# Faster State Preparation with Randomization

Yue Wang,[1,2] Xiao-Ming Zhang,[3,4,*] Xiao Yuan,[5,6,†] and Qi Zhao[1,‡]

[1] *QICI Quantum Information and Computation Initiative, School of Computing and Data Science, The University of Hong Kong, Pokfulam Road, Hong Kong SAR, China*
[2] *Shenzhen International Quantum Research Institute, Shenzhen 518000, Guangdong, China*
[3] *Key Laboratory of Atomic and Subatomic Structure and Quantum Control (Ministry of Education), Guangdong Basic Research Center of Excellence for Structure and Fundamental Interactions of Matter, School of Physics, South China Normal University, Guangzhou 510006, China*
[4] *Guangdong Provincial Key Laboratory of Quantum Engineering and Quantum Materials, Guangdong-Hong Kong Joint Laboratory of Quantum Matter, Frontier Research Institute for Physics, South China Normal University, Guangzhou 510006, China*
[5] *Center on Frontiers of Computing Studies, Peking University, Beijing 100871, China*
[6] *School of Computer Science, Peking University, Beijing 100871, China*

Quantum state preparation remains a dominant cost in many quantum algorithms. We introduce a randomized protocol that fundamentally improves the accuracy-cost trade-off for states with hierarchical amplitude structures, where amplitudes decay exponentially or by a power law with exponent greater than one. Rather than commonly employed deterministically truncating small amplitudes, we prepare an ensemble of simple circuits: each retains all large amplitudes while amplifying a single small one. We rigorously prove that the randomized ensemble achieves a quadratic improvement in trace-distance error over the deterministic truncation method. This quadratic scaling halves the number of encoded amplitudes for exponential decay and yields polynomial reductions for power-law decay, which directly translates to a reduction in circuit depth. Demonstrations on molecular wavefunctions (LiH), many-body ground states (transverse-field Ising), and machine-learning parameters (ResNet) validate the approach across diverse applications. The method broadens the class of states preparable on near-term and fault-tolerant quantum devices.

*Introduction.*—Quantum state preparation [1–9] is a foundational primitive in quantum computing and the entry point for numerous algorithms in ground-state preparation [10], differential-equation solving [11–16], and quantum machine learning [17]. It is also one of the primary resource bottlenecks for many algorithms. Without solving the state-preparation problem, advantages in quantum algorithms such as HHL for solving linear systems [18] will be negated because state-preparation costs dominate the overall complexity. Many efficient quantum algorithms currently treat the preparation of complex initial states as an oracle and do not account for the implementation cost. However, to demonstrate real quantum advantage and solve practical problems, one must consider both state-preparation and data-readout costs.

Deterministic schemes for general quantum state preparation requires deep circuits or many ancillas [19, 20]. While substantial progress has been made for structured data such as sparse states [1–4, 6], conventional methods still require preparing a state that encodes all nonzero amplitudes, incurring costs proportional to the number of amplitudes. Moreover, many practically relevant states in chemistry, condensed matter, and machine learning exhibit prohibitively large numbers of amplitudes that challenge the scalability of existing techniques.

A common practical strategy is amplitude truncation, which discards coefficients below a threshold and prepare only the remaining state. This approach faces a fundamental accuracy-cost trade-off. To achieve error $\varepsilon$ via deterministic truncation, one must keep enough amplitudes to ensure the discarded $\ell_2$ norm is $\mathcal{O}(\varepsilon)$, directly tying approximation error to circuit depth and gate count. Moreover, preparing states with many small-magnitude amplitudes incurs additional overhead in fault-tolerant settings, where small-angle rotations decompose into long sequences of costly $T$ gates [21, 22]. These compounding resource demands motivate the search for methods that improve the accuracy-cost scaling. Crucially, many practical target states exhibit amplitude hierarchies—coefficients that decay exponentially or by a power law—making them natural candidates for more efficient approximation schemes.

Randomization has emerged as a powerful tool for improving quantum algorithms. Beyond noise tailoring via Pauli twirling [23], randomized approaches have accelerated Hamiltonian simulation [24, 25], phase estimation [26], and tomography [27], while enabling more efficient gate synthesis [28, 29]. Recent work shows that randomized constructions can suppress approximation errors quadratically and effectively halve resource costs for broad task families [30, 31]. Prior randomized state-synthesis methods [32, 33] reduce gate-decomposition errors but do not address the cost of encoding many amplitudes into quantum states—the bottleneck we target.

In this Letter, we introduce a randomized state-preparation protocol that fundamentally improves the truncation trade-off by exploiting amplitude hierarchies. Our key insight is to prepare not a single truncated state, but an ensemble: each state instance retains

all large-magnitude amplitudes and amplifies one small-magnitude amplitude. By sampling from this ensemble with probabilities proportional to the amplified coefficients, the resulting mixed state accurately recovers the ideal state in expectation while keeping individual circuit costs low. Beyond fewer encoded amplitudes, amplifying small coefficients reduces fault-tolerant resource requirements because larger amplitudes require less precise rotations, yielding fewer $T$ gates. We rigorously prove that this randomized scheme achieves trace-distance error scaling as $\mathcal{O}(\varepsilon^2)$, a quadratic improvement over $\mathcal{O}(\varepsilon)$ in deterministic truncation. For states with exponential amplitude decay, this improvement halves the number of amplitudes that must be encoded to reach a target accuracy; for power-law decay with exponent greater than one, it provides a polynomial reduction, translating directly to circuit-depth and gate-count savings. Our contributions are threefold: (i) a general mixing framework with rigorous, state-level trace-norm error bounds; (ii) constructive circuits that reduce small-angle rotations and $T$-gate counts; and (iii) empirical validation on molecular (LiH), many-body (transverse-field Ising), and machine-learning (ResNet) states, demonstrating practical resource savings. These results broaden the class of states feasibly preparable on quantum devices and reduce a dominant cost in quantum algorithms that currently treat state preparation as an oracle [18].

*Conventional Truncation Method.*—Given a general $n$-qubit quantum state

$$|\psi\rangle = \sum_{i=0}^{2^n-1} \alpha_i |i\rangle , \tag{1}$$

the problem we consider is constructing an approximated state $\rho_{\text{approx}}$ (typically from $|0\rangle^{\otimes n}$) such that the error is below a threshold, i.e., $\| |\psi\rangle\langle\psi| - \rho_{\text{approx}} \|_1 \leqslant \varepsilon$. Suppose Eq. (1) has $d$ nonzero amplitudes $\alpha_j$ and no error is allowed; then any state-preparation protocol has a circuit-size lower bound $\Omega(d+n)$[34]. This cost can be prohibitive in practice. A straightforward mitigation is to discard insignificant coefficients thus reduce $d$, introducing a controlled approximation error. Such truncation methods are widely used, including in low-rank approximations for quantum machine learning [17, 35, 36], finite-range interaction truncation in condensed-matter simulations [37, 38], and quantum chemistry [39–42].

Formally, the truncation method works as follows: we partition the coefficients into two subsets based on a cutoff threshold $t$:

$$A = \{i : \alpha_i \geq t\} \qquad B = \{i : \alpha_i < t\}. \tag{2}$$

Therefore, set $A$ contains the indices with significant coefficients, and set $B$ contains the remaining indices. If we discard set $B$ entirely and reconstruct the approximate state as $|\psi_A\rangle = \sum_{i\in A} \alpha_i |i\rangle$, then the cutoff error is bounded by $\mathcal{O}\left(\sqrt{\sum_{j\in B} |\alpha_j|^2}\right)$. This error bound directly follows from the trace distance between pure states: $\| |\psi\rangle\langle\psi| - |\phi\rangle\langle\phi| \|_1 = 2\sqrt{1-|\langle\psi|\phi\rangle|^2}$, where the overlap between $|\psi\rangle$ and $|\psi_A\rangle$ is $|\langle\psi|\psi_A\rangle|^2 = \sum_{i\in A} |\alpha_i|^2$. The specific threshold for partitioning can vary depending on the task, provided the set $B$ contains enough terms to meaningfully reduce complexity. In quantum chemistry, for instance, set $A$ might contain the Hartree–Fock (HF) coefficients, while set $B$ includes higher-order excitations.

**Algorithm 1:** Randomized Quantum State Preparation Protocol

**Input:** Quantum state $|\psi\rangle = \sum_i \alpha_i |i\rangle$ with indices partitioned into $A = \{i : \alpha_i \geq t\}$ and $B = \{i : \alpha_i < t\}$.

**Output:** Approximated quantum state $\rho_{\text{approx}}$ as a randomized ensemble of states $\left|\tilde{\psi}_m\right\rangle$.

1 Draw one element from the set $B$ with probabilities $\{p_m\}$ such that $\sum_m p_m = 1$. Define the corresponding state:

$$|\psi_m\rangle = \sum_{i\in A} \alpha_i |i\rangle + \frac{\alpha_m}{p_m} |m\rangle . \tag{3}$$

2 Normalize the state:

$$\left|\tilde{\psi}_m\right\rangle = \Gamma_m |\psi_m\rangle , \tag{4}$$

where $\Gamma_m = \left(\sum_{i\in A} |\alpha_i|^2 + \frac{|\alpha_m|^2}{p_m^2}\right)^{-1/2}$ is the normalization factor.

3 Sample an index $m$ with probability $p_m$ and prepare $\left|\tilde{\psi}_m\right\rangle$. The final prepared state is a mixture:

$$\rho_{\text{approx}} = \sum_m p_m \left|\tilde{\psi}_m\right\rangle \left\langle\tilde{\psi}_m\right| . \tag{5}$$

*Randomized Approach.*—Instead of discarding the coherence carried by set $B$, we use classical randomization to preserve it in expectation while keeping per-instance circuits simple. We illustrate our framework in Alg. 1. Each state $|\psi_m\rangle$ in the ensemble contains all amplitudes in $A$ and a single element in $B$ with an amplified coefficient. The cost of implementing each $|\psi_m\rangle$ is essentially slightly more than that of preparing $|\psi_A\rangle$, as only one additional coefficient is included. Nevertheless, the resulting probabilistic mixture closely approximates the ideal state. We choose probabilities so that the reconstruction condition holds: $\sum_m \frac{1}{p_m}\alpha_m |m\rangle = \sum_{j\in B} \alpha_j |j\rangle$, where $p_m = |\alpha_m| / \sum_{j\in B} |\alpha_j|$ is the selection probability corresponding to index $m$. Because the coefficients $|\alpha_m|$ are small, the amplification term $\alpha_m/p_m$ is small relative to $\sum_{i\in A} |\alpha_i|^2$. Consequently, $\Gamma_m \approx 1$, and each $\left|\tilde{\psi}_m\right\rangle$ deviates only slightly from the target amplitude structure.

We first derive a state-level mixing lemma that quan-

tifies the error for mixtures of pure states (proof in [43]) and use it to bound our protocol.

**Lemma 1.** *Let $|\psi\rangle$ be a target pure quantum state, and let $\{|\psi_m\rangle\}_m$ be a family of pure states with associated probabilities $\{p_m\}_m$ satisfying: $\| |\psi_m\rangle - |\psi\rangle \| \le a, \forall m$, and $\|\sum_m p_m |\psi_m\rangle - |\psi\rangle\| \le b$, for some $a, b > 0$. Define the target and mixed density matrices $\rho := |\psi\rangle\langle\psi|$, and $\rho_{\text{mix}} := \sum_m p_m |\psi_m\rangle\langle\psi_m|$. Then, the trace distance between the mixed and target states satisfies*

$$\|\rho_{\text{mix}} - \rho\|_1 \le a^2 + 2b. \tag{6}$$

Here $a$ is a per-sample deviation and $b$ is the mean bias. This lemma provides a bound on how well a mixture of approximating pure states reproduces the original pure state. We consider the trace-norm distance between the target state $|\psi\rangle$ and the approximated ensemble state. Using Lemma 1 and carefully accounting for the contribution of each term to the final error, our main result is as follows.

**Theorem 1.** *Let $|\psi\rangle = |\psi_A\rangle + \sum_{j\in B} \alpha_j |j\rangle$, and $\epsilon^2 := \sum_{j\in B} \alpha_j^2$, with real amplitudes $\alpha_j$. Assume the $\ell_1$–smallness condition*

$$S := \sum_{j\in B} |\alpha_j| \le c\epsilon, \qquad c = O(1). \tag{7}$$

*Let $\left|\tilde{\psi}_m\right\rangle := \Gamma^{-1} |\psi_m\rangle$ be the normalized state, where $\Gamma := \|\psi_m\| = \sqrt{1-\epsilon^2+S^2}$. Then*

$$\begin{aligned} a &:= \max_{m\in B} \left\| \left|\tilde{\psi}_m\right\rangle - |\psi\rangle \right\| \le (c+1)\epsilon + O(\epsilon^2), \\ b &:= \left\| \sum_m p_m \left|\tilde{\psi}_m\right\rangle - |\psi\rangle \right\| = O(\epsilon^2), \end{aligned} \tag{8}$$

*and consequently $\|\rho_{\text{approx}} - \rho\|_1 \le a^2 + 2b = O(\epsilon^2)$.*

Theorem 1 (proof in [43]) indicates that our protocol leads to a quadratically improved error scaling compared to naive truncation methods. The assumption of the $\ell_1$-smallness condition is justified in the supplementary material [43]. In short, it holds if the coefficients with indices in $B$ decay faster than a power law with exponent greater than 1. This theorem holds even if an individual amplification factor $p_m$ is exponentially small, because the corresponding coefficient $\alpha_m$ is similarly small, and the collective effect is balanced. Therefore, the final error is well bounded. The trace-norm error directly provides a bound for observable measurements: measuring the ensemble state $\rho_{\text{approx}}$ yields results with accuracy scaling $\left|\text{Tr}[(|\psi\rangle\langle\psi| - \rho_{\text{approx}})O]\right| \le \mathcal{O}\left(\|O\| \sum_{j\in B} \alpha_j^2\right)$.

*Resource Savings.*—We now translate the quadratic error suppression into resource savings. Let the amplitudes be ordered by decreasing magnitude, $|\alpha|_{(1)} \ge |\alpha|_{(2)} \ge \cdots \ge |\alpha|_{(N)}$. Given a cutoff $t$, deterministic truncation yields error $\mathcal{O}(\epsilon)$, whereas the randomized mixture attains $\mathcal{O}(\epsilon^2)$. Let $K$ denote the number of kept amplitudes (i.e., $K = |A|$) and let $\epsilon(K)$ be the tail $\ell_2$ norm after keeping the top-$K$ terms. The required $K$ for a given target accuracy follows by inverting $\epsilon(K)$ under two decay models: (i) exponential decay: $\epsilon(K) = \Theta(r_{\text{exp}}^K)$, where $r_{\text{exp}} < 1$ is the decay rate. To achieve error $\tau$, truncation requires $K_{\text{det}}(\tau) = \Theta\big(\log(1/\tau)/\log(1/r_{\text{exp}})\big)$, while randomized requires $K_{\text{rand}}(\tau) = \Theta\big(\frac{1}{2}\log(1/\tau)/\log(1/r_{\text{exp}})\big)$, and (ii) power-law decay: $\epsilon(K) = \Theta\big(K^{\frac{1-2r_{\text{pow}}}{2}}\big)$, where $r_{\text{pow}} > 1$ is the decay exponent. Thus $K_{\text{det}}(\tau) = \Theta\big(\tau^{-\frac{2}{2r_{\text{pow}}-1}}\big)$ and $K_{\text{rand}}(\tau) = \Theta\big(\tau^{-\frac{1}{2r_{\text{pow}}-1}}\big)$. Consequently, the extra amplitudes a deterministic scheme must include beyond the randomized cutoff to match the same target error are

$$\frac{K_{\text{det}}(\tau)}{K_{\text{rand}}(\tau)} = \begin{cases} 2, & \text{geometric decay}, \\ \Theta\big(\tau^{-\frac{1}{2r_{\text{pow}}-1}}\big), & \text{power-law decay}. \end{cases}$$

Since implementation cost is at least proportional to the number of prepared nonzero amplitudes, these $K$-scalings translate directly to resource savings. Hence the randomized protocol yields a constant-factor halving of prepared amplitudes under exponential decay, and a polynomial reduction by $\tau^{-\frac{1}{2r_{\text{pow}}-1}}$ under power-law decay, directly translating to commensurate circuit-cost savings.

*Quantum Chemistry States.*—The randomized state-preparation approach is particularly beneficial for quantum chemistry applications [39–42], notably in configuration interaction (CI) methods. In CI, the many-electron wavefunction is represented as a linear combination of Slater determinants, inherently exhibiting a hierarchical structure. Typically, the largest amplitudes correspond to the reference determinant (often the HF solution) and low-lying excitations, whereas amplitudes for higher-order excitations diminish rapidly due to large energy denominators.

This hierarchy aligns directly with our randomized method: set $A$ comprises the large coefficients from the reference determinant and significant low-order excitations, whereas higher-order excitations fall into set $B$. The energy-based suppression of high-order excitations ensures that $\sum_{j\in B} |\alpha_j|^2$ is small and decaying.

Preparing the full configuration interaction (FCI) state is computationally prohibitive due to the exponential growth in determinants with electron and orbital counts. Consequently, practical workflows employ truncated expansions (e.g., singles, doubles, or limited higher excitations). Such truncations integrate naturally with our randomized preparation, reducing quantum resources by avoiding explicit reconstruction of the entire FCI state in a single instance.

The many-electron wavefunction is expanded in a basis of Slater determinants, $|\Psi_{\text{CI}}\rangle = \sum_I c_I |\Phi_I\rangle$, where each

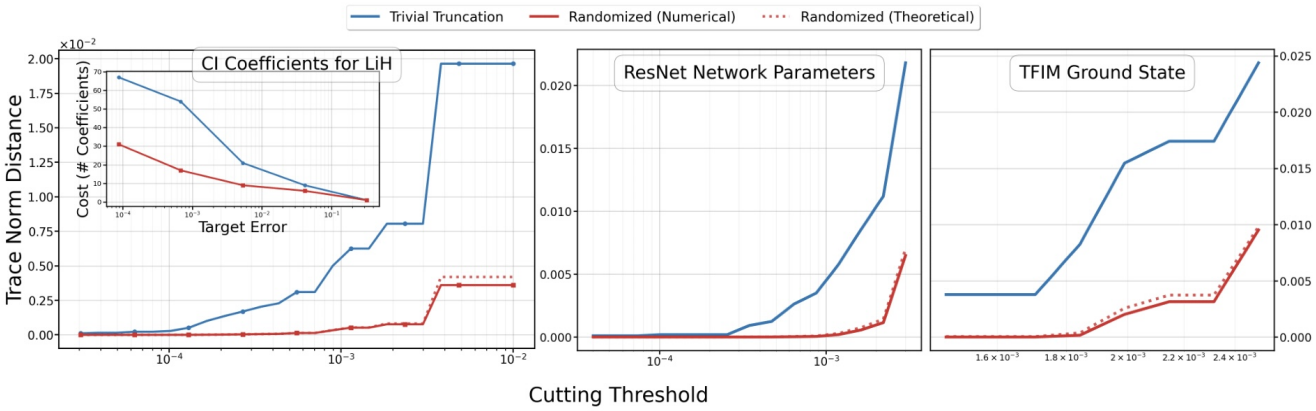


FIG. 1. CI coefficients for LiH: trace-norm distance versus threshold comparing randomized and cutoff reconstructions. As the threshold decreases and fewer terms are discarded, both errors drop, with the randomized method exhibiting a quadratically faster reduction. The inset shows the reduction in the number of coefficients. At a target error of $10^{-4}$, the number of kept coefficients is reduced by 53%. Further numerical results for (i) a state built from a simplified ResNet's parameters using 10 qubits and (ii) the TFIM ground state at $N = 11$ with $J = h = 1.0$ illustrate the accuracy advantages across diverse settings.

$|\Phi_I\rangle$ is a Slater determinant and $c_I$ is its CI coefficient. The reference determinant is $|\Phi_0\rangle$. Excited determinants arise by promoting one or more electrons into unoccupied orbitals. In the supplementary material we show that higher-order determinants have suppressed coefficients, e.g., $|c_I^{(1)}| \leq \|\hat{V}\|/\Delta E_{I0}$, where $\Delta E_{I0} = E_I^{(0)} - E_0^{(0)}$ and $\hat{V}$ is a perturbing operator.

*Numerical Results.*—We validate our framework on three problems: the ground state of LiH, the ground state of the transverse-field Ising model (TFIM), and a quantum state encoding parameters of a trained machine-learning model. These represent both physically structured states and high-dimensional data-driven states, demonstrating broad applicability.

For LiH, we place lithium at the origin and hydrogen 1.6 Å apart, adopting the minimal STO-3G basis, which yields twelve spin orbitals for four electrons. Starting from atomic orbital integrals, we perform a restricted HF calculation to obtain canonical molecular orbitals, then construct the FCI Hamiltonian. Diagonalization on a classical computer produces the ground-state wavefunction, $|\Psi_{\mathrm{CI}}\rangle = \sum_I c_I |\Phi_I\rangle$, where $\{|\Phi_I\rangle\}$ are Slater determinants (configurations) labeled by occupied orbitals, and $c_I$ are the associated CI coefficients.

We map the wavefunction to 12 qubits via Jordan–Wigner encoding and form the ensemble density matrix $\rho_{\mathrm{approx}} = \sum_m p_m \left|\tilde{\psi}_m\right\rangle \left\langle\tilde{\psi}_m\right|$, where each $\left|\tilde{\psi}_m\right\rangle$ incorporates the amplified amplitude for $|m\rangle$ plus all large amplitudes in $A$. Comparing to the FCI state using the trace-norm distance, we consistently achieve near-FCI accuracy at substantially reduced gate cost, in line with the theoretical bounds.

For the one-dimensional TFIM with periodic boundary conditions on $N = 11$ qubits, the Hamiltonian is $H = -J\sum_{i=0}^{N-1} Z_i Z_{i+1} - h\sum_{i=0}^{N-1} X_i$, where $X_i$ and $Z_i$ act on qubit $i$. At $J = h = 1.0$, the ground state is a dense superposition over all $2^{11}$ computational-basis states. We obtain it by diagonalizing the $2048 \times 2048$ Hamiltonian, yielding the coefficients $\{\alpha_i\}$ used in our protocol.

Finally, we construct a quantum state from the parameters of a trained ResNet on CIFAR-10. We load the final parameter vector (526 floating-point values), normalize it, and map it to amplitudes of a 10-qubit state, filling the first 526 computational-basis states. This yields a dense state $|\Psi_{\mathrm{ML}}\rangle = \sum_{i=0}^{526} w_i |i\rangle$ with $w_i$ the normalized weights. As shown in Fig. 1, the results exhibit the predicted quadratic improvement; the observed trace-norm distance closely follows the theoretical upper bound $((c+2)^2 + c/2)\sum_{j\in B} |\alpha_j|^2$.

*Discussion.*—We have presented a randomized state-preparation approach that reduces circuit complexity while maintaining high trace norm error. The method leverages hierarchical amplitude structures common in practice, such as chemistry, condensed-matter systems, and machine learning. Our analysis establishes error bounds proportional to the square of the discarded tail weight, yielding a quadratic improvement over direct truncation. Numerical simulations corroborate these gains in both accuracy and resource utilization.

The core intuition behind our randomized method exploits the amplitude hierarchy inherent in realistic quantum states. Typically, the complexity of state preparation is directly tied to the number of nonzero amplitudes required for state description. However, many amplitudes are negligible in magnitude yet numerous; each contributes little on its own, but together they exert a meaningful collective influence. Randomized approach avoid preparing all tiny amplitudes in one run thus reduce the resource needed. Unlike traditional deterministic methods, which uniformly address all coefficients, our

approach prioritizes amplitude subsets based on significance, achieving a practical balance between trace norm error and circuit complexity.

Moreover, amplifying small coefficients can reduce the quantum resources needed for state preparation because we can employ less accurate gate synthesis. In the fault-tolerant regime, a rotation by a very small angle must be decomposed into a sequence of Clifford and $T$ gates [21, 22], where $T$ gates are widely regarded as an expensive resource [44, 45]. Recent advances, such as magic-state cultivation [46], suggest that $T$ gates might eventually be implemented with overhead comparable to CNOT gates; nevertheless, such techniques are not yet fully mature, and $T$ gates remain more resource-intensive than simpler Clifford operations. Consequently, by amplifying the small-magnitude terms, one effectively raises the minimum amplitude in the resulting density matrix. This, in turn, reduces the number of $T$ gates required for small-angle rotations, thereby offering a more resource-efficient pathway to quantum state preparation in fully fault-tolerant architectures. Such an idea has also been employed to reduce $T$-gate counts in Trotter-based Hamiltonian simulation [47].

Previous applications of randomization to state-synthesis error [32, 33] differ from our work in scope and results, and thus barely overlap with our randomized state-preparation framework. Those works consider a semidefinite programming optimization problem to quadratically improve the worst-case approximation error by identifying a set of allowed quantum states that are synthesis-error-free in the fault-tolerant setting. Since gate-synthesis error commonly scales as $\tilde{\mathcal{O}}(1/\epsilon)$, this approach halves the compilation length. We, in contrast, consider the cost of encoding a classical vector description in a quantum state. Our cost savings depend on the decay pattern of the amplitudes: with exponential decay, we also halve the cost, but with power-law decay, we achieve a polynomial speedup in cost, greatly outperforming their result. Moreover, since their work and ours target separate sources of error in the state-preparation problem, the results can be combined to achieve greater cost savings.

Our framework is well suited to practical implementation. The states in the ensemble $\left|\tilde{\psi}_m\right\rangle$ are very similar, differing only in which sign on the state from $B$ and in the corresponding normalization factor. Thus, the state-preparation procedures for the states in the ensemble are highly similar, favoring practical implementation of the randomized scheme. In summary, our approach reduces error for truncated state preparation, making it promising for near-term and fault-tolerant devices. The rigorous guarantees ensure that prepared states closely approximate targets with controllable error, enabling more efficient implementations of algorithms that rely on complex state preparation.

*Note added.*—During the preparation of the manuscript, we note a recent work by Harrow et al. [48], who studied optimal randomized approximation of arbitrary pure states by convex mixtures of $k$-sparse states. While the studied problems are related, the two methods are different and complementary.

Q.Z. acknowledges funding from Innovation Program for Quantum Science and Technology via Project 2024ZD0301900, National Natural Science Foundation of China (NSFC) via Project No. 12347104 and No. 12305030, Guangdong Basic and Applied Basic Research Foundation via Project 2023A1515012185, Hong Kong Research Grant Council (RGC) via No. 27300823, N_HKU718/23, and R6010-23, Guangdong Provincial Quantum Science Strategic Initiative No. GDZX2303007, HKU Seed Fund for Basic Research for New Staff via Project 2201100596.

* phyxmz@gmail.com
† xiaoyuan@pku.edu.cn
‡ zhaoqi@cs.hku.hk

[1] N. Gleinig and T. Hoefler, An Efficient Algorithm for Sparse Quantum State Preparation, in *2021 58th ACM/IEEE Design Automation Conference (DAC)* (IEEE, San Francisco, CA, USA, 2021) pp. 433–438.
[2] X.-M. Zhang, T. Li, and X. Yuan, Quantum State Preparation with Optimal Circuit Depth: Implementations and Applications, Physical Review Letters **129**, 230504 (2022).
[3] T. M. L. de Veras, L. D. da Silva, and A. J. da Silva, Double sparse quantum state preparation, Quantum Information Processing **21**, 204 (2022).
[4] R. Mao, G. Tian, and X. Sun, Towards Optimal Circuit Size for Sparse Quantum State Preparation (2024), arXiv:2404.05147 [quant-ph].
[5] I. F. Araujo, D. K. Park, F. Petruccione, and A. J. da Silva, A divide-and-conquer algorithm for quantum state preparation, Scientific Reports **11**, 6329 (2021).
[6] J. Luo and L. Li, Circuit Complexity of Sparse Quantum State Preparation (2024), arXiv:2406.16142 [quant-ph].
[7] C. F. Kane, N. Gomes, and M. Kreshchuk, Nearly-optimal state preparation for quantum simulations of lattice gauge theories (2024), arXiv:2310.13757 [hep-lat, physics:quant-ph].
[8] X. Sun, G. Tian, S. Yang, P. Yuan, and S. Zhang, Asymptotically Optimal Circuit Depth for Quantum State Preparation and General Unitary Synthesis, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems , 1 (2023).
[9] X.-M. Zhang, M.-H. Yung, and X. Yuan, Low-depth quantum state preparation, Physical Review Research **3**, 043200 (2021).
[10] L. Lin and Y. Tong, Near-optimal ground state preparation, Quantum **4**, 372 (2020).
[11] D. W. Berry, A. M. Childs, A. Ostrander, and G. Wang, Quantum algorithm for linear differential equations with exponentially improved dependence on precision, Communications in Mathematical Physics **356**, 1057 (2017),

arxiv:1701.03684 [quant-ph].
[12] J.-P. Liu, H. Ø. Kolden, H. K. Krovi, N. F. Loureiro, K. Trivisa, and A. M. Childs, Efficient quantum algorithm for dissipative nonlinear differential equations, Proceedings of the National Academy of Sciences **118**, e2026805118 (2021).
[13] D. An and L. Lin, Quantum linear system solver based on time-optimal adiabatic quantum computing and quantum approximate optimization algorithm, ACM Transactions on Quantum Computing **3**, 1 (2022), number: 2 arXiv:1909.05500 [quant-ph].
[14] D. An, J.-P. Liu, D. Wang, and Q. Zhao, A theory of quantum differential equation solvers: Limitations and fast-forwarding (2023), arXiv:2211.05246 [quant-ph].
[15] H. Krovi, Improved quantum algorithms for linear and nonlinear differential equations, Quantum **7**, 913 (2023), arXiv:2202.01054 [physics, physics:quant-ph].
[16] D. Fang, L. Lin, and Y. Tong, Time-marching based quantum solvers for time-dependent linear differential equations, Quantum **7**, 955 (2023), arXiv:2208.06941 [quant-ph].
[17] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Quantum machine learning, Nature **549**, 195 (2017).
[18] A. W. Harrow, A. Hassidim, and S. Lloyd, Quantum algorithm for solving linear systems of equations, Physical Review Letters **103**, 150502 (2009), arXiv:0811.3171 [quant-ph].
[19] M. Plesch and i. c. v. Brukner, Quantum-state preparation with universal gate decompositions, Phys. Rev. A **83**, 032302 (2011).
[20] V. V. Shende, I. L. Markov, and S. S. Bullock, Smaller two-qubit circuits for quantum communication and computation, in *Proceedings of the Design, Automation and Test in Europe Conference (DATE)* (IEEE, 2004) pp. 980–985.
[21] G. J. Mooney, C. D. Hill, and L. C. L. Hollenberg, Cost-optimal single-qubit gate synthesis in the Clifford hierarchy, Quantum **5**, 396 (2021).
[22] N. J. Ross and P. Selinger, Optimal ancilla-free Clifford+T approximation of z-rotations (2016), arXiv:1403.2975 [quant-ph].
[23] J. J. Wallman and J. Emerson, Noise tailoring for scalable quantum computation via randomized compiling, Physical Review A **94**, 052325 (2016).
[24] A. M. Childs, A. Ostrander, and Y. Su, Faster quantum simulation by randomization, Quantum **3**, 182 (2019).
[25] E. Campbell, A random compiler for fast Hamiltonian simulation, Physical Review Letters **123**, 070503 (2019), arXiv:1811.08017 [quant-ph].
[26] K. Wan, M. Berta, and E. T. Campbell, Randomized Quantum Algorithm for Statistical Phase Estimation, Physical Review Letters **129**, 030503 (2022).
[27] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, Nature Physics **16**, 1050 (2020).
[28] M. B. Hastings, Turning Gate Synthesis Errors into Incoherent Errors, https://arxiv.org/abs/1612.01011v1 (2016).
[29] E. Campbell, Shorter gate sequences for quantum computing by mixing unitaries, Physical Review A **95**, 042306 (2017).
[30] Y. Wang and Q. Zhao, Faster Quantum Algorithms with "Fractional"-Truncated Series (2024), arXiv:2402.05595 [quant-ph].
[31] J. M. Martyn and P. Rall, Halving the Cost of Quantum Algorithms with Randomization (2024), arXiv:arXiv:2409.03744 [quant-ph].
[32] S. Akibue, G. Kato, and S. Tani, Probabilistic state synthesis based on optimal convex approximation, npj Quantum Information **10**, 3 (2024), arXiv:2303.10860 [quant-ph].
[33] S. Akibue, G. Kato, and S. Tani, Quadratic improvement on accuracy of approximating pure quantum states and unitary gates by probabilistic implementation (2022), arXiv:2111.05531 [quant-ph].
[34] This is a trivial lower bound on the cost of state preparation; the state-of-the-art protocol achieves $\mathcal{O}\left(dn/\log(dn)+n\right)$ [6].
[35] S. Lloyd, M. Mohseni, and P. Rebentrost, Quantum principal component analysis, Nature Physics **10**, 631 (2014).
[36] J. Liu, M. Liu, J.-P. Liu, Z. Ye, Y. Wang, Y. Alexeev, J. Eisert, and L. Jiang, Towards provably efficient quantum algorithms for large-scale machine-learning models, Nature Communications **15**, 434 (2024).
[37] R. Babbush, N. Wiebe, J. McClean, J. McClain, H. Neven, and G. K.-L. Chan, Low Depth Quantum Simulation of Electronic Structure, https://arxiv.org/abs/1706.00023v3 (2017).
[38] R. Babbush, C. Gidney, D. W. Berry, N. Wiebe, J. McClean, A. Paler, A. Fowler, and H. Neven, Encoding Electronic Spectra in Quantum Circuits with Linear T Complexity, Physical Review X **8**, 041015 (2018).
[39] B. P. Lanyon, J. D. Whitfield, G. G. Gillet, M. E. Goggin, M. P. Almeida, I. Kassal, J. D. Biamonte, M. Mohseni, B. J. Powell, M. Barbieri, A. Aspuru-Guzik, and A. G. White, Towards Quantum Chemistry on a Quantum Computer, Nature Chemistry **2**, 106 (2010), arXiv:0905.0887 [quant-ph].
[40] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, Quantum computational chemistry, Reviews of Modern Physics **92**, 015003 (2020).
[41] J. Romero, R. Babbush, J. R. McClean, C. Hempel, P. J. Love, and A. Aspuru-Guzik, Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz, Quantum Science and Technology **4**, 014008 (2018).
[42] T. Helgaker, P. Jorgensen, and J. Olsen, *Molecular electronic-structure theory* (John Wiley & Sons, 2013).
[43] See supplemental material, .
[44] C. Gidney, Inplace Access to the Surface Code Y Basis, Quantum **8**, 1310 (2024).
[45] T. Itogawa, Y. Takada, Y. Hirano, and K. Fujii, Even more efficient magic state distillation by zero-level distillation (2024), arXiv:2403.03991 [quant-ph].
[46] C. Gidney, N. Shutty, and C. Jones, Magic state cultivation: Growing T states as cheap as CNOT gates (2024), arXiv:2409.17595 [quant-ph].
[47] E. Granet and H. Dreyer, Hamiltonian dynamics on digital quantum computers without discretization error, npj Quantum Information **10**, 1 (2024).
[48] A. W. Harrow, A. Lowe, and F. Witteveen, Randomized truncation of quantum states (2025), preprint; details to be updated, arXiv:placeholder [quant-ph].

## Appendix A: State-Level Mixing Lemma

We first propose a state-mixing lemmas.

**Lemma 2.** *Let $|\psi\rangle$ be a target pure quantum state, and let $\{|\psi_m\rangle\}_m$ be a family of pure states with associated probabilities $\{p_m\}_m$ satisfying:*

$$\| \, |\psi_m\rangle - |\psi\rangle \, \| \leq a \quad \textit{for all } m, \qquad \left\| \sum_m p_m \, |\psi_m\rangle - |\psi\rangle \right\| \leq b, \tag{A1}$$

*for some constants $a, b > 0$. Define the target and mixed density matrices as*

$$\rho := |\psi\rangle \langle\psi| \, , \qquad \rho_{\text{mix}} := \sum_m p_m \, |\psi_m\rangle \langle\psi_m| \, . \tag{A2}$$

*Then, the trace norm between the mixed and target states satisfies*

$$\|\rho_{\text{mix}} - \rho\|_1 \leq a^2 + 2b. \tag{A3}$$

*Proof.* Define the deviation vectors $\delta_m := |\psi_m\rangle - |\psi\rangle$, which satisfy $\|\delta_m\| \leq a$. Let their average deviation be

$$\Delta := \sum_m p_m \delta_m = \sum_m p_m (|\psi_m\rangle - |\psi\rangle), \tag{A4}$$

which satisfies $\|\Delta\| \leq b$ by assumption (A1).

Expanding each projector:

$$|\psi_m\rangle \langle\psi_m| = (|\psi\rangle + \delta_m)(\langle\psi| + \delta_m^\dagger) = \rho + |\psi\rangle \langle\delta_m| + |\delta_m\rangle \langle\psi| + |\delta_m\rangle \langle\delta_m| \, . \tag{A5}$$

Taking the weighted sum and subtracting $\rho$, we obtain:

$$\begin{aligned} \rho_{\text{mix}} - \rho &= \sum_m p_m \left( |\psi\rangle \langle\delta_m| + |\delta_m\rangle \langle\psi| + |\delta_m\rangle \langle\delta_m| \right) \\ &= |\psi\rangle \langle\Delta| + \Delta \langle\psi| + \sum_m p_m \, |\delta_m\rangle \langle\delta_m| \, . \end{aligned} \tag{A6}$$

We now bound the trace norm of each term in Eq. (A6).

Define the operator

$$X := |\psi\rangle \langle\Delta| + \Delta \langle\psi| \, . \tag{A7}$$

This is a Hermitian operator of rank at most 2. By the triangle inequality and the identity $\| \, |u\rangle \langle v| \, \|_1 = \|u\| \|v\|$, we get

$$\|X\|_1 \leq \| \, |\psi\rangle \langle\Delta| \, \|_1 + \|\Delta \langle\psi| \, \|_1 = 2\|\psi\| \cdot \|\Delta\| = 2\|\Delta\| \leq 2b. \tag{A8}$$

Each term $|\delta_m\rangle \langle\delta_m|$ has trace norm $\|\delta_m\|^2 \leq a^2$. Thus,

$$\left\| \sum_m p_m \, |\delta_m\rangle \langle\delta_m| \right\|_1 = \sum_m p_m \|\delta_m\|^2 \leq a^2 \sum_m p_m = a^2. \tag{A9}$$

Combining Eqs. (A8) and (A9) in Eq. (A6), we obtain the desired bound:

$$\|\rho_{\text{mix}} - \rho\|_1 \leq \|X\|_1 + \left\| \sum_m p_m \, |\delta_m\rangle \langle\delta_m| \right\|_1 \leq 2b + a^2. \tag{A10}$$

This proves the lemma. □

This lemma provides a bound on how well a convex mixture of approximating pure states reproduces the original pure state at the level of density operators. Importantly, the bound depends only on the worst-case deviation $a$ and the average deviation $b$. The probabilities $p_m$ may be arbitrarily small—the proof depends only on their normalization $\sum_m p_m = 1$. This lemma is particularly useful when designing randomized state preparation protocols where each sampled state is close to the target, but their average is even closer.

## Appendix B: Trace Norm Error Bound

With lemma.1, we can bound the error in the random mixing protocol by finding the bound on $a$ and $b$ in our case. We begin by re-stating the definition of our quantum state ensemble.

**Definition 1** (Large-/Small-Amplitude Partition)**.** *An $n$-qubit state with non-zero real coefficients $\alpha_i$*

$$|\psi\rangle = \sum_{i=0}^{2^n-1} \alpha_i |i\rangle \tag{B1}$$

*is split into two disjoint index sets by a threshold value $t$:*

$$A = \{i : |\alpha_i| \geq t\},$$
$$B = \{j : |\alpha_j| < t\}, \quad A \cap B = \varnothing, \quad A \cup B = \{0, \ldots, 2^n - 1\}.$$

*Then, we create an ensemble of quantum states $\{|\psi_m\rangle = \sum_{i\in A} \alpha_i |i\rangle + \frac{\alpha_m}{p_m} |m\rangle\}$, where $p_m = \frac{|\alpha_m|}{\sum_{j\in B} |\alpha_j|}$ is the selection probability of choosing the corresponding quantum state in the random mixing protocol.*

**Theorem 2** (Random mixing error bound)**.** *Let*

$$|\psi\rangle \;=\; |\psi_A\rangle \;+\; \sum_{j\in B} \alpha_j |j\rangle\,, \qquad \|\psi_A\|^2 = 1 - \epsilon^2, \qquad \epsilon^2 := \sum_{j\in B} \alpha_j^2, \tag{B2}$$

*with real amplitudes $\alpha_j$. Assume the $\ell_1$–smallness condition*

$$S \;:=\; \sum_{j\in B} |\alpha_j| \;\leq\; c\epsilon, \qquad c = O(1). \tag{B3}$$

*For every $m \in B$ define the probability $p_m := |\alpha_m|/S$ and the unnormalized state*

$$|\psi_m\rangle := |\psi_A\rangle \;+\; \frac{\alpha_m}{p_m} |m\rangle = |\psi_A\rangle + S\,sgn(\alpha_m)\,|m\rangle\,, \qquad \Gamma := \|\psi_m\| = \sqrt{1 - \epsilon^2 + S^2}. \tag{B4}$$

*Let $|\psi'_m\rangle := \Gamma^{-1} |\psi_m\rangle$ be the normalized state. Then*

$$a := \max_{m\in B} \||\psi'_m\rangle - |\psi\rangle\| \;\leq\; (c+2)\epsilon + O(\epsilon^2), \qquad b := \left\| \sum_m p_m |\psi'_m\rangle - |\psi\rangle \right\| \;=\; O(\epsilon^2), \tag{B5}$$

*and consequently $\|\rho_{\text{mix}} - \rho\|_1 \leq a^2 + 2b = O(\epsilon^2)$.*

*Proof.* Using $|\psi_m\rangle = |\psi_A\rangle + \alpha_m/p_m |m\rangle$ and $|\psi\rangle = |\psi_A\rangle + \alpha_m |m\rangle + \sum_{k\neq m} \alpha_k |k\rangle$ we obtain

$$\begin{aligned} |\psi'_m\rangle - |\psi\rangle &= \Gamma^{-1} |\psi_m\rangle - |\psi\rangle \\ &= \left(\Gamma^{-1} - 1\right) |\psi_A\rangle + \left(\tfrac{\alpha_m}{p_m \Gamma} - \alpha_m\right) |m\rangle - \sum_{k\neq m} \alpha_k |k\rangle\,. \end{aligned} \tag{B6}$$

By (B4), $\Gamma = 1 + O(\epsilon^2)$, so $|\Gamma^{-1} - 1| = O(\epsilon^2)$. Because $p_m = |\alpha_m|/S$, the middle term in Eq. (B6)

$$\left| \tfrac{\alpha_m}{p_m \Gamma} - \alpha_m \right| \leq \left| \tfrac{S}{\Gamma} \right| + |\alpha_m| \leq (c+1)\epsilon + O(\epsilon^2), \tag{B7}$$

where we used $|\alpha_m| \leq \epsilon$ and (B3). The remaining $B$-components satisfy $\left\| \sum_{k\neq m} \alpha_k |k\rangle \right\| = \sqrt{\epsilon^2 - \alpha_m^2} \leq \epsilon$. Combining the three parts gives

$$\||\psi'_m\rangle - |\psi\rangle\| \leq (c+2)\epsilon + O(\epsilon^2) \quad \forall m \in B, \tag{B8}$$

hence $a \leq (c+2)\epsilon + O(\epsilon^2)$.

The unnormalized average equals the target state: $\sum_m p_m |\psi_m\rangle = |\psi\rangle$. Since $\Gamma$ is the same for every $m$

$$\left\| \sum_m p_m |\psi'_m\rangle - |\psi\rangle \right\| = \left\| (\Gamma^{-1} - 1) |\psi\rangle \right\| \leq \left\| \left( \frac{c^2}{2} \epsilon^2 + O(\epsilon^4) \right) |\psi\rangle \right\| = O(\epsilon^2), \tag{B9}$$

so $b = O(\epsilon^2)$.

Applying the state-level mixing lemma yields $\|\rho_{\text{mix}} - \rho\|_1 \leq a^2 + 2b = O(\epsilon^2)$. □

This Theorem implies that we indeed achieves a quadratic speed up using the randomized scheme upon the assumption $l_1$-smallness condition is satisfied. In short, this assumption is justified if the coefficients with index in $B$ satisfies exponential or greater than 1 power-law decay.

We first verify that the condition does not hold in general. Let $K := |B|$ and $\boldsymbol{\alpha} = (\alpha_j)_{j\in B}$. By Cauchy–Schwarz,

$$S = \|\boldsymbol{\alpha}\|_1 \;\le\; \sqrt{K}\,\|\boldsymbol{\alpha}\|_2 \;=\; \sqrt{K}\,\epsilon. \tag{B10}$$

Thus a *dimension-free* constant $c$ in (B3) cannot be guaranteed from $\|\boldsymbol{\alpha}\|_2 = \epsilon$ alone unless the tail on $B$ is $\mathcal{O}(1)$ sparse.

Below we explain when (B3) is satisfied under decay assumption of the tail coefficients, which are common in practices. Let $|\alpha|_{(1)} \ge |\alpha|_{(2)} \ge \cdots$ denote the decreasing rearrangement of $\{|\alpha_j|\}_{j\in B}$, and write

$$S = \sum_{j\in B} |\alpha_j| = \sum_{k\ge 1} |\alpha|_{(k)}, \qquad \epsilon^2 = \sum_{j\in B} \alpha_j^2 = \sum_{k\ge 1} |\alpha|_{(k)}^2. \tag{B11}$$

We now connect the explicit decay models to the $\ell_1$-smallness $S \le c\,\epsilon$. Begining by Geometric (exponential) decay. Assume

$$|\alpha|_{(k)} \le C\, r^{\,k-1}, \qquad 0 < r < 1. \tag{B12}$$

Then

$$\frac{S}{\epsilon} \;\le\; \frac{\sum_{k\ge1} C r^{k-1}}{\left(\sum_{k\ge1} C^2 r^{2(k-1)}\right)^{1/2}} = \frac{1-r^K}{1-r}\sqrt{\frac{1-r^2}{1-r^{2K}}} \le \sqrt{\frac{1+r}{1-r}} =: c(r), \tag{B13}$$

which is finite and independent of $|B|$.

If we have a less strong power-law decay for the coefficients, assumption (B3) still holds if the decay power $r$ is greater than 1. Assume

$$|\alpha|_{(k)} \le C\, k^{-r}, \qquad r > \tfrac{1}{2}. \tag{B14}$$

Then, in the infinite case, where $|B| \to \infty$,

$$\frac{S}{\epsilon} \;\sim\; \frac{\sum_{k\ge1} k^{-r}}{\left(\sum_{k\ge1} k^{-2r}\right)^{1/2}} \;=\; \begin{cases} \dfrac{\zeta(r)}{\sqrt{\zeta(2r)}}, & r > 1, \\ \text{diverges like } \log|B|, & r = 1, \\ \text{diverges like } |B|^{\,1-r}, & \tfrac{1}{2} < r < 1, \end{cases} \tag{B15}$$

where $\zeta(x) = \sum_{n=1}^{\infty} \frac{1}{n^x}$ is the Riemann zeta function.

## Appendix C: Configuration Interaction and Implementation

Configuration interaction (CI) is a many-body method used to capture electron correlation beyond the mean-field Hartree–Fock (HF) approximation. In CI, the exact wavefunction $|\Psi\rangle$ of a molecule is expressed as a linear combination of Slater determinants:

$$|\Psi_{\mathrm{CI}}\rangle = \sum_I c_I\,|\Phi_I\rangle\,, \tag{C1}$$

where $|\Phi_I\rangle$ represents a determinant formed by occupying a subset of spin orbitals.

Each Slater determinant $|\Phi_I\rangle$ corresponds to a configuration of electrons distributed over spin orbitals. In the full configuration interaction (FCI) method, all possible electron configurations within the basis set are included, yielding an exact solution within that basis. However, the number of such determinants scales exponentially with system size, making FCI feasible only for small molecules.

For illustration, we simulate LiH with 4 electrons in 12 spin orbitals (from 6 spatial orbitals using the STO-3G basis). The HF reference state $|\Phi_0\rangle$ occupies the lowest-energy molecular orbitals. All singly and multiply excited determinants $|\Phi_I\rangle$ are constructed by promoting electrons into virtual orbitals.

The molecular Hamiltonian is expressed in second quantization as

$$\hat{H} = \sum_{p,q} h_{pq} a_p^\dagger a_q + \frac{1}{2} \sum_{p,q,r,s} \langle pq|rs\rangle a_p^\dagger a_q^\dagger a_s a_r, \tag{C2}$$

where $h_{pq}$ and $\langle pq|rs\rangle$ are one- and two-electron integrals.

Matrix elements $H_{IJ} = \langle \Phi_I | \hat{H} | \Phi_J \rangle$ define the CI Hamiltonian. Diagonalizing it yields coefficients $c_I$ and the energy eigenvalues.

To prepare $|\Psi_{\mathrm{CI}}\rangle$ on a quantum computer, we encode each determinant $|\Phi_I\rangle$ as a computational basis state $|x_I\rangle$. The full wavefunction is then

$$|\Psi_{CI}\rangle = \sum_{I=1}^{\dim} c_I |x_I\rangle . \tag{C3}$$

We define a threshold $t$ and split coefficients:

$$A = I : |c_I| \geq t, \qquad B = I : |c_I| < t. \tag{C4}$$

Partition $B = \bigcup_m B_m$ and define

$$|\psi_A\rangle = \sum_{I \in A} c_I |x_I\rangle , \qquad |\psi_m\rangle = \sum_{I \in B_m} c_I |x_I\rangle . \tag{C5}$$

Prepare the normalized ensemble

$$\left|\tilde{\psi}_m\right\rangle = \Gamma_m \left( |\psi_A\rangle + \frac{1}{p_m} |\psi_m\rangle \right) \tag{C6}$$

and form the mixture:

$$\rho = \sum_m p_m \left|\tilde{\psi}_m\right\rangle \left\langle \tilde{\psi}_m \right| . \tag{C7}$$

This randomized method preserves fidelity while reducing state preparation complexity.

For the coefficient hierarchy of the CI coefficients, we can show the higher-order Slater determinants have suppressed coefficients via Rayleigh–Schrödinger perturbation theory. Suppose the molecular Hamiltonian can be written as

$$\hat{H} = \hat{H}_0 + \hat{V},$$

where $\hat{H}_0$ is is the Fock operator, whose ground-state eigenfunction $|\Phi_0\rangle$ is the HF determinant, and whose excited eigenfunctions $|\Phi_I\rangle$ correspond to orbital excitations. $\hat{V}$ is the difference between the full molecular Hamiltonian and $\hat{H}_0$. Practically, $\hat{V}$ typically includes the parts of the electron-electron repulsion that go beyond the mean-field approximation in $\hat{H}_0$.

Denote by $E_0^{(0)}$ the unperturbed ground-state energy of $\hat{H}_0$, and let $E_I^{(0)}$ be the energy of an excited configuration $|\Phi_I\rangle$ under $\hat{H}_0$. The exact eigenstate $|\Psi_{\mathrm{CI}}\rangle$ can be expanded as:

$$|\Psi_{\mathrm{CI}}\rangle = c_0 |\Phi_0\rangle + \sum_{I \neq 0} c_I |\Phi_I\rangle .$$

In first-order perturbation theory, one has

$$c_I^{(1)} = -\frac{\langle \Phi_I | \hat{V} | \Phi_0 \rangle}{E_I^{(0)} - E_0^{(0)}} = -\frac{V_{I0}}{\Delta E_{I0}},$$

where $\Delta E_{I0} = E_I^{(0)} - E_0^{(0)}$ is the unperturbed energy gap. Furthermore, if $\|\hat{V}\|$ is the operator norm of the perturbation, then $|V_{I0}| = \left|\langle \Phi_I | \hat{V} | \Phi_0 \rangle\right| \leq \|\hat{V}\|$. Hence, any coefficient $c_I^{(1)}$ for an excited determinant satisfies

$$|c_I^{(1)}| \leq \frac{\|\hat{V}\|}{\Delta E_{I0}}.$$

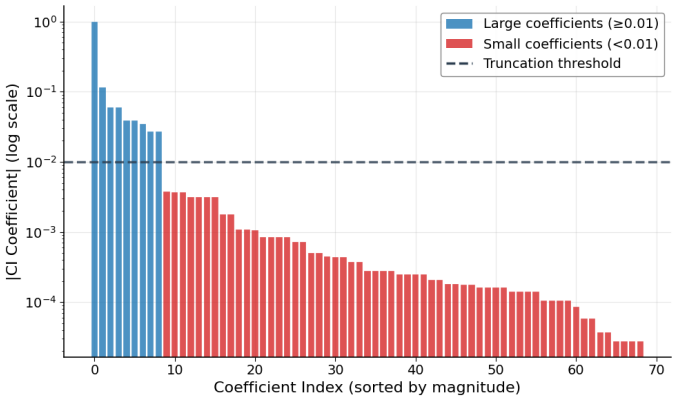


FIG. 2. Coefficient distribution of the CI state for LiH molecule.

Even in higher-order perturbation theory or in full CI solutions, large denominator gaps $\Delta E_{I0}$ strongly suppress amplitudes for high-lying excited determinants. We can see through Fig. 2 that the majority of coefficients have small magnitudes, while there are in total 69 coefficients in the FCI state of LiH molecule, only 2 of them are greater than 0.1 and only 9 of them are greater than 0.01.

In what follows, we let $n$ be the number of spin orbitals, and we index Slater determinants by $I \subset \{1, \ldots, n\}$ to indicate which orbitals are occupied. We denote $c_I$ as the CI coefficient of determinant $|\Phi_I\rangle$, and $\gamma = \sum_{I \in B} |c_I|^2$. we have

$$\gamma \leq C \left( \frac{\|\hat{V}\|}{\Delta_{\min}} \right)^2, \tag{C8}$$

where $\Delta_{\min} = \min_{I \neq 0} \left( E_I^{(0)} - E_0^{(0)} \right) > 0$ and $C$ absorbs constants from second-order (or higher) perturbation expansions. The factor $C$ encapsulates the higher-order terms of Rayleigh–Schrödinger perturbation expansions. More precisely second-order (and higher) corrections introduce additional sums over intermediate excitations, each contributing denominators of the form $\Delta E_{I0} \pm \Delta E_{J0}$ in the denominator and factors like $\langle \Phi_I | \hat{V} | \Phi_J \rangle$ in the numerator. Bounding these systematically leads to combinatorial and system-dependent prefactors, which are collectively denoted by $C$.