C}$ such that $|f|$ decreases monotonically on the interval $[a, b]$. Then performing the operation $f(H)$ on the state $|\psi\rangle$ yields the unnormalized state

$$f(H)|\psi\rangle = \sum_{j=1}^D \mu_j f(\lambda_j) |\lambda_j\rangle, \quad (1)$$

whose normalized version is

$$\frac{f(H)|\psi\rangle}{\|f(H)|\psi\rangle\|} = \sum_{j=1}^D \mu'_j |\lambda_j\rangle = \frac{1}{\sqrt{Z}} \sum_{j=1}^D \mu_j f(\lambda_j) |\lambda_j\rangle, \quad (2)$$

where $Z = \sum_{j=1}^D |\mu_j f(\lambda_j)|^2$ and $\mu'_j = \mu_j f(\lambda_j) / \sqrt{Z}$. By the assumption on f , we have $\left| \frac{\mu_i f(\lambda_i)}{\mu_j f(\lambda_j)} \right| \geq \left| \frac{\mu_i}{\mu_j} \right|, \forall i \leq j$. It follows that $|\mu'_j| \geq |\mu_j|$ for $j = 1, 2, \dots, k$, for some $1 \leq k \leq D$. In other words, the amplitudes of the lower-energy eigenstates of H in $\frac{f(H)|\psi\rangle}{\|f(H)|\psi\rangle\|}$ are larger than their counterparts in $|\psi\rangle$. In particular, the overlap between $\frac{f(H)|\psi\rangle}{\|f(H)|\psi\rangle\|}$ and the ground state of H is larger than the one for $|\psi\rangle$. We note that although monotonicity can ensure some degree of performance, a non-monotonic booster may be favorable if it can be implemented with shorter-depth quantum circuits. For example, the ‘‘truncated Gaussian boosters’’ that we explore later are technically non-monotonic (due to the truncation), but their corresponding functions have steeply decreasing envelopes, making them nonetheless performant.

Although in principle every function f that satisfies the monotonicity condition leads to a booster, the cost of implementing the operation $f(H)$ vary significantly for different f 's. Furthermore, in general, $f(H)$ is non-unitary and hence we cannot implement it with certainty. So we need to choose f carefully to achieve a balance among three factors:

- To what extent can $f(H)$ increase the overlap with the ground state (or the low-energy eigenstates) of H ?
- What is the depth of the circuit for implementing $f(H)$?
- What is the probability of this circuit successfully implementing $f(H)$?

Before addressing this issue in Section 2.2, we first describe a method to physically implement $f(H)$ once f is determined.

2.1 Implementing the booster operation

Suppose the booster function f is known. We implement the booster operation $f(H)$ by obtaining a Fourier approximation of f and applying the linear combination of unitaries

(LCU) method to it. Specifically, let \hat{f} be the Fourier transform of f , i.e.

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x) e^{-i2\pi x \xi} dx, \quad (3)$$

$$f(x) = \int_{-\infty}^{\infty} \hat{f}(\xi) e^{i2\pi x \xi} d\xi. \quad (4)$$

We convert the RHS of Eq. (4) into a finite sum by truncating the integral to the region $[-T, T]$ for some $T > 0$ and then discretizing the integral on that region. Specifically, let

$$f_T(x) = \int_{-T}^T \hat{f}(\xi) e^{i2\pi x \xi} d\xi, \quad (5)$$

and for any $N \in \mathbb{Z}^+$, let

$$f_{T,N}(x) = \frac{T}{N} \sum_{j=-N}^{N-1} \hat{f}(\xi_j) e^{i2\pi x \xi_j}, \quad (6)$$

where $\xi_j = (j + 1/2)\frac{T}{N}$ for $j = -N, -N + 1, \dots, N - 1$. We would like to choose some T and N such that $f \approx f_T \approx f_{T,N}$. Note that

$$\epsilon_T := \max_{x \in \mathbb{R}} |f(x) - f_T(x)| \quad (7)$$

$$= \max_{x \in \mathbb{R}} \left| \int_{-\infty}^{-T} \hat{f}(\xi) e^{i2\pi x \xi} d\xi + \int_T^{\infty} \hat{f}(\xi) e^{i2\pi x \xi} d\xi \right| \quad (8)$$

$$\leq \int_{-\infty}^{-T} |\hat{f}(\xi)| d\xi + \int_T^{\infty} |\hat{f}(\xi)| d\xi. \quad (9)$$

We will focus on the case where $|\hat{f}|$ decays exponentially on \mathbb{R} , i.e., there exists constant $c > 1$ such that $|\hat{f}(\xi)| = O(c^{-|\xi|})$ as $\xi \rightarrow \pm\infty$. Then Eq. (9) implies that $\epsilon_T = O(c^{-T})$ as $T \rightarrow +\infty$. As a consequence, choosing some $T = O(\log(1/\epsilon))$ ensures that $\max_{x \in \mathbb{R}} |f(x) - f_T(x)| \leq \epsilon$, for arbitrary $\epsilon > 0$. Meanwhile, in Appendix B, we show that when f satisfies certain (weak) conditions, choosing some $N = \tilde{O}(1/\epsilon)$ ensures that $\max_{x \in [0,1]} |f_T(x) - f_{T,N}(x)| \leq \epsilon$, for arbitrary $\epsilon > 0$. Then for such T and N , we have

$$\left| f(x) - \frac{T}{N} \sum_{j=-N}^{N-1} \hat{f}(\xi_j) e^{i2\pi x \xi_j} \right| \leq 2\epsilon, \quad \forall x \in [0, 1]. \quad (10)$$

This gives a Fourier approximation of f on the interval $[0, 1]$, which is sufficient for our purpose.

Now suppose we have a Fourier approximation of the booster function f :

$$f(x) \approx \sum_{j=1}^K \alpha_j e^{it_j x} \quad (11)$$

for some $\alpha_j \in \mathbb{C} \setminus \{0\}$ and $t_j \in \mathbb{R}$. Then we have a Fourier approximation of the booster operation $f(H)$ as well:

$$f(H) \approx \sum_{j=1}^K \alpha_j e^{it_j H}. \quad (12)$$

That is, $f(H)$ is close to a linear combination of the unitary operations $e^{it_j H}$'s. This prompts us to use the LCU method to implement $f(H)$ approximately and probabilistically. Specifically, suppose we want to implement a linear operation $A = \sum_{j=1}^K \alpha_j U_j$ on a Hilbert space \mathcal{H} , where $\alpha_j \in \mathbb{C} \setminus \{0\}$ and U_j is a unitary operation on \mathcal{H} for each j . The LCU method requires a K -dimensional ancilla and a unitary operation V on the ancilla such that

$$V |0\rangle = \frac{1}{\sqrt{|\vec{\alpha}|_1}} \sum_{j=1}^K \sqrt{|\alpha_j|} |j\rangle, \quad (13)$$

where $|\vec{\alpha}|_1 := \sum_{j=1}^K |\alpha_j|$. Moreover, let O be a unitary operation on the joint system such that

$$O = \sum_{j=1}^K |j\rangle\langle j| \otimes \frac{\alpha_j}{|\alpha_j|} U_j. \quad (14)$$

Then a direct calculation verifies that

$$\langle 0| (V^\dagger \otimes I) O (V \otimes I) |0\rangle |\psi\rangle = \frac{A|\psi\rangle}{|\vec{\alpha}|_1} \quad (15)$$

for all $|\psi\rangle \in \mathcal{H}$. That is, after applying V on the ancilla, O on the joint system and V^\dagger on the ancilla, we measure the ancilla in the standard basis, and if the measurement outcome corresponds to $|0\rangle$, the state of the main system becomes $\frac{A|\psi\rangle}{\|A|\psi\rangle\|}$, and this event happens with probability $\frac{\|A|\psi\rangle\|^2}{|\vec{\alpha}|_1^2}$. Figure 1 illustrates the quantum circuit for the LCU method.

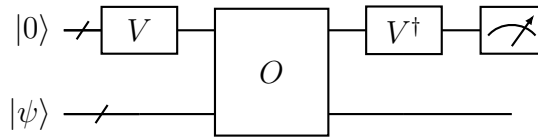


Figure 1: The LCU circuit for implementing a linear operation $A = \sum_{j=1}^K \alpha_j U_j$ probabilistically. Here V and O are given by Eqs. (13) and (14) respectively.

Now we apply the LCU method to Eq. (12) in which α_j and t_j are given by Eq. (10). Then we obtain a quantum circuit for implementing the booster operation $f(H)$ approximately and probabilistically. This circuit uses $O(\log N)$ ancilla qubits and needs

to evolve the Hamiltonian H up to time $O(T)$. When the LCU procedure succeeds, it produces the quantum state

$$\frac{f_{T,N}(H) |\psi\rangle}{\|f_{T,N}(H) |\psi\rangle\|} \approx \frac{f_T(H) |\psi\rangle}{\|f_T(H) |\psi\rangle\|} \approx \frac{f(H) |\psi\rangle}{\|f(H) |\psi\rangle\|} \quad (16)$$

and this event happens with probability

$$p_{succ}(f_{T,N}) := \frac{\langle \psi | f_{T,N}^\dagger(H) f_{T,N}(H) | \psi \rangle}{\left(\frac{T}{N} \sum_{j=-N}^{N-1} \left| \hat{f} \left((j + 1/2) \frac{T}{N} \right) \right| \right)^2} \quad (17)$$

$$\approx \frac{\langle \psi | f_T^\dagger(H) f_T(H) | \psi \rangle}{\left(\int_{-T}^T |\hat{f}_T(\xi)| d\xi \right)^2} \quad (18)$$

$$\approx \frac{\langle \psi | f^\dagger(H) f(H) | \psi \rangle}{\left(\int_{-\infty}^{\infty} |\hat{f}(\xi)| d\xi \right)^2}. \quad (19)$$

In particular, if \hat{f} is real and even, then f is also real and even. In this case, if \hat{f} is also non-negative, then the denominator on the RHS of Eq. (19) can be simplified by noting that

$$\int_{-\infty}^{\infty} \hat{f}(\xi) d\xi = f(0). \quad (20)$$

Consequently, the success probability of the LCU method in this case is

$$p_{succ}(f_{T,N}) \approx \frac{\langle \psi | f^2(H) | \psi \rangle}{f^2(0)}. \quad (21)$$

2.2 Designing the booster function

It remains to design the booster function f . To this end, we need certain prior information about the eigenvalues of the Hamiltonian H and the overlaps between the input state $|\psi\rangle$ and the eigenstates of H . Specifically, we assume without loss of generality that H has eigenvalues between 0 and 1². Furthermore, to facilitate the construction of f , we use models for the spectrum of H and the overlaps between H 's eigenstates and $|\psi\rangle$. Specifically, we approximate H as

$$H = \int_0^1 z 1_S(z) |z\rangle \langle z| dz, \quad (22)$$

²If H has eigenvalues between λ_{\min} and λ_{\max} , then we transform it into $\bar{H} := (H - \lambda_{\min}I)/(\lambda_{\max} - \lambda_{\min})$ which shares the same eigenstates with H , and apply our method to \bar{H} .

where $S \subseteq [0, 1]$ consists of H 's eigenvalues, $\{|z\rangle : z \in S\}$ are orthonormal states, 1_S is the indicator function of S , and we approximate $|\psi\rangle$ as

$$|\psi\rangle = \int_0^1 \psi(z) |z\rangle dz, \quad (23)$$

and $q(z) := |\psi(z)|^2$ is a probability distribution on $[0, 1]$ such that $q(z) = q(z) 1_S(z)$ (i.e. $q(z) = 0$ for all $z \notin S$).

In practice, we often deal with quantum systems with finite-dimensional Hilbert spaces. For such a system, H has a finite number of eigenvalues and eigenstates, and 1_S should be replaced by a sum of Dirac delta functions, and $q(z)$ should be a discrete distribution. However, if we treat $q(z)$ this way, many properties of $f(H)|\psi\rangle$ that we are interested in will be expressed as functions of a discrete sum over exponentially many terms which is difficult to compute. To mitigate this issue, we approximate $q(z)$ with a continuous distribution so that those properties of $f(H)|\psi\rangle$ can be approximated by functions of a continuous integral which is often easier to calculate. This makes the design of the booster function f more convenient, as will be seen later.

Let us consider the impact of the booster operation $f(H)$ on the input state $|\psi\rangle$. Applying $f(H)$ on $|\psi\rangle$ yields the unnormalized state

$$f(H)|\psi\rangle = \int_0^1 f(z)\psi(z) |z\rangle dz, \quad (24)$$

whose squared norm is

$$\|f(H)|\psi\rangle\|^2 = \langle\psi| f^\dagger(H)f(H)|\psi\rangle = \int_0^1 |f(z)|^2 q(z) dz. \quad (25)$$

The expectation of H with respect to $f(H)|\psi\rangle$ is

$$\langle\psi| f^\dagger(H)Hf(H)|\psi\rangle = \int_0^1 z|f(z)|^2 q(z) dz. \quad (26)$$

Thus, the energy of the normalized state $\frac{f(H)|\psi\rangle}{\|f(H)|\psi\rangle}$ with respect to H is

$$E(f) := \frac{\langle\psi| f^\dagger(H)Hf(H)|\psi\rangle}{\langle\psi| f^\dagger(H)f(H)|\psi\rangle} = \frac{\int_0^1 z|f(z)|^2 q(z) dz}{\int_0^1 |f(z)|^2 q(z) dz}. \quad (27)$$

Meanwhile, for arbitrary $\lambda \in [0, 1]$, let

$$P_{\leq\lambda} = \int_0^\lambda 1_S(z) |z\rangle\langle z| dz \quad (28)$$

be the projection operator onto the subspace spanned by the eigenstates of H whose energies are at most λ . We are interested in the total overlap between the state $\frac{f(H)|\psi\rangle}{\|f(H)|\psi\rangle}$ and such eigenstates of H , i.e.

$$O_{\leq\lambda}(f) := \frac{\langle\psi|f^\dagger(H)P_{\leq\lambda}f(H)|\psi\rangle}{\langle\psi|f^\dagger(H)f(H)|\psi\rangle} = \frac{\int_0^\lambda |f(z)|^2 q(z) dz}{\int_0^1 |f(z)|^2 q(z) dz}, \quad (29)$$

where in the second step we use the fact that $q(z) = q(z)1_S(z)$ for all $z \in [0, 1]$. One can see that if $|f|$ decreases monotonically on $[0, 1]$, then $E(f) \leq \langle\psi|H|\psi\rangle$ and $O_{\leq\lambda}(f) \geq \langle\psi|P_{\leq\lambda}|\psi\rangle$. In other words, $f(H)$ reduces the energy of the input state with respect to H while increasing its overlap with the low-energy eigenstates of H .

Ideally, we want $|f|$ to decay as fast as possible on the interval $[0, 1]$, so that the high-energy eigenstates of H get suppressed to a large extent. On the other hand, steep $|f|$ often means that $f \approx f_T$ only for large T , and hence it is expensive to implement $f(H)$ using the LCU method. Furthermore, we need to make sure that $f(H)$ can be realized with sufficiently high probability, as it is non-unitary and cannot be realized with certainty. Our goal is to find a function f that achieves the balance among these factors.

Let us first mention a strategy that attempts to directly optimize $f_{T,N}$ (as defined in Eq. (6)). Recall that the LCU method allows us to implement $f_{T,N}(H)$ probabilistically. Given prior information S and $q(x)$ about the Hamiltonian H and the input state $|\psi\rangle$, as well as T (which affects the maximal evolution time of H in the LCU circuit) and N (which determines the number of ancilla qubits in the LCU circuit), we find a high-quality $f_{T,N}$ by solving the following optimization problem:

$$\begin{aligned} \max_{\{f_j\}} \quad & O_{\leq\lambda}(f_{T,N}), \\ \text{s.t.} \quad & p_{succ}(f_{T,N}) \geq p_0, \end{aligned} \quad (30)$$

where the variables are $\{\hat{f}_j := \hat{f}((j+1/2)T/N) : j = -N, \dots, N-1\}$ (which are complex numbers), $f_{T,N}$ is given by Eq. (6), λ is an upper bound on the ground state energy of H , $O_{\leq\lambda}$ is given by Eq. (29), p_{succ} is given by Eq. (19), and p_0 is the minimum desired success probability. This is a high-dimensional continuous and constrained optimization problem, so in principle it can be solved by any well-established optimization algorithm on a classical computer. However, given the large number of complex variables and the non-convexity of the objective function and the constraint in these variables, this approach takes a long time and hence is impractical.

To circumvent the above issue and make our scheme practical, we confine \hat{f} to a class of parameterized functions $\hat{f}_{\vec{\theta}}$ (e.g. Gaussian functions), where $\vec{\theta} = (\theta_1, \dots, \theta_k)$ are the parameters, and find the optimal solution of Problem (30) within this class of functions. This way we only need to solve a low-dimensional optimization problem, which does not take much time on a classical computer.

Formally, let $f_{\vec{\theta}}$ be the inverse Fourier transform of $\hat{f}_{\vec{\theta}}$, i.e.

$$f_{\vec{\theta}}(x) = \int_{-\infty}^{\infty} \hat{f}_{\vec{\theta}}(\xi) e^{i2\pi x\xi} d\xi. \quad (31)$$

To make the booster operation $f_{\vec{\theta}}(H)$ perform well and easy to implement, we demand that $f_{\vec{\theta}}$ and $\hat{f}_{\vec{\theta}}$ satisfy the following requirements:

- $|\hat{f}_{\vec{\theta}}|$ needs to decay exponentially on \mathbb{R} , so that we can guarantee that $\|f_{\vec{\theta}}(H) - f_{\vec{\theta};T}(H)\| \leq \epsilon$ by choosing some $T = O(\log(1/\epsilon))$ as $\epsilon \rightarrow 0$, which implies that $f_{\vec{\theta}}(H)$ can be approximately implemented by a limited-depth circuit.
- $|f_{\vec{\theta}}|$ needs to decrease monotonically on $[0, 1]$, so that $f_{\vec{\theta}}(H)$ suppresses the high-energy eigenstates of H . In fact, we want $|f_{\vec{\theta}}|$ to decrease as fast as possible on $[0, 1]$, so that $f_{\vec{\theta}}(H)$ suppresses the high-energy eigenstates of H to a large extent.

Here we recommend two classes of functions that meet these conditions: Gaussian and hyperbolic secant. Both of them also have the nice property that the Fourier transform returns a function of the same form as the input function. That is,

- Gaussian function and its Fourier transform:

$$f_a(x) = e^{-ax^2} \leftrightarrow \hat{f}_a(\xi) = \sqrt{\frac{\pi}{a}} e^{-\frac{(\pi\xi)^2}{a}}; \quad (32)$$

- Hyperbolic secant (hsec) function and its Fourier transform:

$$f_a(x) = \operatorname{sech}(ax) \leftrightarrow \hat{f}_a(\xi) = \frac{\pi}{a} \operatorname{sech}\left(\frac{\pi^2}{a}\xi\right). \quad (33)$$

By contrast, the exponential function $f_a(x) = e^{-a|x|}$ can be an effective booster choice for suppressing the high-energy eigenstates, but its Fourier transform decays more slowly (i.e. quadratically), resulting in a higher-depth circuit for some fixed a .

We acknowledge that the optimal choice of the parameterized functions depends on the specific situation (including S and $q(x)$), and leave it as future work to develop a method for this task.

Once we determine the class of parameterized functions $f_{\vec{\theta}}$, we find the optimal parameters $\vec{\theta}$ by solving the following optimization problem:

$$\begin{aligned} \max_{\vec{\theta}} \quad & O_{\leq\lambda}(f_{\vec{\theta}}), \\ \text{s.t.} \quad & p_{\text{succ}}(f_{\vec{\theta}}) \geq p_0, \\ & \int_{-\infty}^{-T} |\hat{f}_{\vec{\theta}}(\xi)| d\xi + \int_T^{\infty} |\hat{f}_{\vec{\theta}}(\xi)| d\xi \leq \delta, \\ & |f_{\vec{\theta}}(0)| = 1. \end{aligned} \quad (34)$$

Here $p_{succ}(f_{\vec{\theta}})$ is defined as

$$p_{succ}(f_{\vec{\theta}}) := \frac{\langle \psi | f_{\vec{\theta}}^\dagger(H) f_{\vec{\theta}}(H) | \psi \rangle}{\left(\int_{-\infty}^{\infty} |\hat{f}_{\vec{\theta}}(\xi)| d\xi \right)^2}. \quad (35)$$

In Problem (34), the second constraint is simply "RHS of Eq. (9) $\leq \delta$ ", where $\delta > 0$ is determined by the method in Appendix C. The idea here is that when δ is small, Eq. (9) implies that $f_{\vec{\theta}} \approx f_{T;\vec{\theta}}$ and hence $O_{\leq \lambda}(f_{\vec{\theta}}) \approx O_{\leq \lambda}(f_{T;\vec{\theta}})$. In other words, this constraint guarantees that the objective function $O_{\leq \lambda}(f_{\vec{\theta}})$ accurately captures the quality of the real booster implemented by the LCU circuit. Appendix C gives a detailed relationship between δ and $|O_{\leq \lambda}(f_{\vec{\theta}}) - O_{\leq \lambda}(f_{T;\vec{\theta}})|$ from which an appropriate δ can be derived. From now on, we will refer to δ as the *implementation error*, as it is an upper bound on $\|f_{\vec{\theta}}(H) - f_{T;\vec{\theta}}(H)\|$. Meanwhile, the last constraint $|f_{\vec{\theta}}(0)| = 1$ is not strictly necessary, but we add it to normalize $f_{\vec{\theta}}$, as δ depends on $|f_{\vec{\theta}}(0)| = \max_{x \in [0,1]} |f_{\vec{\theta}}(x)|$. Then provided that N is sufficiently large, we have $f_{\vec{\theta}} \approx f_{T;\vec{\theta}} \approx f_{T,N;\vec{\theta}}$. It follows that $O_{\leq \lambda}(f_{\vec{\theta}}) \approx O_{\leq \lambda}(f_{T;\vec{\theta}}) \approx O_{\leq \lambda}(f_{T,N;\vec{\theta}})$ and $p_{succ}(f_{\vec{\theta}}) \approx p_{succ}(f_{T;\vec{\theta}}) \approx p_{succ}(f_{T,N;\vec{\theta}})$. In other words, the actions of $f(H)$, $f_T(H)$ or $f_{T,N}(H)$ on the input state $|\psi\rangle$ will be similar. Note that if $\hat{f}_{\vec{\theta}}$ is real and even, then $f_{\vec{\theta}}$ is also real and even. In this case, if $\hat{f}_{\vec{\theta}}$ is also non-negative, then $p_{succ}(f_{\vec{\theta}})$ can be approximated by the RHS of Eq. (21).

2.3 Example: Gaussian booster

Now we demonstrate our method on a simple example. Suppose $H = \sum_{j=1}^D \lambda_j |\lambda_j\rangle \langle \lambda_j|$ has eigenvalues λ_j 's and eigenstates $|\lambda_j\rangle$'s, where $0 \leq \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_D \leq 1$ and the λ_j 's spread almost uniformly between 0 and 1. Moreover, suppose $|\psi\rangle = \sum_{j=1}^D \mu_j |\lambda_j\rangle$ is the state produced by an ansatz circuit, where $\gamma_j := |\mu_j|^2$ decays almost exponentially in j . In this case, we set $S = [0, 1]$ and $q(x) = \beta e^{-\beta x} / (1 - e^{-\beta})$ for some $\beta > 0$ in our framework.

We confine the booster function f to Gaussian functions $f_a(x) = e^{-ax^2}$ and find the optimal parameter a as follows. By direct calculation, we get

$$\langle \psi | f_a^2(H) | \psi \rangle = \frac{\beta}{1 - e^{-\beta}} \int_0^1 e^{-2ax^2 - \beta x} dx \quad (36)$$

$$= \frac{\sqrt{\pi} \beta e^{\beta^2/(8a)}}{\sqrt{8a}(1 - e^{-\beta})} \left[\text{erf}(\sqrt{2a} + \beta/\sqrt{8a}) - \text{erf}(\beta/\sqrt{8a}) \right], \quad (37)$$

and

$$\langle \psi | f_a(H) P_{\leq \lambda} f_a(H) | \psi \rangle = \frac{\beta}{1 - e^{-\beta}} \int_0^\lambda e^{-2ax^2 - \beta x} dx \quad (38)$$

$$= \frac{\sqrt{\pi} \beta e^{\beta^2/(8a)}}{\sqrt{8a}(1 - e^{-\beta})} \left[\text{erf}(\lambda\sqrt{2a} + \beta/\sqrt{8a}) - \text{erf}(\beta/\sqrt{8a}) \right]. \quad (39)$$

In addition, we have

$$\int_{-\infty}^{-T} |\hat{f}_a(\xi)| d\xi + \int_T^{\infty} |\hat{f}_a(\xi)| d\xi = 1 - \operatorname{erf}(\pi T/\sqrt{a}). \quad (40)$$

Also, note that $f_a(0) = 1$. So we solve the following optimization problem:

$$\begin{aligned} \max_{a>0} \quad & \frac{\text{RHS of Eq. (39)}}{\text{RHS of Eq. (37)}}, \\ \text{s.t.} \quad & \text{RHS of Eq. (37)} \geq p_0, \\ & \text{RHS of Eq. (40)} \leq \delta, \end{aligned} \quad (41)$$

where $\lambda \in \mathbb{R}$ is the best known upper bound on λ_1 , and $\delta > 0$ is determined by the method in Appendix C.

This optimization problem can be solved efficiently on a classical computer as follows. Note that RHS of Eq. (37) is a monotonically decreasing function of a , and RHS of Eq. (40) is a monotonically increasing function of a . Let a_1 and a_2 be the solutions of RHS of Eq. (37) = p_0 and RHS of Eq. (40) = δ , respectively. Both a_1 and a_2 can be computed (within high accuracy) quickly on a classical computer via binary search. Then the set of feasible solutions to Problem 41 is $(0, \min(a_1, a_2)]$. Meanwhile, we claim that the objective function $\frac{\text{RHS of Eq. (39)}}{\text{RHS of Eq. (37)}}$ is a monotonically increasing function of a . This implies that the optimal solution to Problem 41 is $a = \min(a_1, a_2)$. To prove this claim, note that

$$\frac{\text{RHS of Eq. (39)}}{\text{RHS of Eq. (37)}} = \frac{\int_0^\lambda e^{-2ax^2 - \beta x} dx}{\int_0^1 e^{-2ax^2 - \beta x} dx} = \frac{g(a)}{g(a) + h(a)}, \quad (42)$$

where $g(a) := \int_0^\lambda e^{-2ax^2 - \beta x} dx$ and $h(a) := \int_\lambda^1 e^{-2ax^2 - \beta x} dx$. For arbitrary $a, \zeta > 0$, we have

$$\frac{g(a + \zeta)}{g(a)} = \frac{\int_0^\lambda e^{-2(a+\zeta)x^2 - \beta x} dx}{\int_0^\lambda e^{-2ax^2 - \beta x} dx} \geq e^{-2\zeta\lambda^2} \geq \frac{\int_\lambda^1 e^{-2(a+\zeta)x^2 - \beta x} dx}{\int_\lambda^1 e^{-2ax^2 - \beta x} dx} = \frac{h(a + \zeta)}{h(a)}, \quad (43)$$

which implies

$$\frac{g(a + \zeta)}{g(a + \zeta) + h(a + \zeta)} \geq \frac{g(a)}{g(a) + h(a)}, \quad (44)$$

as claimed.

Once we find the optimal parameter a , the corresponding booster operation $f_a(H)$ is implemented by the LCU method based on the equation:

$$f_a(x) = \sqrt{\frac{\pi}{a}} \int_{-\infty}^{\infty} e^{-\frac{(\pi\xi)^2}{a}} e^{2\pi i x \xi} d\xi \approx f_{T,N;a}(x) = \frac{T}{N} \sqrt{\frac{\pi}{a}} \sum_{j=-N}^{N-1} e^{-\frac{(\pi\xi_j)^2}{a}} e^{2\pi i x \xi_j}, \quad (45)$$

where $\xi_j = (j + 1/2)T/N$, and N is sufficiently large so that $f_a(x) \approx f_{T,a}(x) \approx f_{T,N;a}(x)$ for all $x \in [0, 1]$. Detailed analysis of Gaussian boosters is given in Appendix D.

We follow the methodology in Section 2.1 to construct the quantum circuit for implementing $f_{T,N;a}(H)$ probabilistically. For ease of implementation, we choose N to be a power of 2, i.e. $N = 2^n$ for an integer $n \geq 1$. The circuit for $f_{T,N;a}(H)$ is a special instance of the circuit in Figure 1, where V is an $(n + 1)$ -qubit unitary operation such that

$$V |0^{n+1}\rangle = \frac{1}{\sqrt{\sum_{k=0}^{2N-1} \alpha_k}} \sum_{k=0}^{2N-1} \sqrt{\alpha_k} |k\rangle \quad (46)$$

in which $\alpha_k := \frac{T}{N} \hat{f}_a(\xi_{k-N}) = \frac{T}{N} \sqrt{\frac{\pi}{a}} e^{-\frac{(\pi \xi_{k-N})^2}{a}} = \frac{T}{N} \sqrt{\frac{\pi}{a}} e^{-\frac{\pi^2 (k-N+1/2)^2 T^2}{aN^2}}$, and O is a unitary operation on the joint system such that

$$O = \sum_{k=0}^{2N-1} |k\rangle\langle k| \otimes U^{2k-2N+1} \quad (47)$$

in which $U := e^{i\pi HT/N}$. Here we have used the fact that α_k is real and positive for each k in defining V and O . Note that O is similar to the unitary operation in quantum phase estimation (QPE) and can be implemented as follows:

$$O = U^{-2N+1} O_n O_{n-1} \cdots O_0, \quad (48)$$

where

$$O_p := |0\rangle\langle 0| \otimes \mathbb{I} + |1\rangle\langle 1| \otimes U^{2^{p+1}} \quad (49)$$

acts on the p -th ancilla qubit (which corresponds to k_p assuming $k = \sum_{j=0}^n k_j 2^j$ for $k_0, k_1, \dots, k_n \in \{0, 1\}$) and the main system, for $p = 0, 1, \dots, n$. Figure 2 illustrates a circuit for implementing O in the case $n = 2$. Note that the final time evolution operation U^{-2N+1} applied to the main system can be dropped because it does not alter the ground state overlap of the output state.

2.4 Simplified optimization scheme for the Gaussian booster

In the asymptotic regime of large T , we expect that the optimized booster will produce an output state with weight concentrated mostly on the ground state and the small remainder on the first excited state. This allows us to forgo using a weight model of the input state (e.g. the exponential decay model described above) in the optimization.

Furthermore, we will remove the constraints on the success probability p_0 and implementation error δ in the optimization. The success probability is removed so as to focus on the quality of the prepared state instead of the time (in repetitions) needed to implement it. While the implementation error is ignored as a constraint, we will

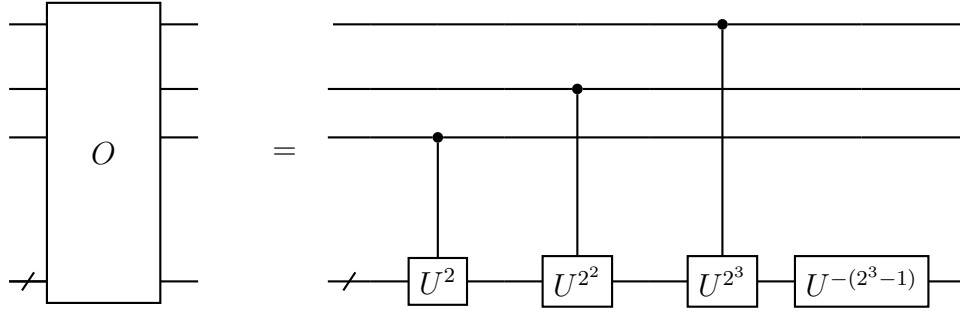


Figure 2: The quantum circuit for implementing the unitary operation O for Gaussian boosters in the case $n = 2$ has the same structure as the quantum phase estimation (QPE) circuit. The final time evolution U^{-7} on the system register can be removed because it does not alter the weights of the eigenstates.

modify the cost function so that the implementation error plays a direct role in the optimization.

We consider maximizing a lower bound on the ground state overlap that can be evaluated independently of the input state,

$$\begin{aligned}
O_{\leq E_0}(f_{T;a}) &= \frac{|\langle \psi | f_{T;a}(H) | \lambda_1 \rangle|^2}{\langle \psi | f_{T;a}^2(H) | \psi \rangle} \\
&= |\langle b_{T;a} | \lambda_1 \rangle|^2 \\
&\geq (\text{Re}(\langle b_{T;a} | \lambda_1 \rangle))^2 \\
&= \left(1 - \frac{1}{2} \|\lambda_1 - |b_{T;a}\rangle\|^2\right)^2 \\
&\geq \left(1 - \frac{1}{2} (\|\lambda_1 - |b_a\rangle\| + \||b_a\rangle - |b_{T;a}\rangle\|)^2\right)^2, \tag{50}
\end{aligned}$$

where we've defined $|b_{T;a}\rangle \equiv \frac{f_{T;a}(H)|\psi\rangle}{\|f_{T;a}(H)|\psi\rangle}$. This lower bound to the cost function is maximized by minimizing the sum of the two quantities in Eq. (50): the ideal ground state preparation error, $\|\lambda_1 - |b_a\rangle\|$, and the truncation error, $\||b_a\rangle - |b_{T;a}\rangle\|$. As described in Appendix D.1, each of these quantities can be upper bounded by expressions involving the Gaussian parameter a , the energy gap Δ , the initial ground state overlap γ , and the truncation level T . The upper bound for the ideal ground state preparation is:

$$\|\lambda_1 - |b_a\rangle\| \leq 2e^{-a\Delta^2}/\gamma, \tag{51}$$

and the upper bound for the truncation error is:

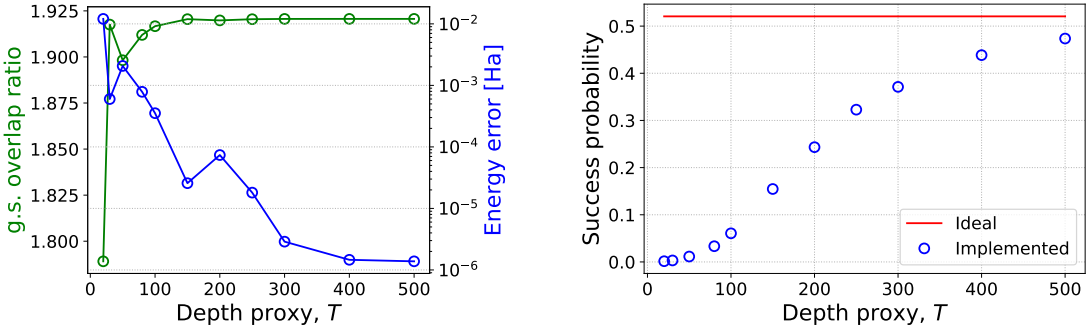
$$\||b_a\rangle - |b_{T;a}\rangle\| \leq 2 \left(1 - \text{erf}(\pi T/\sqrt{a})\right) / \gamma. \tag{52}$$

Putting this together, the simplified optimization strategy finds the a which minimizes the sum of these upper bounds on the state preparation error and the truncation error. Noticing that the initial overlap does not impact the optimization, the optimal a is obtained from:

$$\min_a \left(e^{-a\Delta^2} - \text{erf}(\pi T/\sqrt{a}) \right). \quad (53)$$

In the following section we assess the performance of the resulting boosters through numerical simulation on a specific small molecule example. For these numerics we use the simplified version of the optimization problem described above, which is suited to the asymptotic performance of the Gaussian booster as described in Appendix D.1.

3 Numerical Simulations



(a) Ground state overlap ratio (green) and error in energy (blue) as a function of truncation limit/depth proxy T .

(b) Resulting success probability of the LCU circuits. We show the expected success probability of the “ideal” booster (i.e. un-truncated Fourier expansion) in red. In practice, we truncate the Fourier series, and we show the success probability of such “implemented” boosters in blue.

Figure 3: Evaluating the performance of the Gaussian booster for preparing the ground state of the N_2 molecular system.

In this section, we provide numerical simulations of the Gaussian booster described in Section 2.3 using the optimization strategy described in Section 2.4. We consider the task of preparing the ground state of the N_2 molecular system, considering 6 active electrons in 12 spin-orbitals or qubits in an adapted basis [22] at a bond length of 2.0 Å. We prepare a mean-field state as the initial state $|\phi_0\rangle$, in which the initial ground state overlap, $\gamma \equiv |\langle \lambda_1 | \phi_0 \rangle|$, is 0.72. The ground state fidelity is thus $\gamma^2 = 0.52$. Alternatively, there are methods which ensure better overlap with the ground state using short circuit depth (e.g. SPA [23]). If we applied boosters to such initializations, compared to boosting the Hartree-Fock state, we would expect to be able to achieve better overlap.

To determine which Gaussian boosters to use in this setting, as described in Section 2.4 we choose a based on the optimization problem of Eq. (53). In our simulations, we set the spectral gap to 80% of its value (i.e. 0.8Δ) to reflect that in practice, the spectral gap is estimated with limited accuracy.³

Once an optimal a is determined, the performance of the booster can be characterized using several metrics: (1) the ground state overlap ratio, defined as $\frac{|\langle \lambda_1 | b_{T;a} \rangle|^2}{|\langle \lambda_1 | \psi \rangle|^2}$, where $|b_{T;a}\rangle$ is the normalized state after applying the Gaussian booster and $|\lambda_1\rangle$ is the ground state, (2) the error in energy, i.e. $\langle \phi | H | \phi \rangle - \lambda_1$, and (3) the success probability of implementing the booster. In Fig. 3, we plot the three quantities as a function of T , the truncation level in the truncated Fourier expansion of $f(H)$, or $f_T(H)$. We also consider T as a proxy for the circuit depth as it defines the maximum evolution time in the LCU circuit. The circuit implementing the Gaussian booster, as described in the end of Section 2.3, involves the same accumulated time evolution duration as the operation $c\text{-exp}(4\pi iHT)$. Therefore, measured in terms of number of accumulated $c\text{-exp}(2\pi iH)$ operations, the circuit depth is $2T$. For any fixed compilation scheme applied to this controlled time evolution of depth D , the circuit depth used in the state preparation booster will be $2TD$.

From our results, we observe that the booster converts circuit depth into overlap, in which a greater value of T corresponds to a greater ground state overlap ratio or lower error in estimating the ground state energy. We emphasize that this reliable depth-to-overlap conversion is not a feature of near-term ground state preparation algorithms such as VQE. Using the error function as our objective, the resulting optimal values of a are on the order of 10^6 for the values of T we consider. At such large values of a , for ideal boosters, the ground state overlap ratio saturates to $1/\gamma^2 \approx 1.93$ and the success probability saturates to the initial ground state probability $\gamma^2 = 0.72^2 \approx 0.52$. This latter value is shown in Fig. 3b as a red horizontal line. For the implemented boosters (i.e. with truncation T), we observe that for low values of T , the success probabilities are significantly lower than the ideal value. This is because at low values of T , the actual booster function applied deviates significantly from the ideal booster function of the form e^{-ax^2} as shown in Fig. 4. As T increases, the success probability approaches the ideal value. We also note that the success probability can be further increased by using fixed-point amplitude amplification [24] at the cost of increased circuit depth.

Next, we plot the relationship between T and the ground state infidelity $\tau \equiv 1 - |\langle \lambda_1 | \phi \rangle|^2$ in Fig. 5. We show in Appendix D.1 that it is sufficient to choose some $T = \tilde{O}(\log(1/\epsilon))$ where ϵ and τ are related by some constant factor. We compare the numerical results from this specific example with the analytical predictions by fitting the same functional form in Fig. 5 and observe a good fit. While it might be possible to achieve better performing boosters using a more refined optimization strategy, our results demonstrate the systematic increase in overlap or decrease in error as T increases.

We briefly provide a back-of-the-envelope calculation comparing the relative cost

³The spectral gap, the difference between the lowest and second-lowest energy, is 0.021 Ha before re-scaling.

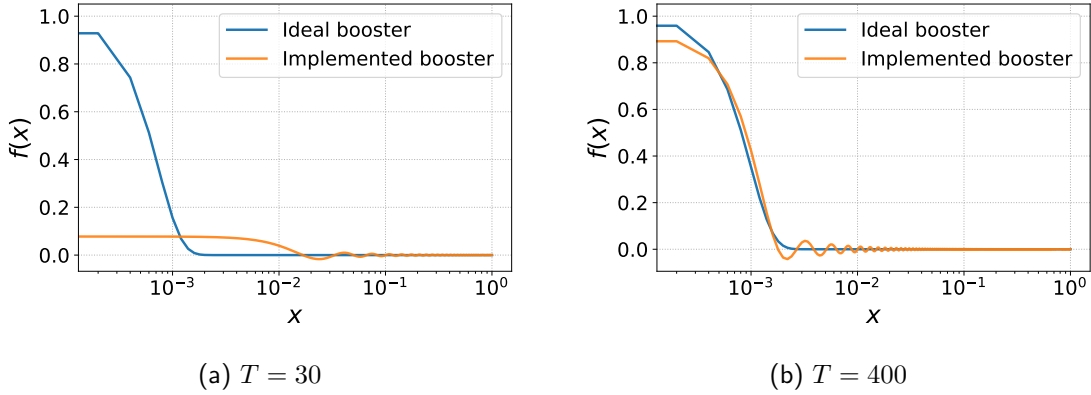


Figure 4: Comparing ideal versus implemented booster functions at two different values of T .

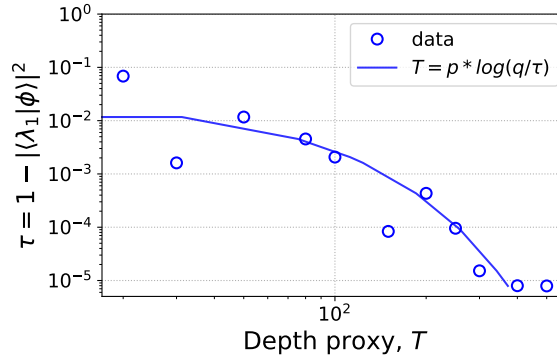


Figure 5: Ground state infidelity τ , as a function of the depth proxy T . We fit the function of the form $T = p \log(q/\tau)$, with $p \approx 46.72$ and $q \approx 0.023$, shown using a solid line.

of the Gaussian booster to that of the Quantum Phase Estimation algorithm (QPE). As explained in the end of Section 2.3, the circuit operation O used in the Gaussian booster has the same structure as the phase estimation circuit used in QPE. Therefore, we can compare the depths of these two algorithms in terms of the number of $c-e^{2\pi i H}$ applications they accumulate. Assuming that H has eigenspectrum between 0 and 1, QPE uses 2^{q+1} applications of $c-e^{2\pi i H}$ to obtain a q -bit precision estimate of the phase with high probability. To compare the circuit depths of the state preparation booster with QPE, we consider the same N_2 molecule analyzed above and calculate the number of $c-e^{2\pi i H}$ operations needed by QPE to get a chemically-accurate estimate of an eigenenergy. Chemical accuracy is roughly 10^{-3} mHa and the spectrum of the 12-qubit N_2 molecule ranges over roughly 10 mHa. Therefore, after shifting and rescaling the Hamiltonian such that its spectrum lies within 0 to 1, an energy estimate to within chemical accuracy requires $q = -\log_2(10^{-3}/10)$, or more than 13 bits of precision. The number of $c-e^{2\pi i H}$ needed is therefore roughly 20,000. The Gaussian boosters considered in Figure 3 use up to $T = 500$, which, as described earlier, is roughly the circuit depth of $2T = 1000$ sequential applications of $c-e^{2\pi i H}$.

Further investigation is necessary for a more precise comparison of the two methods. For example, the number of ancilla qubits used in the Gaussian booster is greater than that used in QPE (c.f. Appendix D). However, our rough calculation suggests the circuit depth, required for implementing a booster is expected to be a small fraction of the depth needed for QPE. This gives evidence that state preparation boosters are suited for improving quantum algorithms in the era of early-fault tolerant quantum computers.

4 Discussion and Outlook

In this work, we present a framework for improving state preparation with limited quantum circuit depth. The method of *state preparation boosters* reliably converts circuit depth into ground state overlap. This can be used to boost the performance of the variational quantum eigensolver algorithm on early fault-tolerant devices as well as the more far-term method of quantum phase estimation. We leave for future work an analysis of how state preparation boosters can be used to improve the performance of ground state energy estimation. We give a detailed treatment of the specific case of Gaussian boosters, motivated by the performance they achieve while keeping circuit depth relatively low. We carry out simulations of the method on a small (12 qubit) example of the molecule N_2 . These show that the truncated Gaussian boosters can reliably increase the ground state overlap as circuit depth is increased and they also exhibit the asymptotic scaling derived in Appendix D.1. Below we outline several important future directions to explore in order to better understand and assess the capabilities of boosters.

First, we have focused on the specific case of Gaussian boosters using controlled-time-evolution circuit primitives (i.e. $c - \exp(iHt)$). It is worth investigating the performance of alternative booster functions and alternative circuit primitives (e.g. block-encoded Hamiltonians $U = \begin{bmatrix} H & \\ & \cdot \end{bmatrix}$). Also, it might be possible to refine the booster design process, informed by further numerical investigations.

Along similar lines, it may be worth investigating alternative performance metrics for state preparation. While boosters reliably convert circuit depth into ground state overlap, particular classes of booster functions may yield better conversion rates than others. That is, it may be informative to derive a notion of “efficiency” of boosters, e.g. amount of overlap gained per unit depth of a booster. Such an efficiency metric may provide a way to assess and compare general state preparation methods such as VQE or QPE.

To properly assess the performance of state preparation boosters in practice, we should incorporate the impact of circuit error. One potential way to account for noise is to consider the circuit depolarizing model as was considered in Ref. [25]. Using this model, the output state is a mixture of the noiseless output state and the completely mixed state. So far we have fixed the circuit depth (or truncation level T) of the boosters. But, by incorporating a model of noise, we might arrive at an optimal choice

of booster circuit depth where ideal overlap is balanced with the decay of the quantum coherence.

In addition to circuit depth, another metric associated with boosters is the success probability. In the case of ideal boosters (not the truncated boosters analyzed in Fig. 3), the more that a booster suppresses high-energy eigenstates, the lower the success probability will be. In our numerical simulations, we had not designed boosters to accommodate success probability. But, in practice the success probability will lead to a time overhead of $O(1/p_{\text{success}})$ when the booster is used in an algorithm. One solution, as considered in Section 2.2, is to design boosters which yield a success probability above a fixed threshold. In some cases, this approach may not yield boosters with sufficiently high success probabilities. In such cases, an alternative strategy is to boost the success probability using amplitude amplification [24]. Such techniques cost additional circuit depth and we leave it to future work to analyze the cost-benefit trade-off in the setting of state preparation boosters. In general, there appears to be a trade-off between the booster depth and success probability. Depending on the constraint on the circuit depth, it may be preferable to apply multiple rounds of shallower boosters in sequence rather than applying a deep booster, with a higher overall success probability, once. Further studies are necessary to better understand the trade-off and identify when each strategy is preferable.

Our proposed booster method can be extended to prepare excited states. Using existing near-term methods for estimating low-lying excited states [26, 27, 28, 29, 30], one can obtain an initial state that approximates an excited state. Then, with the estimated excited state energy, call it \tilde{x} , one can construct a shifted Gaussian booster $f(x) = e^{-a(x-\tilde{x})^2}$ such that the booster function peaks (approximately) at the excited state energy. We note that many of the steps to implement the booster method for ground state preparation carry over for excited state preparation, including energy re-scaling.

In many applications, we need to estimate the expectation value of an observable O with respect to the ground state of a Hamiltonian H . Our method can be used to improve the accuracy of the estimate of this quantity. Alternatively, one can use the method in Ref. [31] to estimate this quantity without explicitly preparing the ground state. It would be interesting to know which approach is more efficient in practice.

Boosters provide a reliable depth to overlap conversion, which is lacking in near-term state preparation approaches. In particular, Gaussian boosters suppress high energy eigenstates while maintaining a relatively low depth, compared to other booster functions. Based on our theoretical and empirical analysis, boosters have the potential to become a useful and efficient subroutine for various quantum algorithms, aiding in the quest for quantum advantage.

Acknowledgement

We thank Jakob Kottmann, Michał Stęchły, Athena Caesura, Alex Kunitsa, and Daniel Stilck-França for providing valuable feedback to the manuscript. We thank all our colleagues at Zapata Computing for their feedback and support.

A List of symbols

f	Booster function
λ	An upper bound on ground state energy of H
T	Truncation limit; depth proxy
N	Number of points in discretization in approximation of f
Δ	Spectral gap
a	Width parameter of the Gaussian booster function
δ	Target booster implementation error
p_0	Target success probability
β	Parameter for modeling initial state weights using a beta distribution
γ	Initial overlap with the ground state, $ \langle \lambda_1 \psi \rangle $
ϵ	Lower bound on the error between the ground state and the implemented boosted state

B Bounding the distance between f_T and $f_{T,N}$

It is clear from Eqs. (5) and (6) that $f_{T,N} \rightarrow f_T$ as $N \rightarrow \infty$. A concrete bound on the convergence speed can be obtained as follows. Assume that \hat{f} is differentiable on \mathbb{R} . Let $\xi_j = (j + 1/2)T/N$ for $j = -N, -N + 1, \dots, N - 1$. Then for arbitrary $\xi \in [jT/N, (j + 1)T/N]$, the mean value theorem implies that

$$\hat{f}(\xi) = \hat{f}(\xi_j) + \hat{f}'(\eta)(\xi - \xi_j) \quad (54)$$

for some η between ξ and ξ_j . Consequently, we have

$$|\hat{f}(\xi) - \hat{f}(\xi_j)| \leq |\xi - \xi_j| \max_{\xi \in [-T, T]} |\hat{f}'(\xi)|, \quad (55)$$

and hence for arbitrary $x \in [0, 1]$,

$$\left| \hat{f}(\xi) e^{i2\pi x \xi} - \hat{f}(\xi_j) e^{i2\pi x \xi_j} \right| \leq \left| \hat{f}(\xi) e^{i2\pi x \xi} - \hat{f}(\xi_j) e^{i2\pi x \xi} \right| + \left| \hat{f}(\xi_j) e^{i2\pi x \xi} - \hat{f}(\xi_j) e^{i2\pi x \xi_j} \right| \quad (56)$$

$$= \left| \hat{f}(\xi) - \hat{f}(\xi_j) \right| + \left| \hat{f}(\xi_j) \right| \left| e^{i2\pi x \xi} - e^{i2\pi x \xi_j} \right| \quad (57)$$

$$\leq |\xi - \xi_j| \left(\max_{\xi \in [-T, T]} \left| \hat{f}'(\xi) \right| + 2\pi |x| \max_{\xi \in [-T, T]} \left| \hat{f}(\xi) \right| \right) \quad (58)$$

$$\leq \frac{T}{N} \left(\max_{\xi \in [-T, T]} \left| \hat{f}'(\xi) \right| + 2\pi \max_{\xi \in [-T, T]} \left| \hat{f}(\xi) \right| \right), \quad (59)$$

where in the third step we use $\left| e^{i2\pi x \xi} - e^{i2\pi x \xi_j} \right| \leq 2\pi |x| |\xi - \xi_j|$, and in the last step we use $|\xi - \xi_j| \leq T/N$ and $|x| \leq 1$. It follows that

$$|f_T(x) - f_{T,N}(x)| = \left| \int_{-T}^T \hat{f}(\xi) e^{i2\pi x \xi} d\xi - \frac{T}{N} \sum_{j=-N}^{N-1} \hat{f}(\xi_j) e^{i2\pi x \xi_j} \right| \quad (60)$$

$$= \left| \sum_{j=-N}^{N-1} \int_{jT/N}^{(j+1)T/N} \hat{f}(\xi) e^{i2\pi x \xi} d\xi - \frac{T}{N} \sum_{j=-N}^{N-1} \hat{f}(\xi_j) e^{i2\pi x \xi_j} \right| \quad (61)$$

$$\leq \sum_{j=-N}^{N-1} \left| \int_{jT/N}^{(j+1)T/N} \hat{f}(\xi) e^{i2\pi x \xi} d\xi - \frac{T}{N} \hat{f}(\xi_j) e^{i2\pi x \xi_j} \right| \quad (62)$$

$$\leq \sum_{j=-N}^{N-1} \int_{jT/N}^{(j+1)T/N} \left| \hat{f}(\xi) e^{i2\pi x \xi} - \hat{f}(\xi_j) e^{i2\pi x \xi_j} \right| d\xi \quad (63)$$

$$\leq \sum_{j=-N}^{N-1} \int_{jT/N}^{(j+1)T/N} \left| \hat{f}(\xi) e^{i2\pi x \xi} - \hat{f}(\xi_j) e^{i2\pi x \xi} \right| d\xi \quad (64)$$

$$+ \sum_{j=-N}^{N-1} \int_{jT/N}^{(j+1)T/N} \left| \hat{f}(\xi_j) e^{i2\pi x \xi} - \hat{f}(\xi_j) e^{i2\pi x \xi_j} \right| d\xi \quad (65)$$

$$= \sum_{j=-N}^{N-1} \int_{jT/N}^{(j+1)T/N} \left| \hat{f}(\xi) - \hat{f}(\xi_j) \right| d\xi \quad (66)$$

$$+ \sum_{j=-N}^{N-1} \int_{jT/N}^{(j+1)T/N} \left| \hat{f}(\xi_j) \right| \left| e^{i2\pi x \xi} - e^{i2\pi x \xi_j} \right| d\xi \quad (67)$$

$$\leq \frac{2T^2}{N} \left(\max_{\xi \in [-T, T]} \left| \hat{f}'(\xi) \right| + 2\pi \max_{\xi \in [-T, T]} \left| \hat{f}(\xi) \right| \right), \quad (68)$$

where the third, fourth and fifth steps follow from the triangle inequality, and the seventh step follows from

$$\left| \hat{f}(\xi) - \hat{f}(\xi_j) \right| \leq |\xi - \xi_j| \cdot \max_{\xi \in [-T, T]} \left| \hat{f}'(\xi) \right| \leq \frac{T}{N} \cdot \max_{\xi \in [-T, T]} \left| \hat{f}'(\xi) \right| \quad (69)$$

and

$$\left| e^{i2\pi x\xi} - e^{i2\pi x\xi_j} \right| \leq 2\pi |x| |\xi - \xi_j| \leq \frac{2\pi T}{N}. \quad (70)$$

Now let $R := \max_{\xi \in [-T, T]} (|\hat{f}(\xi)| + |\hat{f}'(\xi)|)$. Then we have

$$\max_{x \in [0, 1]} |f_T(x) - f_{T, N}(x)| = O\left(\frac{RT^2}{N}\right). \quad (71)$$

So for arbitrary $\epsilon > 0$, by choosing some $N = O(RT^2/\epsilon)$, we guarantee that $|f_T(x) - f_{T, N}(x)| \leq \epsilon$ for all $x \in [0, 1]$.

Recall that we require that $|\hat{f}|$ decays exponentially on \mathbb{R} , so that $\max_{x \in \mathbb{R}} |f(x) - f_T(x)| \leq \epsilon$ for some $T = O(\log(1/\epsilon))$. In this case, \hat{f} is bounded on \mathbb{R} , i.e., there exists a constant $R_1 > 0$ such that $|\hat{f}(\xi)| \leq R_1$ for all $\xi \in \mathbb{R}$. If we require that \hat{f}' is also bounded on \mathbb{R} , i.e., there exists a constant $R_2 > 0$ such that $|\hat{f}'(\xi)| \leq R_2$ for all $\xi \in \mathbb{R}$, then $R \leq R_1 + R_2 = O(1)$. In this case, it is sufficient to choose some $N = O(T^2/\epsilon) = \tilde{O}(1/\epsilon)$ to ensure that $\max_{x \in [0, 1]} |f_T(x) - f_{T, N}(x)| \leq \epsilon$, as claimed. Note that many exponentially-decaying functions (e.g. $\hat{f}(\xi) = e^{-a|\xi|^b}$ for some $a > 0$ and $b \geq 1$) have bounded derivatives and hence satisfy this additional requirement.

C Choosing an appropriate δ in Problem (34)

Here we describe how to set the parameter δ in Problem (34). For better readability, we drop the subscript $\vec{\theta}$ throughout this section. We aim to choose δ properly so that f_T is close to f , which implies that $O_{\leq \lambda}(f_T)$ is close to $O_{\leq \lambda}(f)$, i.e., the total overlap between the real boosted state and the low-energy eigenstates of H is close to the one for the ideal boosted state.

By Eq. (9), we know that the second constraint of Problem (34) ensures that

$$\|f(H) - f_T(H)\| \leq \delta, \quad (72)$$

which implies that

$$\|f(H)|\psi\rangle - f_T(H)|\psi\rangle\| \leq \delta. \quad (73)$$

Note that $\|f(H)|\psi\rangle\| \leq 1$ since $|f(x)| \leq 1$, for all $x \in [0, 1]$. So we have $\|f_T(H)|\psi\rangle\| \leq 1 + \delta$.

Lemma 1. *Suppose $|\psi_1\rangle$ and $|\psi_2\rangle$ are two vectors such that $\| |\psi_1\rangle \|, \| |\psi_2\rangle \| \leq a$ and $\| |\psi_1\rangle - |\psi_2\rangle \| \leq b$ for some $a, b > 0$, and A is a linear operator such that $\|A\| \leq 1$. Then*

$$|\langle \psi_1 | A | \psi_1 \rangle - \langle \psi_2 | A | \psi_2 \rangle| \leq 2ab. \quad (74)$$

Proof. Using the triangle inequality, we obtain

$$|\langle \psi_1 | A | \psi_1 \rangle - \langle \psi_2 | A | \psi_2 \rangle| \leq |\langle \psi_1 | A | \psi_1 \rangle - \langle \psi_1 | A | \psi_2 \rangle| + |\langle \psi_1 | A | \psi_2 \rangle - \langle \psi_2 | A | \psi_2 \rangle| \quad (75)$$

$$= |\langle \psi_1 | A (|\psi_1\rangle - |\psi_2\rangle)| + |(\langle \psi_1 | - \langle \psi_2 |) A | \psi_2 \rangle| \quad (76)$$

$$\leq ab + ba \quad (77)$$

$$= 2ab. \quad (78)$$

□

It follows from Lemma 1, Eq. (73), $f(H) |\psi\rangle \leq 1$ and $\|f_T(H) |\psi\rangle\| \leq 1 + \delta$ that

$$\left| \langle \psi | f^\dagger(H) P_{\leq \lambda} f(H) | \psi \rangle - \langle \psi | f_T^\dagger(H) P_{\leq \lambda} f_T(H) | \psi \rangle \right| \leq 2\delta(1 + \delta), \quad (79)$$

and

$$\left| \langle \psi | f^\dagger(H) f(H) | \psi \rangle - \langle \psi | f_T^\dagger(H) f_T(H) | \psi \rangle \right| \leq 2\delta(1 + \delta). \quad (80)$$

Meanwhile, the first constraint of Problem (34) ensures that

$$\langle \psi | f^\dagger(H) f(H) | \psi \rangle = p_{succ}(f) \left(\int_{-\infty}^{\infty} |\hat{f}(\xi)| d\xi \right)^2 \geq p_0 \left(\int_{-\infty}^{\infty} |\hat{f}(\xi)| d\xi \right)^2. \quad (81)$$

Let η be a lower bound on $\int_{-\infty}^{\infty} |\hat{f}(\xi)| d\xi$. It can be obtained by, e.g., noting that

$$\int_{-\infty}^{\infty} |\hat{f}(\xi)| d\xi \geq \left| \int_{-\infty}^{\infty} \hat{f}(\xi) d\xi \right| = |f(0)| = 1, \quad (82)$$

where we use the last constraint of Problem (34) in the last step. But one might obtain better bounds by utilizing other information about f . Then combining Eqs. (79), (80)

and (81) yields

$$|O_{\leq\lambda}(f) - O_{\leq\lambda}(f_T)| = \left| \frac{\langle \psi | f^\dagger(H) P_{\leq\lambda} f(H) | \psi \rangle}{\langle \psi | f^\dagger(H) f(H) | \psi \rangle} - \frac{\langle \psi | f_T^\dagger(H) P_{\leq\lambda} f_T(H) | \psi \rangle}{\langle \psi | f_T^\dagger(H) f_T(H) | \psi \rangle} \right| \quad (83)$$

$$\leq \left| \frac{\langle \psi | f^\dagger(H) P_{\leq\lambda} f(H) | \psi \rangle}{\langle \psi | f^\dagger(H) f(H) | \psi \rangle} - \frac{\langle \psi | f_T^\dagger(H) P_{\leq\lambda} f_T(H) | \psi \rangle}{\langle \psi | f^\dagger(H) f(H) | \psi \rangle} \right| \quad (84)$$

$$+ \left| \frac{\langle \psi | f_T^\dagger(H) P_{\leq\lambda} f_T(H) | \psi \rangle}{\langle \psi | f^\dagger(H) f(H) | \psi \rangle} - \frac{\langle \psi | f_T^\dagger(H) P_{\leq\lambda} f_T(H) | \psi \rangle}{\langle \psi | f_T^\dagger(H) f_T(H) | \psi \rangle} \right| \quad (85)$$

$$= \frac{|\langle \psi | f^\dagger(H) P_{\leq\lambda} f(H) | \psi \rangle - \langle \psi | f_T^\dagger(H) P_{\leq\lambda} f_T(H) | \psi \rangle|}{\langle \psi | f^\dagger(H) f(H) | \psi \rangle} \quad (86)$$

$$+ \frac{\langle \psi | f_T^\dagger(H) P_{\leq\lambda} f_T(H) | \psi \rangle}{\langle \psi | f_T^\dagger(H) f_T(H) | \psi \rangle} \cdot \frac{|\langle \psi | f^\dagger(H) f(H) | \psi \rangle - \langle \psi | f_T^\dagger(H) f_T(H) | \psi \rangle|}{\langle \psi | f^\dagger(H) f(H) | \psi \rangle} \quad (87)$$

$$\leq \frac{|\langle \psi | f^\dagger(H) P_{\leq\lambda} f(H) | \psi \rangle - \langle \psi | f_T^\dagger(H) P_{\leq\lambda} f_T(H) | \psi \rangle|}{\langle \psi | f^\dagger(H) f(H) | \psi \rangle} \quad (88)$$

$$+ \frac{|\langle \psi | f^\dagger(H) f(H) | \psi \rangle - \langle \psi | f_T^\dagger(H) f_T(H) | \psi \rangle|}{\langle \psi | f^\dagger(H) f(H) | \psi \rangle} \quad (89)$$

$$\leq \frac{2\delta(1+\delta)}{p_0\eta^2} + \frac{2\delta(1+\delta)}{p_0\eta^2} \quad (90)$$

$$= \frac{4\delta(1+\delta)}{p_0\eta^2}, \quad (91)$$

where the second step follows from the triangle inequality, the fourth step follows from $\langle \psi | f_T^\dagger(H) P_{\leq\lambda} f_T(H) | \psi \rangle \leq \langle \psi | f_T^\dagger(H) f_T(H) | \psi \rangle$, and the fifth step follows from Eqs. (79), (80) and (81) and the definition of η . Consequently, setting $\delta = (\sqrt{\epsilon p_0 \eta^2 + 1} - 1)/2$ ensures that $|O_{\leq\lambda}(f) - O_{\leq\lambda}(f_T)| \leq \epsilon$. This provides a method for picking an appropriate δ in Problem (34).

D Detailed analysis of Gaussian boosters

Here we give detailed analysis of Gaussian boosters. Recall that in Section 2.3, we demonstrate how to find the optimal parameter a for the booster function $f_a(x) = e^{-ax^2}$ in the case where $S = [0, 1]$ and $q(x) = \beta e^{-\beta x}/(1 - e^{-\beta})$ for some $\beta > 0$.

By Eqs. (9) and (40), we have

$$\max_{x \in \mathbb{R}} |f_a(x) - f_{T,a}(x)| \leq 1 - \operatorname{erf}(\pi T/\sqrt{a}) \leq \frac{\sqrt{a} e^{-\pi^2 T^2/a}}{\pi^{3/2} T}, \quad (92)$$

where in the second step we use the fact that

$$1 - \operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} dt \leq \frac{1}{\sqrt{\pi x}} \int_x^\infty 2te^{-t^2} dt = \frac{e^{-x^2}}{\sqrt{\pi x}}. \quad (93)$$

This implies that for some $a = \tilde{\Theta}(T^2/\log(1/\delta))$, we have

$$\|f_a(H) - f_{T,a}(H)\| \leq \max_{x \in \mathbb{R}} |f_a(x) - f_{T,a}(x)| \leq \delta. \quad (94)$$

In this case, we use the results in Appendix B to estimate the smallest N such that

$$\|f_{T,a}(H) - f_{T,N,a}(H)\| \leq \max_{x \in [0,1]} |f_{T,a}(x) - f_{T,N,a}(x)| \leq \delta. \quad (95)$$

Recall that $\hat{f}_a(\xi) = \sqrt{\frac{\pi}{a}} e^{-\frac{(\pi\xi)^2}{a}}$ is the Fourier transform of $f_a(x) = e^{-ax^2}$. Then we have

$$R := \max_{\xi \in [-T,T]} (|\hat{f}_a(\xi)| + |\hat{f}'_a(\xi)|) \quad (96)$$

$$= \max_{\xi \in [-T,T]} \left(\sqrt{\frac{\pi}{a}} e^{-\frac{(\pi\xi)^2}{a}} + \frac{2\pi^{3/2}}{a^{3/2}} e^{-\frac{(\pi\xi)^2}{a}} |\xi| \right) \quad (97)$$

$$\leq \max_{\xi \in [-T,T]} \left(\sqrt{\frac{\pi}{a}} e^{-\frac{(\pi\xi)^2}{a}} \right) + \max_{\xi \in [-T,T]} \frac{2\pi^{3/2}}{a^{3/2}} e^{-\frac{(\pi\xi)^2}{a}} |\xi| \quad (98)$$

$$= \sqrt{\frac{\pi}{a}} + \frac{\sqrt{2\pi}}{\sqrt{ea}} \quad (99)$$

$$= O\left(\frac{1}{\sqrt{a}}\right) \quad (100)$$

$$= \tilde{O}\left(\frac{\sqrt{\log(1/\delta)}}{T}\right). \quad (101)$$

It follows that Eq. (95) holds for some $N = O(RT^2/\delta) = \tilde{O}(T/\delta)$.

D.1 Asymptotic performance of Gaussian boosters

Now suppose we want the boosted state $\frac{f_{T,a}(H)|\psi\rangle}{\|f_{T,a}(H)|\psi\rangle\|}$ to be ϵ -close to the ground state of H , for arbitrary $\epsilon > 0$. How large does T need to be as a function of ϵ ? We will show that $T = \tilde{O}(\log(1/\epsilon))$ is sufficient for this purpose.

Formally, suppose $H = \sum_{j=1}^D \lambda_j |\lambda_j\rangle \langle \lambda_j|$ has eigenvalues λ_j 's and eigenstates $|\lambda_j\rangle$'s, where $0 = \lambda_1 < \lambda_2 = \Delta \leq \dots \leq \lambda_D \leq 1$. Moreover, suppose $|\psi\rangle = \sum_{j=1}^D \mu_j |\lambda_j\rangle$ is the state produced by an ansatz circuit such that $\gamma := \mu_1 > 0$. Then applying $f_a(H)$ or $f_{T,a}(H)$ on $|\psi\rangle$ yields the unnormalized state $f_a(H)|\psi\rangle = \sum_{j=1}^D f_a(x)\mu_j |\lambda_j\rangle$ or $f_{T,a}(H)|\psi\rangle = \sum_{j=1}^D f_{T,a}(x)\mu_j |\lambda_j\rangle$, respectively. We want to choose appropriate T and a such that

$$\left\| |\lambda_1\rangle - \frac{f_{T,a}(H)|\psi\rangle}{\|f_{T,a}(H)|\psi\rangle\|} \right\| \leq \epsilon \quad (102)$$

for given $\epsilon > 0$. By the triangle inequality, we know that Eq. (102) holds if we have both

$$\left\| |\lambda_1\rangle - \frac{f_a(H) |\psi\rangle}{\|f_a(H) |\psi\rangle\|} \right\| \leq \epsilon/2 \quad (103)$$

and

$$\left\| \frac{f_a(H) |\psi\rangle}{\|f_a(H) |\psi\rangle\|} - \frac{f_{T;a}(H) |\psi\rangle}{\|f_{T;a}(H) |\psi\rangle\|} \right\| \leq \epsilon/2. \quad (104)$$

It remains to find T and a that satisfy these equations. First, note that

$$\|\gamma |\lambda_1\rangle - f_a(H) |\psi\rangle\| = \left\| \sum_{j=2}^D \mu_j e^{-a\lambda_j^2} |\lambda_j\rangle \right\| = \sqrt{\sum_{j=2}^D |\mu_j|^2 e^{-2a\lambda_j^2}} \leq e^{-a\Delta^2}, \quad (105)$$

where the last step follows from $\lambda_j \geq \Delta$ for $j = 2, 3, \dots, D$ and $\sum_{j=2}^D |\mu_j|^2 \leq 1$. This implies that

$$|\gamma - \|f_a(H) |\psi\rangle\|| = \|\gamma |\lambda_1\rangle\| - \|f_a(H) |\psi\rangle\| \leq e^{-a\Delta^2}. \quad (106)$$

As a result, we have

$$\left\| |\lambda_1\rangle - \frac{f_a(H) |\psi\rangle}{\|f_a(H) |\psi\rangle\|} \right\| \leq \left\| |\lambda_1\rangle - \frac{f_a(H) |\psi\rangle}{\gamma} \right\| \quad (107)$$

$$+ \left\| \frac{f_a(H) |\psi\rangle}{\gamma} - \frac{f_a(H) |\psi\rangle}{\|f_a(H) |\psi\rangle\|} \right\| \quad (108)$$

$$= \frac{\|\gamma |\lambda_1\rangle - f_a(H) |\psi\rangle\|}{\gamma} \quad (109)$$

$$+ \frac{|\gamma - \|f_a(H) |\psi\rangle\||}{\gamma} \quad (110)$$

$$\leq \frac{e^{-a\Delta^2}}{\gamma} + \frac{e^{-a\Delta^2}}{\gamma} \quad (111)$$

$$= \frac{2e^{-a\Delta^2}}{\gamma}, \quad (112)$$

where the first step follows from the triangle inequality, and the third step follows from Eqs. (105) and (106). Thus, Eq. (103) holds for some $a = O(\log(1/(\gamma\epsilon))/\Delta^2)$. Meanwhile, Eq. (92) implies that

$$\|f_a(H) |\psi\rangle - f_{T;a}(H) |\psi\rangle\| \leq 1 - \operatorname{erf}(\pi T/\sqrt{a}) \leq \frac{\sqrt{a}e^{-\pi^2 T^2/a}}{\pi^{3/2}T}, \quad (113)$$

which in turn implies that

$$\left| \|f_a(H) |\psi\rangle\| - \|f_{T;a}(H) |\psi\rangle\| \right| \leq 1 - \operatorname{erf}(\pi T/\sqrt{a}) \leq \frac{\sqrt{a}e^{-\pi^2 T^2/a}}{\pi^{3/2}T}. \quad (114)$$

Then we get

$$\left\| \frac{f_a(H) |\psi\rangle}{\|f_a(H) |\psi\rangle\|} - \frac{f_{T;a}(H) |\psi\rangle}{\|f_{T;a}(H) |\psi\rangle\|} \right\| \leq \left\| \frac{f_a(H) |\psi\rangle}{\|f_a(H) |\psi\rangle\|} - \frac{f_{T;a}(H) |\psi\rangle}{\|f_a(H) |\psi\rangle\|} \right\| \quad (115)$$

$$+ \left\| \frac{f_{T;a}(H) |\psi\rangle}{\|f_a(H) |\psi\rangle\|} - \frac{f_{T;a}(H) |\psi\rangle}{\|f_{T;a}(H) |\psi\rangle\|} \right\| \quad (116)$$

$$= \frac{\|f_a(H) |\psi\rangle - f_{T;a}(H) |\psi\rangle\|}{\|f_a(H) |\psi\rangle\|} \quad (117)$$

$$+ \frac{\left| \|f_a(H) |\psi\rangle\| - \|f_{T;a}(H) |\psi\rangle\| \right|}{\|f_a(H) |\psi\rangle\|} \quad (118)$$

$$\leq \frac{1 - \operatorname{erf}(\pi T/\sqrt{a})}{\gamma} + \frac{1 - \operatorname{erf}(\pi T/\sqrt{a})}{\gamma} \quad (119)$$

$$= \frac{2(1 - \operatorname{erf}(\pi T/\sqrt{a}))}{\gamma} \quad (120)$$

$$\leq \frac{2\sqrt{a}e^{-\pi^2 T^2/a}}{\pi^{3/2}T\gamma}, \quad (121)$$

where the first step follows from the triangle inequality, and the third step follows from Eqs. (113) and (114) and $\|f_a(H) |\psi\rangle\| = \sqrt{\gamma^2 + \sum_{j=2}^D |\mu_j|^2 e^{-2a\lambda_j^2}} \geq \gamma$. So Eq. (104) holds for some $T = \tilde{O}(\sqrt{a \log(1/(\gamma\epsilon))}) = \tilde{O}(\log(1/(\gamma\epsilon))/\Delta)$. Overall, by choosing some $T = \tilde{O}(\log(1/(\gamma\epsilon))/\Delta)$ and $a = O(\log(1/(\gamma\epsilon))/\Delta^2)$, we ensure that Eq. (102) holds, as claimed.

References

- [1] Alán Aspuru-Guzik, Anthony D Dutoi, Peter J Love, and Martin Head-Gordon. Simulated quantum computation of molecular energies. *Science*, 309(5741):1704–1707, 2005.
- [2] Yudong Cao, Jonathan Romero, Jonathan P Olson, Matthias Degroote, Peter D Johnson, Mária Kieferová, Ian D Kivlichan, Tim Menke, Borja Peropadre, Nicolas PD Sawaya, et al. Quantum chemistry in the age of quantum computing. *Chemical Reviews*, 119(19):10856–10915, 2019.
- [3] Ivan Kassal, Stephen P Jordan, Peter J Love, Masoud Mohseni, and Alán Aspuru-Guzik. Polynomial-time quantum algorithm for the simulation of chemical dynamics. *Proceedings of the National Academy of Sciences*, 105(48):18681–18686, 2008.

- [4] Alexei Yu Kitaev, Alexander Shen, Mikhail N Vyalyi, and Mikhail N Vyalyi. *Classical and quantum computation*. Number 47. American Mathematical Society, 2002.
- [5] 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.
- [6] Lin Lin and Yu Tong. Heisenberg-limited ground state energy estimation for early fault-tolerant quantum computers. *PRX Quantum* 3(1):010318, 2022.
- [7] Emanuel Knill, Gerardo Ortiz, and Rolando D Somma. Optimal quantum measurements of expectation values of observables. *Physical Review A*, 75(1):012328, 2007.
- [8] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J Love, Alán Aspuru-Guzik, and Jeremy L O’Brien. A variational eigenvalue solver on a photonic quantum processor. *Nature Communications*, 5:4213, 2014.
- [9] Jarrod R McClean, Jonathan Romero, Ryan Babbush, and Alán Aspuru-Guzik. The theory of variational hybrid quantum-classical algorithms. *New Journal of Physics*, 18(2):023023, 2016.
- [10] Jarrod R McClean, Sergio Boixo, Vadim N Smelyanskiy, Ryan Babbush, and Hartmut Neven. Barren plateaus in quantum neural network training landscapes. *Nature Communications*, 9(1):1–6, 2018.
- [11] Sukin Sim, Jonathan Romero, Jérôme F Gonthier, and Alexander A Kunitsa. Adaptive pruning-based optimization of parameterized quantum circuits. *Quantum Science and Technology*, 6(2):025019, 2021.
- [12] Xavier Bonet-Monroig, Hao Wang, Diederick Vermetten, Bruno Senjean, Charles Moussa, Thomas Back, Vedran Dunjko, and Thomas E. O’Brien. Performance comparison of optimization methods on variational quantum algorithms. *arXiv preprint arXiv:2111.13454*, 2021.
- [13] David Poulin and Pawel Wocjan. Preparing ground states of quantum many-body systems on a quantum computer. *Physical Review Letters*, 102(13):130503, 2009.
- [14] Lin Lin and Yu Tong. Near-optimal ground state preparation. *Quantum*, 4:372, 2020.
- [15] Phillip WK Jensen, Lasse Bjørn Kristensen, Jakob S Kottmann, and Alán Aspuru-Guzik. Quantum computation of eigenvalues within target intervals. *Quantum Science and Technology*, 6(1):015004, 2020.
- [16] Anurag Anshu, David Gosset, Karen J Morenz Korol, and Mehdi Soleimanifar. Improved approximation algorithms for bounded-degree local hamiltonians. *Physical Review Letters*, 127(25):250502, 2021.

- [17] Mario Motta, Chong Sun, Adrian TK Tan, Matthew J O'Rourke, Erika Ye, Austin J Minnich, Fernando GSL Brandão, and Garnet Kin-Lic Chan. Determining eigenstates and thermal states on a quantum computer using quantum imaginary time evolution. *Nature Physics*, 16(2):205–210, 2020.
- [18] Sam McArdle, Tyson Jones, Suguru Endo, Ying Li, Simon C Benjamin, and Xiao Yuan. Variational ansatz-based quantum simulation of imaginary time evolution. *npj Quantum Information*, 5(1):1–6, 2019.
- [19] David Amaro, Carlo Modica, Matthias Rosenkranz, Mattia Fiorentini, Marcello Benedetti, and Michael Lubasch. Filtering variational quantum algorithms for combinatorial optimization. *Quantum Science and Technology*, 7(1):015021, 2021.
- [20] Min-Quan He, Dan-Bo Zhang, and Z. D. Wang. Quantum Gaussian filter for exploring ground-state properties. *Physical Review A*, 106(3):032420, 2022.
- [21] 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. *arXiv preprint arXiv:2204.05955*, 2022.
- [22] Jakob S. Kottmann, Philipp Schleich, Teresa Tamayo-Mendoza, and Alán Aspuru-Guzik. Reducing Qubit Requirements while Maintaining Numerical Precision for the Variational Quantum Eigensolver: A Basis-Set-Free Approach. *The Journal of Physical Chemistry Letters*, 12(1):663–673, 2021.
- [23] Jakob S. Kottmann and Alán Aspuru-Guzik. Optimized Low-Depth Quantum Circuits for Molecular Electronic Structure using a Separable Pair Approximation. *Physical Review A*, 105(3):032449, 2022.
- [24] Theodore J Yoder, Guang Hao Low, and Isaac L Chuang. Fixed-point quantum search with an optimal number of queries. *Physical Review Letters*, 113(21):210501, 2014.
- [25] Guoming Wang, Dax Enshan Koh, Peter D Johnson, and Yudong Cao. Minimizing estimation runtime on noisy quantum computers. *PRX Quantum*, 2(1):010346, 2021.
- [26] Oscar Higgott, Daochen Wang, and Stephen Brierley. Variational Quantum Computation of Excited States. *Quantum*, 3:156, 2019.
- [27] Robert M. Parrish, Edward G. Hohenstein, Peter L. McMahon, and Todd J. Martínez. Quantum Computation of Electronic Transitions Using a Variational Quantum Eigensolver. *Physical Review Letters*, 122(23):230401, 2019.
- [28] Ken M. Nakanishi, Kosuke Mitarai, and Keisuke Fujii. Subspace-search variational quantum eigensolver for excited states. *Physical Review Research*, 1(3), 2019.
- [29] Joonho Lee, William J. Huggins, Martin Head-Gordon, and K. Birgitta Whaley. Generalized Unitary Coupled Cluster Wave functions for Quantum Computation. *Journal of Chemical Theory and Computation*, 15(1):311–324, 2019.

- [30] Jakob S. Kottmann, Abhinav Anand, and Alán Aspuru-Guzik. A feasible approach for automatically differentiable unitary coupled-cluster on quantum computers. *Chemical Science*, 12(10):3497–3508, 2021.
- [31] Ruizhe Zhang, Guoming Wang, and Peter Johnson. Computing Ground State Properties with Early Fault-Tolerant Quantum Computers. *Quantum*, 6:761, 2022.