Exploting Sparsity in Quantum Imaginary Time Evolution

Lita, Victor[1] and Stair, Nicholas H.[1, a)]

Department of Physics, California Polytechnic State University, San Luis Obispo, CA, 93410

This work explores methods for exploiting sparsity in the quantum imaginary time evolution (QITE) algorithm.

I. INTRODUCTION

In modern quantum chemistry, solving the Schrödinger equation efficiently for molecules with strongly correlated electrons remains an important open challenge[?][?]. Here, "strong correlation" refers to scenarios where the energy required to promote electrons to higher orbitals rivals the Coulomb repulsion between them, resulting in electronic states that must be described by an extensive mixture of Slater determinants[?][?]. This complexity is central to phenomena such as bond dissociation and photochemical reactions[?][?], as well as to the behavior of systems exhibiting molecular magnetism[?] and high-temperature superconductivity[?]. Conventional single-reference Configuration Interaction (CI) or Coupled Cluster (CC) methods, which generally build excitations out of the Hartree–Fock (HF) determinant, often fail to capture these effects with sufficient accuracy. Additionally, if trying to capture excited state properties, popular classical approaches like time-dependent density functional theory (TD-DFT)[?][?] and equation-of-motion coupled cluster (EOM-CC)[?][?] often falter in the presence of near-degenerate states, such as those occurring at avoided crossings and conical intersections[?][?][?][?][?]. Although methods which can accurately treat strong correlation do exist—namely, the complete active space (CAS) family, density matrix renormalization group (DMRG), and selected configuration interaction (CI)—their memory requirements, especially when approaching full configuration interaction (FCI) limit, scale prohibitivly with active space size, rendering them impractical except under very specific conditions[?][?][?][?][?][?][?][?].

Quantum computational algorithms offer a compelling solution to at least the storage issue associated with the exponential number of FCI determinants by placing the many-particle wave function into a register of qubits.[?][?] Harnessing quantum-mechanical resources in this way holds the potential for more efficient description of strongly correlated electronic states in particular, as they are generally less amenable to classical compression or approximate representation. Significant progress in developing quantum-classical hybrid methods for ground- and excited-state electronic structure has emerged in recent years, either as stand alone algorithms or as state-preparation metnods for quantum phase estimation (QPE). Variational quantum eigensolver (VQE) approaches,[?][?][?][?] which optimize the parameters of an ansatz circuit on classical hardware, have shown promise on today's noisy devices, but face complications associated with high-dimensional, nonlinear optimizations that can easily become trapped in local minima.

By contrast, imaginary time evolution (ITE) has a long history in classical simulations of correlated fermionic systems,[?] and its quantum analogue (QITE) has been introduced[?] to obviate many of the classical optimization challenges. Recent advances in QITE have focused on reducing resource demands and accelerating convergence for both ground and excited state preparation, via QITE's Lanczos and/or Folded Spectrum (FS) formulations. A stochastic variational QITE approaches significantly reduces measurement overhead on large-scale devices[?], while step-merging techniques have been successfully applied to combinatorial optimization problems such as MaxCut[?]. Composite formulations that extend qDRIFT-based methods into the imaginary-time regime offer rigorous error bounds and improved circuit efficiency[?][?]. QITE has also been applied to diverse domains, including nuclear density functional theory[?] and molecular systems, where improved derivations of the QITE equations in second quantization have enable larger time steps and reduced circuit depth[?]. In addition, alternative strategies based on drifting real-time evolution lower gate and measurement complexity[?]. Finally, adaptive variational QITE techniques and reinforcement learning-assisted QITE variants bridge QITE and variational approaches, further enhancing ground state preparation on near-term quantum devices[?][?].

With only the notable exception of **??**, virtually all current QITE implementations rely on a *fixed* pool of anti-Hermitian operators $\mathcal{A}$ to approximate the imaginary time propagation at each time step.[?][?][?][?] The dimension of $\mathcal{A}$ directly dictates both the quantum and classical costs, and practical simulations often restrict the operator pool to include only one- and two-body excitations. Such a static pool can be suboptimal in strongly correlated systems, where higher-than-two-body excitations may be required to accurately reproduce the evolving state. Furthermore, as the system is propagated toward the ground (or an excited) eigenstate, the QITE solution vector $\mathbf{x}$ naturally becomes sparser in the later stages of imaginary time. In principle, fewer operators would then be needed to describe the update, suggesting that a dynamic pool selection scheme could yield substantial savings in both accuracy and computational resources. Introducing a selection protocol at each time step is therefore a very sensible extension to QITE in order to accurately capture strong correlation early in the imaginary time evolution and to exploit the sparsity that emerges later in the process.

Here, we present an improved QITE algorithm that *iteratively selects* the most significant excitation and de-excitation operators at each time step. Our approach identifies important contributions to the anti-Hermitian generator in an efficient manner, requiring significantly fewer quantum resources than those demanded by the conventional full update of $\mathbf{x}$ at every iteration. Improve the above, discuss general idea of selec-

a)Electronic mail: nstair@calpoly.edu

toin based on satisfaction of the residual condition (similar to PQE) Notably, by adaptively growing (or shrinking) the operator pool, the method balances the requirement to capture high-order effects in the early imaginary time regime and the opportunity to eliminate negligible operators once the system is close to an eigenstate. This dynamical selection of operators is especially pertinent for challenging excited states where multireference effects come into play and for near-degenerate situations in which physically relevant interactions of higher order can become indispensable. Thus, the proposed selection strategy adds a powerful layer of flexibility to QITE, paving the way for more accurate and resource-efficient computations of ground and excited states on near-term and early fault-tolerant quantum hardware.

   a.   *Significance of Our Results (Template).*

- **Overview of Key Contributions:** Summarize the main algorithmic or theoretical developments introduced in this work, highlighting how they address current limitations in QITE for strongly correlated systems.

- **Performance Advantages:** Emphasize improvements in accuracy, computational cost, or scalability relative to existing QITE approaches and other quantum/classical methods.

- **Applications and Broader Impact:** Discuss the relevance of these results to realistic molecular or materials systems, including potential extensions to larger, more complex problems.

- **Future Directions:** Suggest how the proposed strategies might be generalized or combined with other quantum algorithms, as well as possible avenues for error mitigation or fault-tolerance.

This is a citation:[1]

II. THEORY

A. Quantum imaginary time evolution in second quantization

The quantum imaginary time evolution algorithm is predicated on the principle that the ground state of a Hamiltonian $\hat{H}$ can be found by evolving a trial state $|\Phi_o\rangle$ with the imaginary time evolution operator in the infinite time-step limit

$$|\Psi_0\rangle = \lim_{\beta \to \infty} \frac{1}{\sqrt{c(\beta)}} e^{-\beta \hat{H}} |\Phi_0\rangle, \qquad (1)$$

so long as $\langle \Psi_0 | \Phi_0 \rangle \neq 0$. The factor of $1/\sqrt{c(\beta)} = 1/\sqrt{\langle \Phi_0 | e^{-2\beta \hat{H}} | \Phi_0 \rangle}$ normalizes the evolved state, and it is required that $\langle \Psi_o | \Phi_o \rangle \neq 0$.

In this work we will consider $\hat{H}$ as the electronic structure hamiltonian

$$\hat{H} = \sum_{pq} h_{pq} \hat{a}_p^\dagger \hat{a}_q + \frac{1}{4} \sum_{pqrs} v_{pqrs} \hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_s \hat{a}_r, \qquad (2)$$

where $\hat{a}_p^\dagger$ ($\hat{a}_q$) is a fermionic annihilation (creation) operator, h_{pq} are one-electron integrals and v_{pqrs} are antisymmetrized two-electron integrals[?] .

The imaginary time evolution operator is non-unitary, making it impractical for implementation on quantum computers. However, one may approximate the action of the imaginary time evolution operator with time step $\Delta\beta$ using a unitary operation of the form

$$c(\Delta\beta)^{-1/2} e^{-\Delta\beta \hat{H}} |\Phi\rangle \approx e^{-i\Delta\beta \hat{A}} |\Phi\rangle, \qquad (3)$$

where $\hat{A}$ is Hermitian. The operator $\hat{A}$, of central importance to this work, can generally be written as a linear expansion of simple hermitian operators $\hat{\rho}_\mu$ contained in the set $\mathcal{P}$ as

$$\hat{A} = \sum_{\hat{\rho}_\mu \in \mathcal{P}} \alpha_\mu \hat{\rho}_\mu, \qquad (4)$$

where the expansion coefficients α_μ comprising the vector $\boldsymbol{\alpha}$ are real, and the set has dimension M. The operators $\hat{\rho}_\mu \in \mathcal{P}$ greatly influance the both the computational cost and effectiveness of the QITE algorithm, and thus their selection is of principal importance. In the original implementation outlined by Motta *et. al.*[2], $\mathcal{P}$ is chosed as the set Q of all possible $4^{N_{qb}}$ Pauli operator products $\hat{\rho}_\mu = \prod_l \hat{\sigma}_{\mu_l}^{(l)}$ such that $\mu \equiv (\mu_1, \mu_2, .., \mu_{N_{qb}})$ is a multi-index describing a unique Pauli operator product, and $\mu_l \in \{I, X, Y, Z\}$.

In this work, during evolution under $\hat{A}$ we care to to preserve particle-number and spin-projection symmetry. Optimally, would like to also maintain the feature that the wave function remains real. As such, we write $\hat{A} = i(\hat{T} - \hat{T}^\dagger)$ as as en expansion of unitary coupled-cluster like terms (multiplied by the imaginary number i) such that each individual operator in the expansion can take the form

$$\hat{\rho}_\mu = -i(\hat{\tau}_\mu - \hat{\tau}_\mu^\dagger). \qquad (5)$$

The general excitation operator $\hat{\tau}_\mu$ can subsequently be expressed as

$$\hat{\tau}_\mu \equiv \hat{\tau}_{rs\cdots}^{pq\cdots} = \hat{a}_p^\dagger \hat{a}_q^\dagger \cdots \hat{a}_s \hat{a}_r,$$

where, $\hat{\tau}_{rs\cdots}^{pq\cdots}$ denotes an operator that annihilates particles in the orbitals indexed by $(r, s, \ldots)$ and creates particles in the orbitals indexed by $(p, q, \ldots)$. In this case multi-index μ is redefined as $\mu \equiv ((p, q, \ldots), (r, s, \ldots))$, representing the unique set of excitations from spin orbitals $(\phi_p \phi_q \cdots)$ to spin orbitals $(\phi_r \phi_s \cdots)$. In this second-quantized representation, the operators $\hat{\rho}_\mu$ are then given by small *linear combinations* of Pauli operator products given by an appropriate operator mapping, where all sub terms of Pauli operator products commute with one another such that the exponentiation $e^{\alpha_\mu \hat{\rho}_\mu}$ can be implemented via Trotterization without any error.

NOTE: Need to verify if in Tsuchimochi's derivation there simply is no first order error, or if its simply independent of α

As mentioned in the introduction, Tsuchimochi *et. al.*[3] have recently shown that the QITE error function for a vector of expansion coefficients $\boldsymbol{\alpha}$ and a time step $\Delta\beta$ is given to second

order in $\Delta\beta$ by

$$
\begin{aligned}
\mathcal{F}(\boldsymbol{\alpha}, \Delta\beta) &= \| c(\Delta\beta)^{-1/2} e^{-\Delta\beta\hat{H}} |\Phi\rangle - e^{-i\Delta\beta\hat{A}} |\Phi\rangle \|^2 \\
&= 2 - 2c(\Delta\beta)^{-1/2} \Re \langle\Phi| e^{-\Delta\beta\hat{H}} e^{-i\Delta\beta\hat{A}} |\Phi\rangle \quad (6) \\
&\approx \mathcal{F}_o + \Delta\beta^2 \left(\langle\Phi| \hat{A}^2 |\Phi\rangle - i \langle\Phi| [\hat{H}, \hat{A}] |\Phi\rangle \right)
\end{aligned}
$$

Importantly, the term $\mathcal{F}_o$ does not have any dependence on the expansion coefficients $\boldsymbol{\alpha}$, and the second order term does not have dependence on the normaliztion ocefficeiont $c(\Delta\beta)$, so the objective of this second-order QITE becomes minimization of

$$
\mathcal{F}_2(\boldsymbol{\alpha}) = \sum_{\hat{\rho}_\mu, \hat{\rho}_\nu \in \mathcal{P}} \alpha_\mu \alpha_\nu \langle\Phi| \hat{\rho}_\mu \hat{\rho}_\nu |\Phi\rangle - \sum_{\hat{\rho}_\mu \in \mathcal{P}} \alpha_\mu i \langle\Phi| [\hat{H}, \hat{\rho}_\mu] |\Phi\rangle .
\tag{7}
$$

Eq. (7) gives rise to the M dimensional linear system

$$
\mathbf{S}\boldsymbol{\alpha} + \mathbf{b} = \mathbf{0}.
\tag{8}
$$

The elements of $\mathbf{S}$, and $\mathbf{b}$, can then be determined via measurement of the quantum register such that

$$
S_{\mu\nu} = 2\Re \left(\langle\Phi| \hat{\rho}_\mu \hat{\rho}_\nu |\Phi\rangle \right)
\tag{9}
$$

and,

$$
b_\mu = \Im \left(\langle\Phi| [\hat{H}, \hat{\rho}_\mu] |\Phi\rangle \right)
\tag{10}
$$

As pointed out by 3, b_μ is the derivative of the energy with respect to the parameter $\Delta\beta\alpha_\mu$ evaluated at $\boldsymbol{\alpha} = \mathbf{0}$

$$
b_\mu = \left. \frac{\partial E}{\partial(\Delta\beta\alpha_\mu)} \right|_{\boldsymbol{\alpha}=\mathbf{0}},
\tag{11}
$$

which helps elucidate the significance the linear system. This interpretation suggests that the Quantum Imaginary Time Evolution (QITE) algorithm can be seen as a variant of the natural gradient descent method. In this context, noting the condition that $\langle\Phi|\hat{\rho}_\mu|\Phi\rangle = 0$ for real states, the matrix $\mathbf{S}$ is effectively the Fubini-Study metric tensor[4,5] and convergence of the algorithm is achieved when the gradient vector $\mathbf{b}$ reaches zero.

Once the solution vector $\boldsymbol{\alpha}$ is found by solving Eq. (8), the QITE unitary for a time step $\Delta\beta$ is constructed via Trotterization of $\hat{A}$ as

$$
\hat{U}(\Delta\beta) = \prod_{\hat{\rho}_\mu \in \mathcal{P}} e^{-i\alpha_\mu \Delta\beta \hat{\rho}_\mu} = e^{-i\Delta\beta\hat{A}} + O(\Delta\beta)
\tag{12}
$$

The linear system can then be solved for n iterations until a desired total amount of evolution time $\beta = n\Delta\beta$ is achieved. NOTE: should comment on the order of $\Delta\beta$ error for Trotterization vs from the QITE equations.

B. Folded-Spectrum QITE for Excited States

Maybe do this section after adaptive QITE, and write it from that perspective (FS-AQITE) A straightforward strategy to target higher-lying eigenstates in QITE is to employ the *folded-spectrum* approach, which has been explored in both

classical imaginary time evolution and quantum algorithms.? Rather than evolving under $\hat{H}$, one defines a shifted operator

$$
\hat{H}_\omega = (\hat{H} - \omega\hat{I})^2,
\tag{13}
$$

where ω is chosen near the energy of the desired eigenstate. The imaginary time evolution of an initial trial state $|\Phi_0\rangle$ under $e^{-\beta\hat{H}_\omega}$ projects out the eigenstate of $\hat{H}_\omega$ (and thus of $\hat{H}$) whose energy is closest to ω. In analogy with ground-state QITE, the procedure for folded-spectrum QITE (FS-QITE) replaces all occurrences of $\hat{H}$ in the linear system [Eq. (8)] with $(\hat{H} - \omega\hat{I})^2$, thereby requiring measurements of the matrix elements

$$
S_{\mu\nu} = 2\Re \left(\langle\Phi| \hat{\rho}_\mu \hat{\rho}_\nu |\Phi\rangle \right)
\tag{14}
$$

(identical to the standard approach) and,

$$
b_\mu = \Im \left(\langle\Phi| \left[(\hat{H} - \omega\hat{I})^2, \hat{\rho}_\mu \right] |\Phi\rangle \right).
\tag{15}
$$

By carrying out imaginary time steps $\Delta\beta$ with this modified Hamiltonian $\hat{H}_\omega$, one arrives at an approximate excited state $|\Psi_n\rangle$ satisfying

$$
|\Psi_n\rangle \approx \lim_{\beta\to\infty} \frac{e^{-\beta(\hat{H}-\omega\hat{I})^2} |\Phi_0\rangle}{\sqrt{\langle\Phi_0| e^{-2\beta(\hat{H}-\omega\hat{I})^2} |\Phi_0\rangle}},
\tag{16}
$$

provided that $\langle\Psi_n|\Phi_0\rangle \neq 0$. Thus, FS-QITE effectively "folds" the spectrum around ω, isolating the targeted eigenstate even in the presence of lower-lying states. Importantly, the same considerations regarding operator selection (e.g., second-quantized excitations and de-excitations) carry over to FS-QITE, with the only modification being the requisite measurements involving $(\hat{H} - \omega\hat{I})^2$. Hence, all of the symmetry-preserving and real-valued operator constraints discussed above remain valid in the folded-spectrum formalism.

C. Quantum Lanczos for Excited States

Maybe do this section after adaptive QITE, and write it from that perspective (AQLanczos + FS-AQLanczos) Although the primary objective of QITE is to approximate the ground state, the intermediate wave functions generated at each step can be used to form a Krylov subspace for excited-state calculations. In the *Quantum Lanczos* (QLanczos) method,[2] one exploits these QITE-produced states rather than the exact imaginary time propagations. Specifically, define the sequence

$$
|\Phi_k\rangle = \left(\prod_{k=1}^{K} \hat{U}(\Delta\beta)^{(k)} \right) |\Phi_0\rangle, \quad k = 0, 1, \dots, K,
\tag{17}
$$

with $|\Phi_0\rangle$ as the initial reference (e.g., Hartree–Fock). In principle, for small $\Delta\beta$, this unitary product approximates $e^{-k\Delta\beta\hat{H}}$, thus preserving components of higher-lying eigenstates across a moderate range of k.

To perform QLanczos, one measures the overlaps and Hamiltonian matrix elements among $\{|\Phi_k\rangle\}$:

$$
X_{k,k'} = \langle\Phi_k | \Phi_{k'}\rangle,
\tag{18}
$$

$$
H_{k,k'} = \langle\Phi_k| \hat{H} |\Phi_{k'}\rangle,
\tag{19}
$$

where $k, k' \in \{0, 1, \ldots, K\}$. A characteristic simplification of imaginary time evolution is that for k, k' differing by an even integer, one can relate $X_{k,k'}$ and $H_{k,k'}$ to those at intermediate steps, thus reducing the measurement overhead. Because $|\Phi_k\rangle$ is obtained from a sequence of short-time unitaries, QLanczos inherits much of the efficiency of classical Lanczos, requiring far fewer than the full set of pairwise overlaps among $\{|\Phi_k\rangle\}$.

Having constructed the matrices $\mathbf{X}$ and $\mathbf{H}$ from Eqs. (18) and (19), one solves the generalized eigenvalue problem

$$\mathbf{H} \mathbf{c} = E \mathbf{X} \mathbf{c}. \tag{20}$$

The resulting eigenvalues E and eigenvectors $\mathbf{c}$ provide approximations to the eigenenergies and eigenstates of $\hat{H}$:

$$|\Psi_n\rangle \approx \sum_{k=0}^{K} c_k^{(n)} |\Phi_k\rangle. \tag{21}$$

The lowest root E_0 often refines the direct QITE ground-state energy, while higher roots E_n yield approximations to excited states. Crucially, QLanczos reuses states (and associated measurement data) generated by the QITE procedure, thereby offering a cost-effective subspace method for post-processing excited states. In the limit of very small $\Delta\beta$, the set $\{|\Phi_k\rangle\}$ approximates the classical Krylov space spanned by powers of $\hat{H}$, delivering rapid convergence and robust excited-state estimates with relatively few QITE steps.

D. Sparsity and Residual Selected QITE

There is a great deal of flexibility in how one chooses the Pauli product operators that make up the set $\mathcal{P}$. This choice is also important because if greatly impacts the quantum (and classical) resources required for QITE and ultimately determines its usefulness as a hybrid quantum-classical algorithm. For example, setting $\mathcal{P} = Q$ would result in the need to repeatedly solve linear systems of dimension $4^{N_{qb}}$, which scales even more poorly that the dimension of the entire Fock space. In previous work there are two distinct ways to dramatically reduce the dimension of $\mathcal{P}$ without significant loss of accuracy.

The first strategy, which is described in detail in the original QITE article [Ref. 2] is to consider a set of operators $\mathcal{P}_\ell$ specific to each of the N_ℓ terms of the Hamiltonian ($\hat{H}_\ell$), rather than a single set for the entire Hamiltonian. They show that this strategy works quite well for geometric k–local Hamiltonians for which each term $\hat{H}_\ell$ acts only on the nearest k (or fewer) qubits. Specifically, they choose some domain $D > k$ qubits such that each $\mathcal{P}_\ell$ contains 4^D operators such that the total number of QITE parameters is approximated by $nN_\ell 4^D$.

The second strategy, which we also employ in this study (as discussed above), is to limit the operators $\hat{\rho}_\mu$ to only those that preserve particle-number and spin symmetry. Prior work has then considered fixed sets $\mathcal{P}_{sq}$ of second quantized excitation and de-excitation operators. Specific examples include the set of generalized singles and doubles excitation/de-excitation pairs $\mathcal{P}_{sq}^{GSD}$ (as similar operators comprise the hamiltonian), and only particle-hole single and double excitation/de-excitation pairs $\mathcal{P}_{sq}^{SD}$ for a given Hartree-Fock reference determinant.

A key feature of this work is to introduce a selection strategy for constructing an *optimal* set, $\widetilde{\mathcal{P}}_{sq}$, of excitation/de-excitation operator pairs based on the residuals of the projected Schrödinger equation. Although there are several sensible selection criteria (e.g., $|b_\mu|$ or perturbative energy estimates similar to those used in selected configuration interaction[?]), here we employ a protocol closely following the work of Stair *et al.*, which is particularly economical in terms of measurement cost. The rationale is grounded in the so-called *residual condition*: for any unitary $\hat{U}_n$ that transforms a single-determinant reference $|\Phi_0\rangle$ (e.g., Hartree–Fock) into the exact eigenstate $|\Psi_n\rangle$, the residuals

$$r_\mu = \langle \Phi_\mu | \hat{U}_n^\dagger f(\hat{H}) \hat{U}_n | \Phi_0 \rangle \tag{22}$$

must vanish for all determinants $|\Phi_\mu\rangle$ once $|\Psi_n\rangle$ is reached. Here $f(\hat{H})$ is any function of the Hamiltonian sharing the same eigenvectors, and $|\Phi_\mu\rangle$ is the determinant obtained by applying the second-quantized operator $\hat{\rho}_\mu$ to $|\Phi_0\rangle$. Careful with k_μ vs ρ_μ here In magnitude, $|r_\mu|$ indicates how much the determinant $|\Phi_\mu\rangle$ is under- or overrepresented in the current approximation to $|\Psi_n\rangle$. Thus, including $\hat{\rho}_\mu$ in the QITE unitary is sensible because it directly addresses the residual mismatch, enabling a more accurate approach to the target eigenstate.

Importantly, we note that it is not necessary to estimate the numerical values of r_μ by operator averaging (or shadow tomography) if $f(\hat{H})$ is chosen to be unitary, such as the time evolution operator $f\hat{H} = e^{-i\Delta t \hat{H}}$ (or a product-formula approximation). In this situation, one can construct a quantum state on the register

$$|\psi_n^R\rangle = \hat{U}_n^\dagger e^{-i\Delta t \hat{H}} \hat{U}_n |\Phi_0\rangle \tag{23}$$

with coefficients c_μ who's square moduli will also satisfy the above residual condition (i.e. $|c_\mu|^2 = 0$ and $|c_o|^2 = 1$ if $\hat{U}_n$ transforms the single-determinant reference $|\Phi_0\rangle$ into $|\Psi_n\rangle$). This implies that important operators $\hat{\rho}_\mu$ should be added to $\widetilde{\mathcal{P}}_{sq}$ if measuring each qubit (in the Z basis) of $|\psi_n^R\rangle$ gives a bitstring readout corresponding to particle-number representation of $|\phi_\mu\rangle$ with high frequency. The algorithm can be considered converged when, after a user-defined number measurements, only the bitstring corresponding to the reference determinant $|\Phi_o\rangle$ is measured.

In practice, our numerical evidence suggests that summing the magnitudes of these residuals over all QITE iterations, $R_\mu = \sum_k |r_\mu^{(k)}|$, provides a distribution that closely mirrors the accumulated magnitudes of the solution-vector elements, $A_\mu = \sum_k |\alpha_\mu^{(k)}|$, for the same set of candidate operators $\mathcal{P}_{sq}$. In other words, operators that consistently exhibit large residuals also acquire large QITE parameters across the imaginary time evolution. Theoretically, this similarity can be rationalized by noting that both r_μ and x_μ measure departures of the evolving wave function from the exact target state—the former in terms of incomplete representation in the determinant $|\Phi_\mu\rangle$, and the latter in terms of how strongly $\hat{\rho}_\mu$ must act to correct that imbalance.

Because we are also interested in low-lying excited states,

such that the total QITE wave function (starting from a reference state $|\Phi_0\rangle$) is

$$|\Psi_{\text{QITE}}\rangle = \prod_{k=1}^{K} \hat{U}(\Delta\beta)^{(k)} |\Phi_0\rangle = \prod_{k=1}^{K} \prod_{\mu=1}^{M^{(k)}} e^{-i\,\Delta\beta\,\alpha_\mu^{(k)}\,\hat{\rho}_\mu^{(k)}} |\Phi_0\rangle.$$

(26)

Here, $K = \beta/\Delta\beta$ is the total number of imaginary time steps, and each pool $\mathcal{P}^{(k)}$ has size $M^{(k)}$. The cost of performing QITE may be characterized along several axes:

    a.   Classical Parameters.

A central figure of merit is the total number of parameters $\{\alpha_\mu^{(k)}\}$ used in the QITE unitaries across all iterations. Since each QITE step may introduce a new set of $M^{(k)}$ operators, the *total* number of classical parameters is

$$N_{\text{QITE}} = \sum_{k=1}^{K} M^{(k)}.$$

(27)

Larger pools (and hence larger $M^{(k)}$) can improve the flexibility of the QITE ansatz but incur heavier computational and measurement burdens. Designing efficient, possibly adaptive, strategies to select $\mathcal{P}^{(k)}$ is therefore crucial to maintaining tractable classical costs.

    b.   Measurements of Observables.

Each QITE iteration requires measuring matrix elements to construct the linear system in Eq. (24). Concretely, one needs

$$S_{\mu\nu}^{(k)} = 2\,\Re\left[\langle\Phi^{(k)}|\,\hat{\rho}_\mu^{(k)}\,\hat{\rho}_\nu^{(k)}\,|\Phi^{(k)}\rangle\right],$$

(28)

and,

$$b_\mu^{(k)} = \Im\left[\langle\Phi^{(k)}|\,\left[\hat{H},\hat{\rho}_\mu^{(k)}\right]\,|\Phi^{(k)}\rangle\right],$$

(29)

where $|\Phi^{(k)}\rangle \equiv \hat{U}(\Delta\beta)^{(k-1)}\cdots\hat{U}(\Delta\beta)^{(1)}|\Phi_0\rangle$ denotes the current wave function (i.e., the state at the start of the kth iteration). The total number of unique *pairwise* products $\hat{\rho}_\mu^{(k)}\hat{\rho}_\nu^{(k)}$ may be on the order of $(M^{(k)})^2$ per iteration, and one must also measure various commutators $\left[\hat{H},\hat{\rho}_\mu^{(k)}\right]$. Hence, summing over all k, the number of distinct observables can become quite large for big pools.

While straightforward measurement protocols (e.g., measuring each observable individually) could be prohibitively expensive, more advanced methods such as *classical shadows*? can compress the measurement overhead. In essence, classical shadows allow one to efficiently estimate many Pauli expectation values simultaneously using randomized measurements and classical post-processing, reducing the exponential blowup in the number of measurement settings. Emphasize that reguardless, one wants to minimize the total number of Pauli operators that need to be measured

    c.   Circuit Depth and Gate Counts.

From a quantum-circuit perspective, each exponential $e^{-i\,\Delta\beta\,\alpha_\mu^{(k)}\,\hat{\rho}_\mu^{(k)}}$ is typically decomposed into a small network of multi-qubit gates (e.g., CNOTs) and parameterized single-qubit rotations. If $\hat{\rho}_\mu^{(k)}$ preserves fundamental symmetries (particle number, spin projection, etc.) and is built from

FIG. 1. Comparison of the accumulated residual magnitudes R_μ versus the accumulated QITE parameter magnitudes X_μ for all second-quantized operators in $\mathcal{P}_{\text{sq}}^{\text{all}}$. The strong correlation between these two distributions indicates that residual-based selection effectively identifies the same dominant operators as those that ultimately acquire large QITE amplitudes.

we partition the overall transformation into two parts,

$$\hat{U}_n = \hat{U}_n^{\text{QITE}}\,\hat{U}_n^{\text{CIS}},$$

where the first circuit $\hat{U}_n^{\text{CIS}}$ transforms $|\Phi_0\rangle$ into the nth configuration-interaction singles (CIS) state, providing a reasonable zeroth-order approximation to $|\Psi_n\rangle$. The second part, $\hat{U}_n^{\text{QITE}}$, is then adaptively constructed via the QITE procedure described in Section **??**, using the residual-based selection criterion to systematically introduce the most relevant higher-order excitations and de-excitations.

E. Computational resource requirements in QITE

As emphasized above, QITE is an iterative procedure whereby each iteration k (with $k = 1, 2, \ldots, K$) solves the linear system

$$\mathbf{S}^{(k)}\boldsymbol{\alpha}^{(k)} + \mathbf{b}^{(k)} = \mathbf{0}$$

(24)

to obtain the expansion coefficients $\alpha_\mu^{(k)}$ for the operators $\{\hat{\rho}_\mu^{(k)}\}$ in the pool $\mathcal{P}^{(k)}$. These coefficients then define a short-time unitary

$$\hat{U}(\Delta\beta)^{(k)} = \prod_{\mu=1}^{M^{(k)}} e^{-i\,\Delta\beta\,\alpha_\mu^{(k)}\,\hat{\rho}_\mu^{(k)}},$$

(25)

low-rank fermionic excitations (e.g., up to doubles), the Trotter block for $e^{-i\Delta\beta\,\alpha_\mu^{(k)}\,\hat{\rho}_\mu^{(k)}}$ can often be implemented with a bounded number of CNOTs that scales polynomially in the system size. Careful, not true if higher order excitations are included

In the *deepest* QITE circuit (where the largest operator pool is applied), we may have

$$(\# \text{ CNOTs}) \propto \sum_{\mu=1}^{M^{(k_\star)}} \text{cost}(\hat{\rho}_\mu^{(k_\star)}),$$

where $k_\star$ is the iteration featuring the largest pool. This cost may be further reduced if adjacent exponentials can be merged or commuted.[6] Additionally, each operator $e^{-i\Delta\beta\,\alpha_\mu^{(k)}\,\hat{\rho}_\mu^{(k)}}$ introduces at least one parameterized rotation gate, and in practice, multiple rotations may be needed to handle real and imaginary parts or to map the fermionic operator into an appropriate qubit representation. Hence, the total number of *parameterized* gates across all steps is also $O(N_{\text{QITE}})$, which serves as a proxy for the T-gate cost in a fault-tolerant setting.

    d.   Motivation to Reduce Costs.

In summary, the resource requirements of QITE scale with (i) the total number of classical parameters N_{QITE}, (ii) the number of unique observables measured, and (iii) the circuit depth (especially in terms of CNOTs and parameterized rotations). As larger systems with stronger correlations are considered, these costs can quickly become prohibitive. Consequently, much of this work is devoted to devising efficient strategies for *selecting* and *updating* the pool of operators $\{\hat{\rho}_\mu^{(k)}\}$, thereby minimizing N_{QITE} while maintaining sufficient accuracy to capture the essential physics. In parallel, improved measurement protocols (e.g., classical shadows) can mitigate overheads from the exponentially growing number of observables. By combining careful operator selection with more efficient measurement and circuit compilation, one aims to push QITE toward practical regimes even for challenging, strongly correlated electronic structure problems.

NOTE: Think about selection for excited states as well.

III. COMPUTATIONAL CONSIDERATIONS

IV. RESULTS

V. CONCLUSIONS

[1] J. B. Schriber and F. A. Evangelista, J. Chem. Phys. **144**, 161106 (2016).
[2] M. Motta, C. Sun, A. T. Tan, M. J. O'Rourke, E. Ye, A. J. Minnich, F. G. Brandão, and G. K.-L. Chan, Nat. Phys. **16**, 1 (2019).
[3] T. Tsuchimochi, Y. Ryo, S. L. Ten-No, and K. Sasasako, Journal of Chemical Theory and Computation **19**, 503 (2023).
[4] J. Stokes, J. Izaac, N. Killoran, and G. Carleo, Quantum **4**, 269 (2020).
[5] J. Gacon, C. Zoufal, G. Carleo, and S. Woerner, Quantum **5**, 567 (2021).
[6] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, Rev. Mod. Phys. **92**, 015003 (2020).