Cédrick Perron

April 7th 2026

arXiv:2301.07726

Hierarchical Clifford transformations to reduce entanglement in quantum chemistry wavefunctions

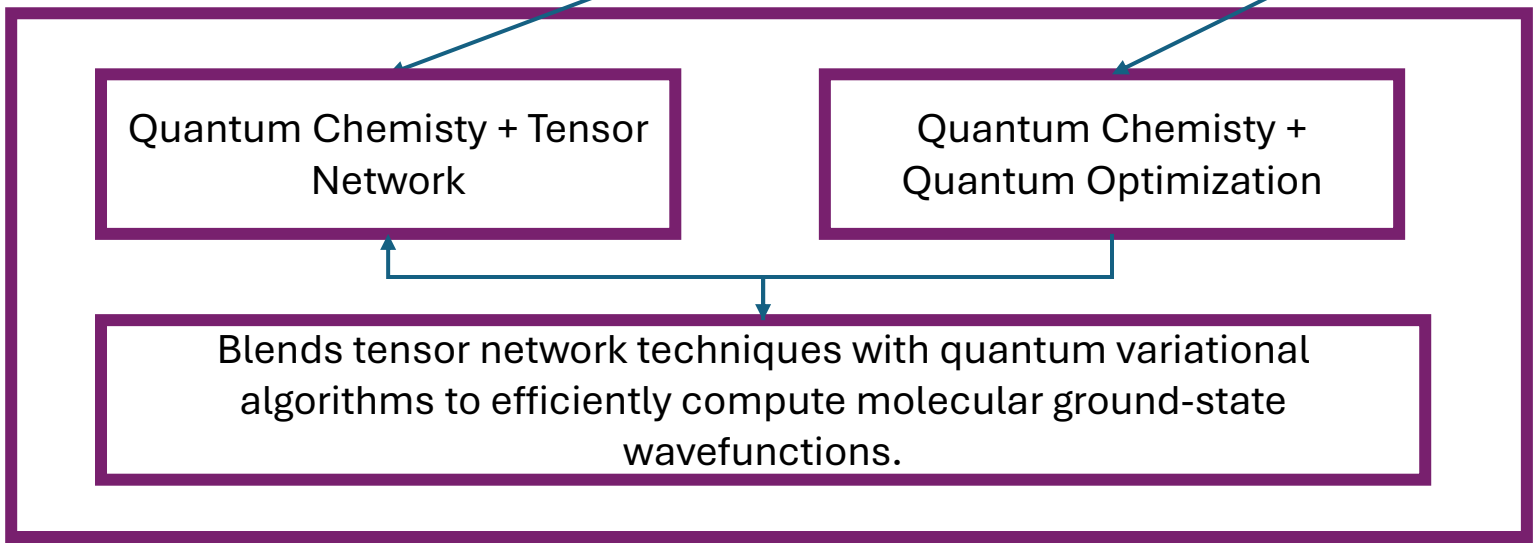
Ryan V. Mishmash,[1,*] Tanvi P. Gujarati,[1] Mario Motta,[1] Huanchen Zhai,[2] Garnet Kin-Lic Chan,[3] and Antonio Mezzacapo[3,1]

[1]*IBM Quantum, Almaden Research Center, San Jose, California 95120, USA*

[2]*Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, United States*

[3]*IBM Quantum, T. J. Watson Research Center, Yorktown Heights, NY 10598, USA*

(Dated: January 20, 2023)



Quantum Chemistry: Choose a parametrized wavefunction ansatz $U(\theta)$ and optimize it:

$$\mathrm{E}_0 \;=\; \arg\min_{\theta} \left\langle \Psi_0 | U^{\mathrm{dag}}(\theta) \mathrm{H} U(\theta) \right| \Psi_0 \Big\rangle$$

Bottleneck is choosing a good choice for $U(\theta)$:
→ Too expressive: Barren plateaus
→ Not expressive enough: can't reach ground state

In the paper, they develop a new way to construct $U(\theta)$:

$$U(\theta) \;=\; U_{clifford} V(\theta)$$

where $U_{clifford}$ is constructed by leveraging *approximate* Clifford symmetry (Qubit tapering). It is classically efficient to implement, capture a huge chunk of the physics upfront, reduces the effective problem size.
$V(\theta)$*:* A dynamically simpler parametrized unitary

So, in the end, you end up with something that is significantly easier to optimize over!

How to construct $U_{clifford}$:

From Qubit Tapering, start with a chemistry system with Hamiltonian given by (Pauli basis):

$$H \;=\; \sum_i c_i \, P_i$$

There exists a symmetry group: $S \;=\; \langle \tau_1, \tau_2, \ldots, \tau_{n\_sym} \rangle \;=\; \{I, \sigma^x, \sigma^y, \sigma^z\}^{\otimes n}$ that commutes with each term of the Hamiltonian: $[\tau, P_i] = 0 \; \forall i$.

This means that eigenvectors of *H*, like the ground state, is also an eigenvector of τ with eigenvalues ± 1. So you can organize the Hamiltonian in $2^{\mathrm{n\_sym}}$ blocks, where each block are labeled by a unique set of $\{\pm 1\}^{\otimes n\_\mathrm{sym}}$ eigenvalues!

The idea of qubit tapering is to assign the parity symmetry generated by τ_{j} to the state of qubit $q(j)$ using the Clifford transformation:

$$C_{tapering} \;=\; \frac{1}{\sqrt{2}} (\sigma^x_{q(j)} + \tau_j)$$

thus: $C^{dag}_{tapering} \tau_j C_{tapering} = \sigma^x_{q(j)}$ and thus the state of qubit $q(j)$ encodes the eigenvalue ± 1 of the symmetry operator. Qubit $q(j)$ is now fixed and can be removed from the simulation. The final ground state will thus be given by:

$$|\Psi_0\rangle \;=\; \left[\prod_{\mathrm{j=1}}^{\mathrm{n\_sym,}} |\pm\rangle_{\mathrm{q(j)}} \right] \otimes |\Psi\rangle$$

The novel feature here is to construct these Clifford transformations hierarchically by compressing the Hamiltonian:

$$H_\varepsilon = \sum_{i:\,|h_i|\geq\varepsilon} h_i P_i$$

The Clifford symmetry group of this Hamiltonian becomes larger $S \subseteq S_\varepsilon$:

$$S_\varepsilon = \{\tau_j\}_{j=1,\ldots n\_sym} \cup \{\tau_j^\varepsilon\}_{j=1,\ldots n_\varepsilon\text{-}n\_sym}$$

Thewhere τ_j^ε are approximate symmetry of the full Hamiltonian H. The idea is that they **approximately block-diagonalize** H.

$$H' = C^{dag}(\varepsilon)\, H\, C(\varepsilon)$$

where $C(\varepsilon) = C_{tapering} \prod_{j=1}^{n\_\varepsilon\text{-}n\_sym,} C_j^\varepsilon$

The goal is to build this as a hierarchy of $\{\varepsilon_m\}$.

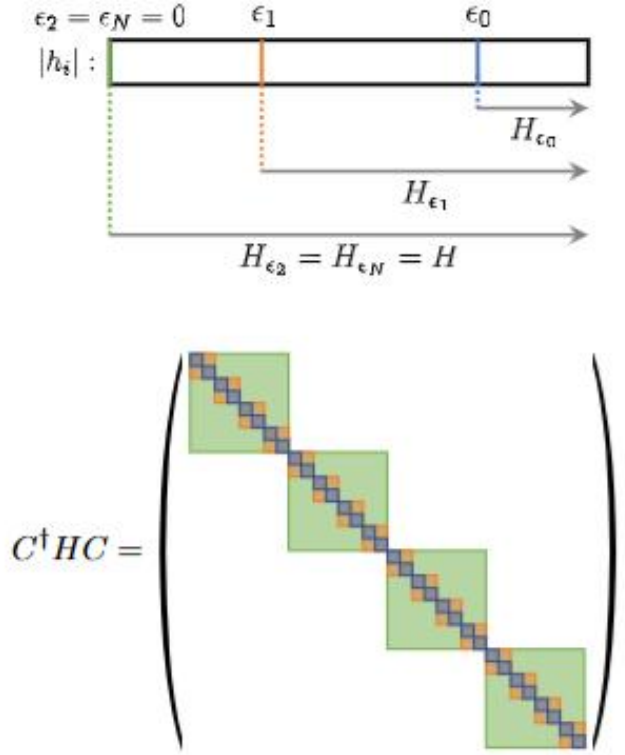


FIG. 1. (top) Schematic of the H_{ϵ_m} defined by truncating the full Hamiltonian $H = \sum_i h_i P_i$ according to the absolute values of the coefficients $|h_i|$ [cf. Eq. (19)] for a minimal example of $N = 2$ thresholds: $\epsilon_2 = \epsilon_N = 0 < \epsilon_1 < \epsilon_2$; $|h_i|$ is represented as increasing from left to right in the diagram. (bottom) Schematic of a $(2^n \times 2^n)$ matrix representation of the untruncated Hamiltonian, H, conjugated by the corresponding hierarchical Clifford unitary, $C = C(\{\epsilon_m\})$, as described in Sec. III. In this hypothetical example, we have $n_{\text{sym}} = 2$ exact symmetries ($2^2 = 4$ blocks; green) at threshold level $\epsilon_2 = \epsilon_N = 0$, $n_1 = 4$ (2 approximate, 2 exact) symmetries ($2^4 = 16$ blocks; orange) derived from ϵ_1, and $n_2 = 5$ (3 approximate, 2 exact) symmetries ($2^5 = 32$ blocks; blue) derived from the highest threshold ϵ_0. Matrix elements with larger magnitude live within smaller blocks by construction.

Results: Both figures show that the change of basis reorganizes the quantum state into a simpler form: the overall entanglement is significantly reduced (top fig), and correlations between qubits become more local (bottom fig). As a result, the wavefunction becomes easier to represent and optimize, which explains the improved performance of even simple ansätze.

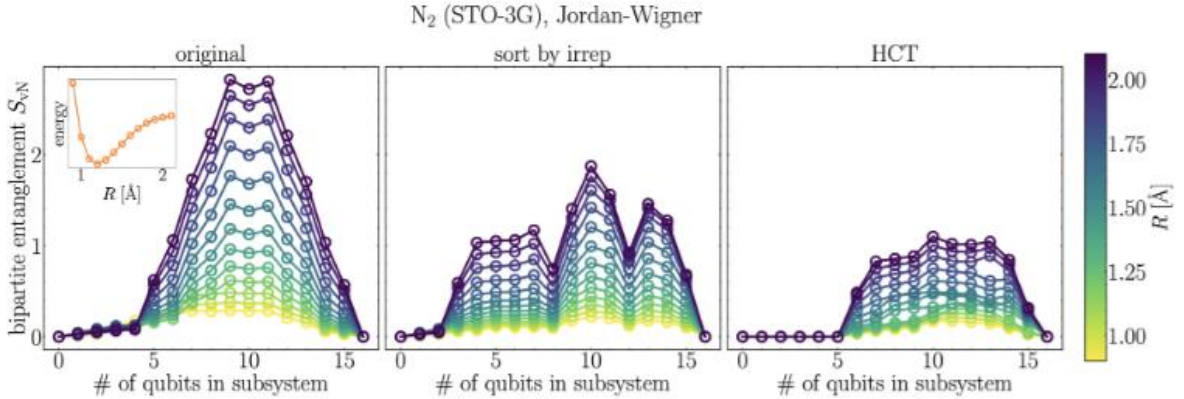


FIG. 4. Bipartite von Neumann entanglement entropy versus subsystem size for the ground state of N_2 represented in the STO-3G single-electron basis under Jordan-Wigner qubit mapping for (left) the original energy-based ordering of Hartree-Fock molecular orbitals, (middle) a sorting of molecular orbitals based on their Abelian point group irrep, and (right) the Hamiltonian under HCT, $C^\dagger HC$, as described in the text. The different curves correspond to different bond lengths R ranging from 0.9 Å to 2.1 Å in increments of 0.1 Å as indicated in the color bar; the corresponding ground state energy versus R (dissociation curve) is shown in the inset of the left panel (ranging from -107.677085 Hartree at $R = 1.2$ Å to -107.292712 Hartree at $R = 0.9$ Å). The sharp increase in entanglement as R is increased in the original representation is mitigated most effectively by the HCT scheme, which performs better than the standard irrep-based ordering strategy.

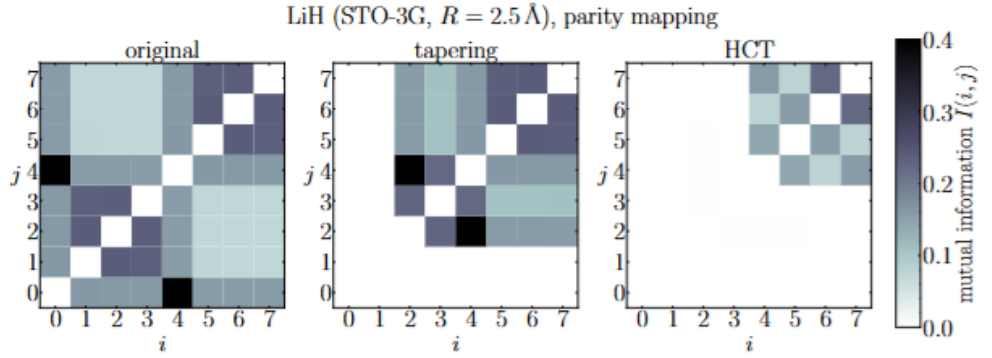


FIG. 7. Quantum mutual information between qubits i and j, $I(i,j)$, for the ground state of LiH in the STO-3G single-electron basis at (stretched) bond length $R = 2.5$ Å under parity qubit mapping for three different Hamiltonian representations: (left) the original qubit Hamiltonian H, (middle) the 'tapering' representation $C^\dagger_{\text{tapering}} H C_{\text{tapering}}$, and (right) the HCT representation $C^\dagger HC$. Two qubits are removed by tapering [16], hence the zeroed out 'empty' regions in the middle panel, while two additional nearly exact symmetries are identified by the HCT algorithm. This is the problem setup used for the (hardware-efficient ansatz HCT-based VQE) data shown in the bottom panel of Fig. 10.

Results: The change-of-basis leads to improved optimization performance even when using naïve ansatz (hardware-efficient ansatz).

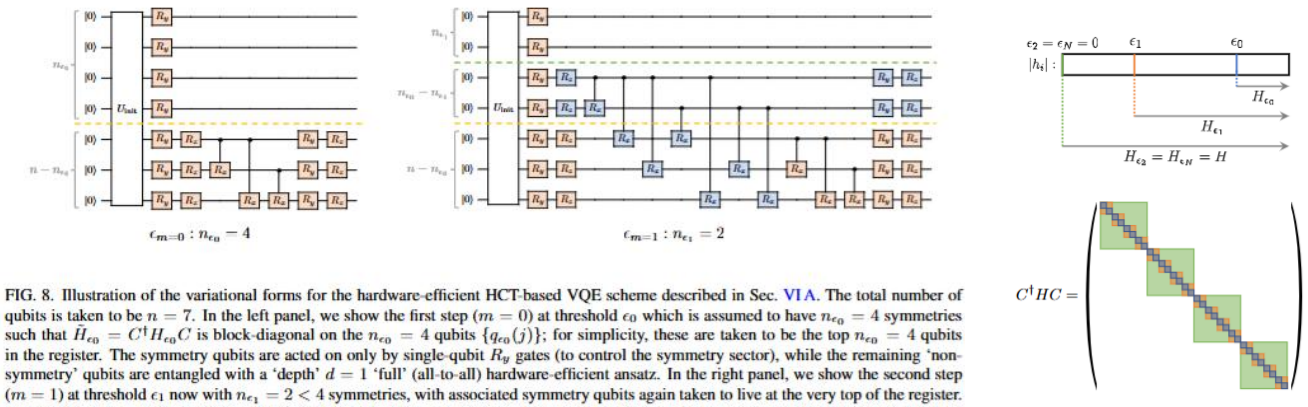


FIG. 8. Illustration of the variational forms for the hardware-efficient HCT-based VQE scheme described in Sec. VI A. The total number of qubits is taken to be $n = 7$. In the left panel, we show the first step ($m = 0$) at threshold ϵ_0 which is assumed to have $n_{\epsilon_0} = 4$ symmetries such that $\tilde{H}_{\epsilon_0} = C^\dagger H_{\epsilon_0} C$ is block-diagonal on the $n_{\epsilon_0} = 4$ qubits $\{q_{\epsilon_0}(j)\}$; for simplicity, these are taken to be the top $n_{\epsilon_0} = 4$ qubits in the register. The symmetry qubits are acted on only by single-qubit R_y gates (to control the symmetry sector), while the remaining 'non-symmetry' qubits are entangled with a 'depth' $d = 1$ 'full' (all-to-all) hardware-efficient ansatz. In the right panel, we show the second step ($m = 1$) at threshold ϵ_1 now with $n_{\epsilon_1} = 2 < 4$ symmetries, with associated symmetry qubits again taken to live at the very top of the register. Any circuit at $m = 0$ can be embedded into this circuit at $m = 1$ by zeroing out the blue gates in the latter and matching parameters for the orange gates. In our simulations, we take the initial prepended circuit to be $U_{\text{init}}|0\rangle^{\otimes n} = C^\dagger|\text{HF}\rangle$.

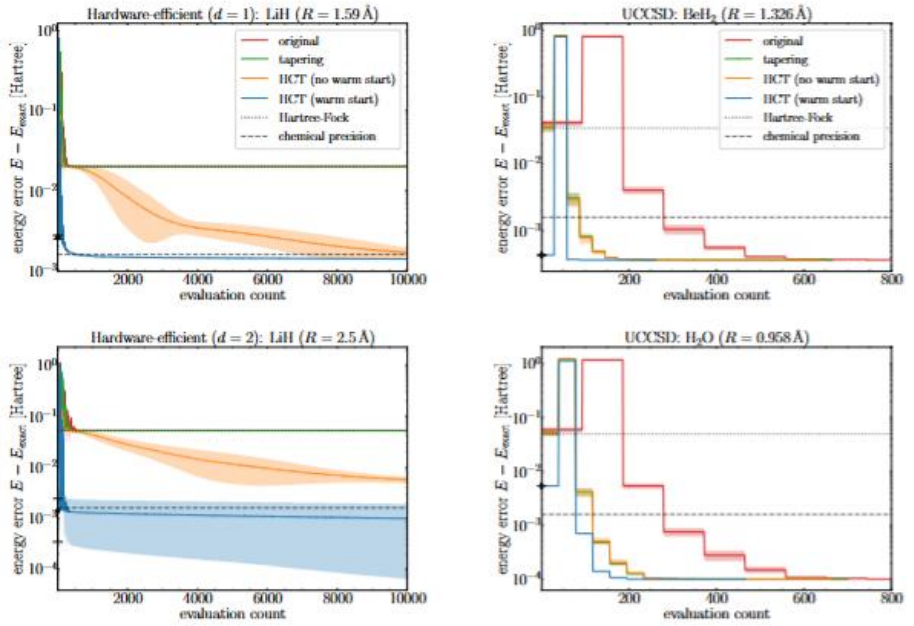


FIG. 10. Variational energy versus energy evaluation count during optimization (training trajectories) for hardware-efficient-based VQE simulations (statevector without noise) of LiH in STO-3G ($n = 8$ qubits; see text for details). Top: $d = 1$ repetition circuits at equilibrium bond length $R = 1.59$ Å; bottom: $d = 2$ repetition circuits at stretched bond length $R = 2.5$ Å. In each panel, we show VQE runs using the original qubit Hamiltonian representation H ('original'), the Hamiltonian in the $\mathbb{Z}_2$ tapering basis [16] $C^\dagger_{\text{tapering}} H C_{\text{tapering}}$ ('tapering'), and our HCT-based VQE with and without warm starting the solution from threshold to threshold (see text). For the HCT-based calculations with warm starting, we only show the trajectory for the final step (solution of $H_{\epsilon_{N-0}} = H$); the energy at the start of the final step optimization is plotted as a black diamond. Here we use the COBYLA optimizer and the shaded regions indicate spread of training trajectories starting from randomly chosen initial parameter points near the Hartree-Fock state (see text for details). Only the HCT-based strategies are able to improve upon the Hartree-Fock energy and are able to achieve 'chemical precision' in both cases.

FIG. 11. Data analogous to Fig. 10 for UCCSD-based VQE simulations of BeH_2 at $R = 1.326$ Å (top) and H_2O at $R = 0.958$ Å (bottom), both using the STO-3G single-electron basis ($n = 10$-qubit problems; see text for details). Here we choose the L-BFGS-B optimizer. For the HCT-based strategy with warm starting, the training for the final step exhibits relatively fast convergence owing to a good initial trial state obtained by the iterative procedure.

data) and the new ST based on approximate symmetries in the absence of iterative warm starting ['HCT (no warm start)' data]. Thus the green and orange training curves in Fig. 11 are equivalent. The influence of the HCT basis change on *depth* of the corresponding (full) UCCSD ansatz when compiled to a native gate set is an interesting detail we leave for future work. In any case, the full iterative scheme with warm starting does however provide potentially improved convergence properties as the solutions of the truncated Hamiltoni-

Conclusion: When you incorporate problem structure into your basis representation, you can turn a hard quantum optimization problem into a much simpler one. This paper describes an efficient way to do this for chemistry problems.