OVERLAPOTRON: ZERO-SHOT MOLECULAR TRANSFER FOR QUANTUM STATE PREPARATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Fault-tolerant quantum algorithms for chemistry need an initial state with high overlap with the target state, and the Hartree–Fock overlap decays exponentially with molecular size. Current methods build that state for one molecule at a time, from a classical solution of that same molecule or from a per-molecule quantum optimization. We propose Overlapotron, a model that predicts the initial-state preparation circuit of an unseen molecule from the molecule alone, with no target state and no per-molecule optimization at test time. An equivariant molecular-orbital encoder feeds heads whose output size follows the molecule while the learned weights stay fixed. We show transfer in two regimes. On exact state-vector targets, the heads emit unitary coupled cluster with singles and doubles (UCCSD) and coupled exchange operator (CEO) circuits. On density matrix renormalization group (DMRG) matrix product state (MPS) targets computed at a fixed bond dimension, we introduce the brickwall family, whose overlap with an MPS contracts as a tensor network, so training scales past the state-vector limit. One brickwall model trained on 25,877 QM7 states of at most 88 qubits transfers to QM9 targets of 90 to 112 qubits. Its predicted circuits exceed the overlap of a target-aware non-variational benchmark that compiles each target state directly, at 43% of its T count and 11% of its T depth. On held-out hydrogen chains of 68 to 100 qubits, the margin over the benchmark averages $5\times$ and grows to $8\times$ on the longest chain. Two new datasets of 30,000 CASCI state vectors and 7,165 DMRG matrix product states, with a 25,877-state geometry sweep, support the training.

1 INTRODUCTION

Fault-tolerant quantum computing (FTQC) promises large speedups over classical methods for many-body problems. Quantum chemistry is a central application, with relevance to solid-state physics, drug discovery, and materials science. The focus of quantum chemistry is to obtain the molecular wavefunction, which completely characterizes the molecule. Fault-tolerant quantum algorithms targeting this wavefunction require preparation of an initial state through the execution of a suitable quantum circuit. That initial state needs a high overlap with the target state (Aspuru-Guzik et al., 2005; Whitfield et al., 2011; Tubman et al., 2018; Fomichev et al., 2024). The overlap of simple initial states decays exponentially with system size, which adds an exponential overhead to the algorithm (Lee et al., 2023). Quantum advantage is contested at molecular active spaces of 70 to 200 qubits (Legeza et al., 2026), so state preparation must scale to more than 100 qubits.

Many classical and quantum methods address initial-state preparation, but each treats one system at a time or has a high classical or quantum cost (Tubman et al., 2018; Fomichev et al., 2024; Rupprecht & Wölk, 2026; Sugisaki et al., 2022). Machine learning methods so far have limited generality: initial-state candidates for one molecule (Nakaji et al., 2024), circuit parameters for geometries of a single molecule (Koziell-Pipe et al., 2026), or hydrogen systems only up to 24 qubits (Bincoletto et al., 2026).

In this paper we introduce Overlapotron, a machine learning model that generates initial-state preparation circuits across molecules and their geometries, for systems up to 112 qubits. Overlapotron uses an equivariant molecular-orbital encoder (Thiede et al., 2026) to predict quantum circuits. To train the model, we create two large, chemically diverse datasets of molecules and geometries paired with classically computed target states. We use two methods to compute the target states: complete

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

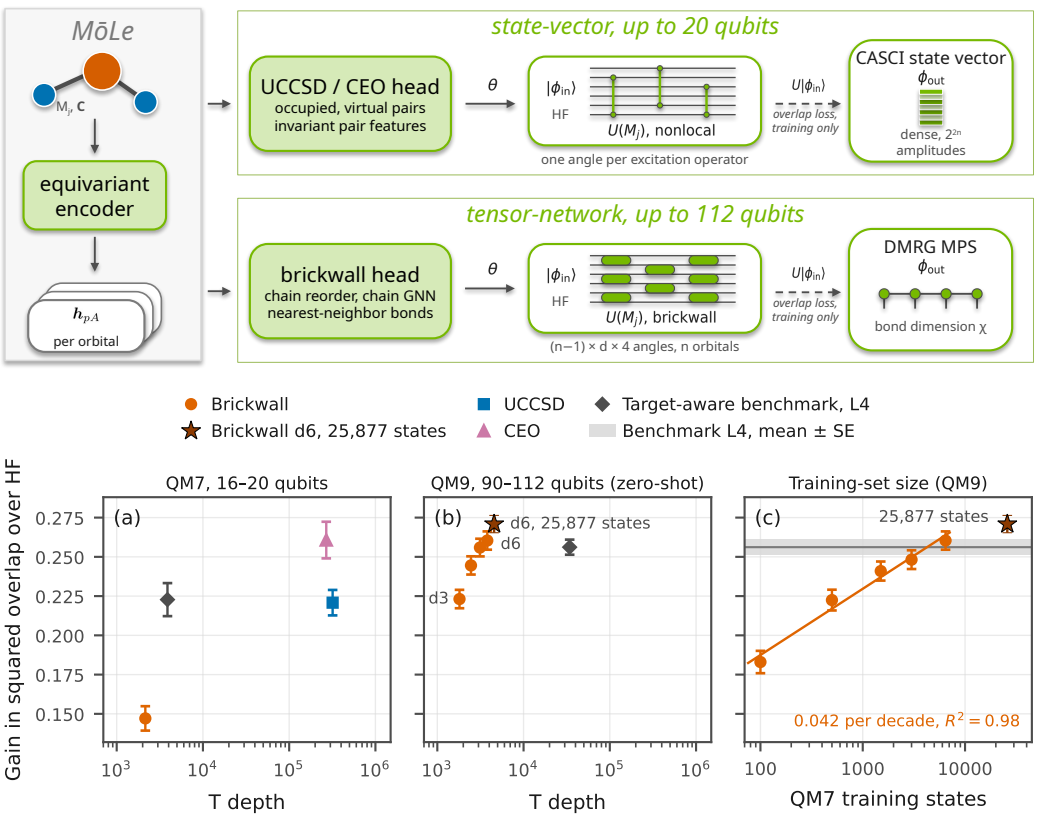


Figure 1: **Overview of Overlapotron.** *Top:* an equivariant molecular-orbital encoder (MōLe) maps a molecule to orbital embeddings. A head for each circuit family predicts the circuit angles θ. The circuit $U(M)$ acts on the Hartree–Fock (HF) state. Target states enter only the training loss, so a test molecule needs no target. *Bottom:* gain in squared overlap over HF against T depth, the number of sequential layers of T gates (a proxy for runtime). (a) QM7 validation molecules, 16–20 qubits. (b) QM9 test molecules, 90–112 qubits, each larger than every training molecule. (c) The depth-six brickwall on QM9 against the number of QM7 training states. The grey benchmark compiles each test molecule's target state directly at four layers (Ran, 2020). Error bars are one standard error over molecules. The star is the depth-six brickwall trained on all 25,877 QM7 training states; in (b) it has the same T depth as the d6 circle and is drawn slightly to its right for legibility. Appendix E gives the cost model.

active space configuration interaction (CASCI), which is exact, and the density matrix renormalization group (DMRG), which is approximate but scales to large systems. DMRG returns the state in a compressed tensor-network form, a matrix product state (MPS).

Molecules differ in atom order and orientation, and larger molecules need circuits with more angles. We design heads whose output is invariant to atom order and rotation. The output size follows the molecule while the number of learned parameters stays fixed. The heads emit circuit angles that run directly, with no further classical optimization (Sec. 4.2). For the exact CASCI targets, the heads emit two circuit families from the literature, unitary coupled cluster with singles and doubles (UCCSD) (Romero et al., 2019) and coupled exchange operators (CEO) (Ramôa et al., 2025). For the DMRG targets, we introduce the brickwall family, whose overlap with an MPS target is computed on a classical computer at a cost polynomial in system size, so we compute the overlap at every size we study, up to 112 qubits. The heads produce circuits with substantial overlap gains over Hartree–Fock in both regimes. On the DMRG targets, the brickwall circuits also outperform a target-aware non-variational benchmark, the sequential disentangler of Ran (2020), which compiles each circuit from the target state itself, information that is not available at deployment. The brickwall circuits do so at a fraction of the benchmark's T count and T depth, the two standard cost measures

for fault-tolerant circuits. We transfer to 112 qubits, where prior transferable circuit-parameter prediction reached 24 qubits. Our hydrogen-chain study extrapolates to unseen chain lengths, and the zero-shot overlap grows with the logarithm of the training-set size.

Our contributions are:

- **Zero-shot circuit prediction problem.** We cast transferable state preparation as prediction of molecule-sized Ansatz parameters from the molecule alone, with no target state and no per-molecule optimization at test time. Overlapotron, an equivariant molecular encoder with invariant, size-independent heads, provides a solution: one model trained on QM7 molecules of at most 88 qubits transfers to QM9 molecules of 90 to 112 qubits, where prior transferable methods reached 24 qubits.
- **Brickwall Ansatz.** We introduce a circuit family whose angles are predicted along the chain of the target matrix product state, so the training loss contracts as a tensor network and training scales past the state-vector limit of about 20 qubits. On 100 QM9 molecules, its predicted circuits exceed the overlap of the target-aware non-variational disentangler of Ran (2020), which compiles each target state directly, at 43% of its T count and 11% of its T depth. On held-out hydrogen chains of 68 to 100 qubits, up to more than twice the longest training length, the margin grows from $3.9\times$ to $8.2\times$.
- **Two datasets of multireference target states.** QM7-StateVec-30k and QM7-MPS-7k pair 30,000 CASCI state vectors and 7,165 DMRG matrix product states with QM7 geometries stretched to increase multireference character, under one uniform active-space construction. A sweep of 25,877 geometries of the same molecules trains the strongest model. Zero-shot overlap grows with the logarithm of the training-set size and has not saturated at the size of QM7.

2 Related work

Quantum phase estimation has a success probability proportional to the squared overlap with the target state (Kitaev, 1996; Aspuru-Guzik et al., 2005; Whitfield et al., 2011), and projector or Monte Carlo methods also benefit from high-quality trial states (Zhang & Krakauer, 2003). The Hartree–Fock (HF) states many algorithms start from are cheap to prepare both classically and on a quantum computer, but often have poor overlap in correlated regimes (Tubman et al., 2018; Lee et al., 2023; Fomichev et al., 2024).

Existing approaches broadly fall into one of two categories. The first is to calculate an approximation on a classical (non-quantum) computer and then prepare that state on a quantum computer (Greene-Diniz et al., 2026; Rupprecht & Wölk, 2026). Classical approximations include selected CI methods (Tubman et al., 2018), spin-coupled initial states (Marti-Dafcik et al., 2025), and DMRG (White, 1992; Chan & Head-Gordon, 2002; Sharma & Chan, 2012), with orbital optimization (Ollitrault et al., 2024) or refinement on a quantum computer also possible (Fomichev et al., 2024; Rupprecht & Wölk, 2026). The second is to do optimization directly on a quantum computer. These methods include adiabatic state preparation techniques (Farhi et al., 2000; Aharonov et al., 2007; Aspuru-Guzik et al., 2005; Sugisaki et al., 2022) and the Variational Quantum Eigensolver (VQE) (Peruzzo et al., 2014). To date these methods need substantial quantum resources, for the preparation itself or for its optimization.

Machine learning methods have recently emerged to address these limitations by training a model to produce quantum circuits by amortizing the costs of the optimization process. One strand of work has focused on learning to generate quantum circuits for a single molecule and geometry. The Generative Quantum Eigensolver (GQE) and its extensions train a generative model to produce a quantum circuit for a provided molecule, shifting the optimization problem from the quantum circuit parameters to the generative model parameters, which are generally better behaved (Nakaji et al., 2024; Keithley et al., 2026; Kemmoku et al., 2026). Other methods have learned to generate the circuit structure with a classical post-processing optimization step for the circuit parameters (Ostaszewski et al., 2021; Patel et al., 2024; Kundu & Sarra, 2026; Zhang et al., 2021; Li et al., 2025) or jointly generate the circuit structure and parameters (Vahedi & Arefi, 2026; Chen et al., 2026). More recent work has investigated generating quantum circuit parameters at different geometries for a single molecule (Tao et al., 2022; Mesman et al., 2025; Zou et al., 2025), or jointly generating circuit

structure and parameters (Krumtünger et al., 2025). GQE has been shown to transfer to different geometries for a single molecule (Koziell-Pipe et al., 2026). Bincoletto et al. (2026) demonstrated transfer across hydrogen systems, reaching H_{12} in a 24-qubit minimal-basis representation. Unlike prior work, we demonstrate chemically diverse transfer, direct overlap optimization, and transfer to active spaces of more than 100 qubits. Our model builds on molecular representations designed for cross-molecule transfer (Behler & Parrinello, 2007; Bartók et al., 2010; Batzner et al., 2022; Thiede et al., 2026).

3 BACKGROUND

3.1 PROBLEM SETTING

Given a fixed *Ansatz*, i.e. a prescription of which gates act on which qubits in which order, together with the parameters modulating them, we define a molecular instance M_j specifying the geometry, atom types, atom-centered orbital basis set, active-space selection, and molecular orbitals used to define an electronic-structure problem. The molecular orbitals represent sites that the electrons can occupy. From this information we construct the model input described in Sec. 3.2. The learning problem is then to train a parametrized predictor that takes M_j as input and outputs the circuit parameters that, together with the fixed Ansatz, fully specify a quantum circuit $U(M_j)$. The size of the circuit specification is determined by the molecule: larger active spaces mean more orbital sites, more allowed gates or bonds, and larger parameter tensors to predict.

The circuit is applied to an initial state, chosen in our case to be the Hartree–Fock state $|\phi_{\mathrm{in}}(M_j)\rangle$, which is the result of solving the electronic structure problem when approximating electron-electron interactions by their mean field. That state is also very easy to prepare in a quantum circuit, requiring a single gate per qubit occupied by an electron. For training and evaluation, we also define a target state $|\phi_{\mathrm{out}}(M_j)\rangle$, which may be represented either as a dense state vector or as a matrix product state, depending on the supervision regime (see Sec. 3.3).

The primary metric is the squared overlap

$$\mathcal{O}_\theta(M_j) = \left|\langle\phi_{\mathrm{out}}(M_j)|U(M_j)|\phi_{\mathrm{in}}(M_j)\rangle\right|^2 . \tag{1}$$

Training minimizes an overlap loss such as $1 - \mathcal{O}_\theta(M_j)$ over a dataset D_{train} by gradient descent through a differentiable overlap estimator, so every training molecule needs a target representation. The zero-shot setting evaluates the trained predictor on a different dataset D_{test} without per-molecule circuit optimization, and the generated state $U(M_j)\,|\phi_{\mathrm{in}}(M_j)\rangle$ serves as the initial state of a downstream quantum algorithm. The intended use is a D_{test} whose target states are classically intractable, so the predictor must generate each circuit without its target. Our experiments stay where the target can still be computed, because the overlap metric requires it.

As a reference we report the overlap with no circuit applied, $\left|\langle\phi_{\mathrm{out}}(M_j)|\phi_{\mathrm{in}}(M_j)\rangle\right|^2$, the value that Hartree–Fock preparation alone achieves.

3.2 MODEL INPUT

As the model input M_j, we use the same input representation as Molecular Orbital Learning (MōLe) (Thiede et al., 2026): given a molecular geometry $\{\mathbf{R}_A\}$, we generate molecular orbitals (MOs) ψ_p, localize the occupied and virtual ones within their respective subspaces (Boys, 1960; Pipek & Mezey, 1989; Knizia, 2013), and feed the model their coefficients $C_{pA,k\ell m}$ in an atom-centered atomic-orbital basis together with $\{\mathbf{R}_A\}$ and atom types. Here p indexes the MO, A indexes atoms, and k, ℓ, m label the radial shell and angular-momentum channel of the atomic basis, zero-padded across atom types to a common channel set per ℓ; App. C.1 gives the expansion and the padding convention, App. A the active-space and MO generation.

We regard the padded input as a collection of molecular graphs, one for each of the N MOs: the nodes are atoms, node A carries $C^{\mathrm{padded}}_{pA,k\ell m}$ together with an atom-type attribute, and edges carry the interatomic distance $d_{AB} = |\mathbf{R}_B - \mathbf{R}_A|$ and direction $\mathbf{R}_B - \mathbf{R}_A$. This input is not tied to a fixed number of atoms: a larger molecule contributes more atomic nodes and more MO graphs, while

the local feature channels remain fixed after padding. Molecule-to-molecule transfer is therefore formulated as applying shared graph-based operations to variable-size molecular inputs.

MōLe uses an equivariant encoder to construct molecular-orbital representations and task-specific heads to read out coupled-cluster amplitudes (Thiede et al., 2026). Overlapotron reuses this encoder, that takes the molecular input and outputs embeddings $\{\mathbf{h}_{pA}(M)\}_{p,A}$, where $\mathbf{h}_{pA}$ is the embedding of orbital p on atom A. The encoder is size-independent: it mixes orbitals by attention and atoms by geometric message passing, and produces a set of embeddings rather than a flattened molecule vector. Encoder internals and the orbital-sign convention are given in App. C.

3.3 Supervision representations

We use two supervision regimes that differ in how the target $|\phi_{\mathrm{out}}(M_j)\rangle$, a superposition of electron configurations, is represented.

State-vector supervision. In the state-vector regime, $|\phi_{\mathrm{out}}(M_j)\rangle$ is the CASCI ground state. CASCI solves the electronic-structure problem of M_j exactly within the active space, at exponential cost, and returns one real amplitude per electron configuration (with no magnetic field). Dense state-vector simulation then evaluates $\mathcal{O}_\theta$ exactly, but it needs $O(2^{2N})$ memory, so this regime covers only small systems.

Tensor-network supervision. In the tensor-network regime the same 2^{2N}-dimensional vector is written as a chain of N site tensors, one per spatial orbital, each with a four-valued occupation index: empty, spin up, spin down, or both. An exact representation needs bonds of exponential dimension. DMRG truncates them at a maximum bond dimension χ, which sets both accuracy and cost, and never materializes the full vector. We fix the particle number and total spin by optimizing in a spin-adapted $SU(2)$ representation and then convert to the occupation form (White, 1992; Chan & Head-Gordon, 2002; Sharma & Chan, 2012). The cost scales as $N\chi^3$, which brings the hundred-qubit benchmark within reach.

4 Methods

We introduce an architecture that produces circuits for state-vector and tensor-network supervision alike. For each case we present the Ansatz and the head that emits its parameters with permutation and size invariance.

4.1 Ansatz

In the state-vector supervision regime, we use either a UCCSD or CEO circuit family (Romero et al., 2019; Ramôa et al., 2025). UCCSD constructs a circuit from exponentiated single- and double-excitation generators, with one predicted angle assigned to each generator in a fixed order (Romero et al., 2019). CEO instead combines qubit excitations acting on the same orbital support into shared generators, reducing the entangling-gate cost while retaining chemically structured excitation operators (Yordanov et al., 2021; Ramôa et al., 2025).

Because these circuit families contain nonlocal generators, their overlaps cannot be evaluated efficiently by tensor-network contraction, even when the target state is represented as an MPS. To enable classically tractable training at the hundred-qubit scale, we introduce the brickwall Ansatz.

The brickwall Ansatz couples neighboring spatial orbitals, where neighboring orbitals are defined as adjacent sites in the one-dimensional MPS ordering. Each active bond carries four angles per layer, so the parameter count grows linearly with the number of orbitals. Let d denote the circuit depth, $\mathcal{B}_\ell(M)$ the set of disjoint neighboring-orbital pairs in layer ℓ, and $R_g(\theta)$ the two-site rotation of type g. The indices $g = 1, \ldots, 4$ refer, in order, to the spin-up Givens, spin-down Givens, pair-hopping, and spin-exchange rotations. The parameter $\theta^{\mathrm{BW}}_{ij\ell g}(M)$ is the angle of rotation g on bond (i, j) in layer ℓ, giving

$$U_{\mathrm{BW}}(M) = U_{d-1}(M) \cdots U_0(M), \qquad U_\ell(M) = \prod_{(i,j)\in\mathcal{B}_\ell(M)} \left[R_4\big(\theta^{\mathrm{BW}}_{ij\ell 4}\big) \cdots R_1\big(\theta^{\mathrm{BW}}_{ij\ell 1}\big)\right]_{ij}. \quad (2)$$

4.2 CIRCUIT PREDICTION HEADS

For circuit family a, a head maps the inherited encoder output to the molecule-sized angles $\theta^{(a)}(M)$ that define the corresponding circuit. Three requirements fix the form of such a head: *invariance* with respect to input permutations and rotations; compatibility with variable-size input and output; and output angles that are directly usable in the circuit, in the correct order.

4.2.1 FIXED-ORDER UCCSD AND CEO HEADS

These two heads follow the amplitude readouts of MōLe (Thiede et al., 2026), which map the same embeddings to coupled-cluster amplitudes. The scalar readout is unchanged; the heads change only its target, a circuit angle, and the order of the output. Writing $\mathbf{h}_{pA}$ for the encoder embedding of orbital p on atom A, a single-excitation angle is read from an occupied–virtual pair (i, a) by an equivariant projection of $\mathbf{h}_{iA}$ and $\mathbf{h}_{aA}$, an inner product, and a sum over atoms, Eq. (11); the sum over atoms makes the readout invariant to atom permutations and independent of the atom count. Double excitations contract two such pair features, and both pools resolve spin, so one orbital tuple carries several angles: singles have α and β; doubles have $\alpha\beta$, $\alpha\alpha$ and $\beta\beta$, which CEO subdivides further still. Each angle is one unconstrained scalar from a zero-initialized projection. Appendix C gives the pair feature, the quadruple contraction, the spin sectors, and the pool order and block lengths of each head.

4.2.2 BRICKWALL HEAD

Brickwall angles are indexed by chain bonds instead of by excitations, so this head reorders the orbital embeddings onto the chain of Sec. 4.1 and propagates information along it; Fig. 4 in App. C shows the resulting data flow. Abbreviating $\mathbf{h}_i = \{\mathbf{h}_{iA}\}_A$ for the i-th chain site, a residual chain graph network applies

$$\mathbf{h}_{iA}^{(t+1)} = \mathbf{h}_{iA}^{(t)} + W_{\mathrm{s}}\mathbf{h}_{iA}^{(t)} + W_{\mathrm{L}}\mathbf{h}_{i-1,A}^{(t)} + W_{\mathrm{R}}\mathbf{h}_{i+1,A}^{(t)}, \tag{3}$$

with equivariant maps W shared by every site and every circuit layer. Each angle is then read out from the two endpoints of its own bond. Angles are periodic and network outputs are not, so the readout emits two real components and takes the angle as their direction in the plane:

$$\left(r^{(1)}, r^{(2)}\right)_{i,i+1,\ell,g}(M) = f_{g\ell}(\mathbf{h}_i, \mathbf{h}_{i+1}, e_\ell)\,, \qquad g = 1, \ldots, 4, \tag{4}$$

$$\theta^{\mathrm{BW}}_{i,i+1,\ell,g}(M) = \mathrm{atan2}\Big(r^{(2)}_{i,i+1,\ell,g}(M),\, 1 + r^{(1)}_{i,i+1,\ell,g}(M)\Big)\,. \tag{5}$$

Here $f_{g\ell}$ builds an invariant bond feature by the inherited construction of Eq. (11), concatenates the learned embedding e_ℓ of circuit layer ℓ, and applies a residual multilayer perceptron whose two outputs are $r^{(1)}$ and $r^{(2)}$. The offset of one places a zero-initialized readout at $(1, 0)$, so every generated circuit starts at the identity and the loss at the Hartree–Fock overlap, as for the dense heads (App. C). One readout is kept per rotation type and circuit layer and shared by all bonds; the head emits every bond of every layer, and the parity mask $\mathcal{B}_\ell$ of Eq. (2) selects those that act in layer ℓ. The new element is the chain propagation: the readout is the same inner product as in the dense heads, but it reads embeddings that Eq. (3) has propagated along the chain. Without Eq. (3) and the orbital-interaction channels below, the head reduces to the MōLe readout on adjacent pairs, which barely leaves the Hartree–Fock state (App. F). The hydrogen-chain models use the head as written. The molecular models add the orbital-interaction matrix that also fixes the Fiedler order, as a local bond weight and a nonlocal message in both equations, so strongly coupled orbitals far apart on the chain inform each other's angles at no extra Ansatz parameters (App. C).

5 EXPERIMENTS

5.1 DATASETS

Transfer across molecules needs training data that share atom types and orbital construction with the test molecules and that contain strong correlation. We build two datasets from QM7 (Blum & Reymond, 2009; Rupp et al., 2012) to this end, one with exact state-vector targets and one with

DMRG targets. Both stretch the molecular geometry to increase multireference character (Karton, 2025), and both use one active-space and orbital construction (App. A).

QM7-StateVec-30k contains 30,000 records from 6,838 C/H/N/O parent molecules, each at up to five uniform bond-stretch factors from 1.2 to 2.0. The active spaces have 16, 18, or 20 qubits, with 10,000, 10,000, and 9,999 records; one faulty record was removed. The target is the lowest CASCI singlet state in the same localized-orbital basis as the Hartree–Fock input and the circuit operators. We split by parent molecule, so every geometry of a held-out molecule stays outside training.

QM7-MPS-7k sweeps the stretch factor and keeps one geometry per molecule, chosen so that the Hartree–Fock determinant no longer dominates the DMRG ground state (App. A). It contains 7,165 active spaces of 16 to 88 qubits, with all STO-3G orbitals except frozen cores, split into 6,449 fitting and 716 checkpoint-selection molecules. A larger training set keeps every swept geometry of the same fitting molecules with a Hartree–Fock overlap amplitude of at least 0.20, a squared overlap of at least 0.04: 25,877 states over 6,447 molecules, about four per molecule. Two fitting molecules have no geometry above that threshold and drop out of this set. We evaluate the selected models without fine-tuning on QM9-MPS-100, 100 uniformly deformed QM9 molecules of 90 to 112 qubits, four to six isomers of each of twenty molecular formulas. Their targets enter neither optimization nor checkpoint selection.

Hydrogen chains isolate size extrapolation from chemical diversity. We fit on H_6, H_{10}, H_{14}, H_{18}, and H_{22} (12–44 qubits), select checkpoints on H_{26} and H_{30}, and hold out H_{34}–H_{50} (68–100 qubits). Appendix A gives the geometry, active space, orbital ordering, and DMRG bond dimension.

5.2 State-vector targets

Across all experiments we use the iterative sequential matrix-product disentangler of Ran (2020) as a target-aware non-variational benchmark method for initial state preparation. It compiles the complete target $|\phi_{\text{out}}(M_j)\rangle$ of each test molecule directly, information that Overlapotron does not have at test time, and it does no learning across molecules. We keep the benchmark non-variational, rather than refining its layers against the target (Rudolph et al., 2024), because Overlapotron likewise emits its circuit in a single forward pass and spends no optimization on the test molecule; the two then differ in what they know about that molecule, not in how much optimization each is given on it.

We evaluate the CASCI state-vector regime of Sec. 3.1 against it, with Overlapotron trained to produce UCCSD, CEO, and depth-four brickwall Ansatz families on QM7-StateVec-30k (Tab. 1). Scoring uses 300 validation records of the 30,000-record dataset and reports the sample standard deviation across five training seed means. Note that these molecules exclude every training parent molecule but were used to select the checkpoints. All three heads gain substantially over the Hartree–Fock reference. CEO reaches the highest overlap of the four methods, with UCCSD and the benchmark following within 0.04 of it. Brickwall loses 0.11 of overlap but costs only 3% of the T count of the dense families and 55% of the benchmark's T depth (App. G gives results by molecule).

5.3 DMRG-MPS targets

We study brickwall in three settings: (1) size extrapolation without molecular diversity; (2) size extrapolation and molecular diversity; and (3) scaling with circuit depth and training-set size.

Size extrapolation. Hydrogen chains isolate size extrapolation from chemical diversity as every length shares one composition and one orbital ordering rule. Difficulty increases with chain length: the Hartree–Fock overlap falls from 0.178 on H_6 to 1.01×10^{-7} on H_{50} (Fig. 2). We train depth-six brickwall on H_6–H_{22} and use H_{26} and H_{30} as validation data to select a checkpoint. Brickwall reaches a mean squared overlap of 0.4082 ± 0.0179 on the five strictly held-out chains H_{34}–H_{50}, a factor of 2.5×10^4 to 3.1×10^6 above the Hartree–Fock reference. By comparison, the target-aware benchmark with six layers (the highest we compiled, App. D) reaches 0.0781 on those chains, making brickwall $5.2\times$ higher by a margin that grows with length, from $3.9\times$ on H_{34} to $8.2\times$ on H_{50}. Brickwall circuits are less costly, at 29% of the benchmark's T count and 12% of its T depth; the T-depth ratio grows from $7.0\times$ on H_{34} to $10.0\times$ on H_{50}.

Table 1: Squared overlap, gain over Hartree–Fock (HF), and T costs under one synthesis budget of 10^{-3} per circuit (App. E). The target-aware benchmark is Ran (2020) at four layers (L4). On QM7-StateVec-30k, learned entries are means over five training seeds $\pm$ the standard deviation of the seed means, on the 300 validation records. On QM9-MPS-100, entries are means over the 100 molecules $\pm$ one standard error, and the brickwall is the depth-six model trained on all 25,877 states; Tab. 2 lists the other models.

Method	Depth	Squared overlap	Gain over HF	T count	T depth
QM7-StateVec-30k, state-vector targets					
Hartree–Fock	—	0.5986	0	0	0
Target-aware benchmark	L4	0.8214	+0.2228	18,807	3,880
brickwall	d4	0.7457 ± 0.0006	+0.1471	10,442	2,148
UCCSD	—	0.8195 ± 0.0034	+0.2208	321,172	318,997
CEO	—	0.8593 ± 0.0022	+0.2607	321,172	269,149
QM9-MPS-100, DMRG-MPS targets					
Hartree–Fock	—	0.3683 ± 0.0101	0	0	0
Target-aware benchmark	L4	0.6245 ± 0.0119	+0.2562	$250{,}101 \pm 1{,}864$	$34{,}237 \pm 240$
brickwall	d6	0.6392 ± 0.0123	+0.2709	$108{,}402 \pm 824$	$3{,}796 \pm 3$

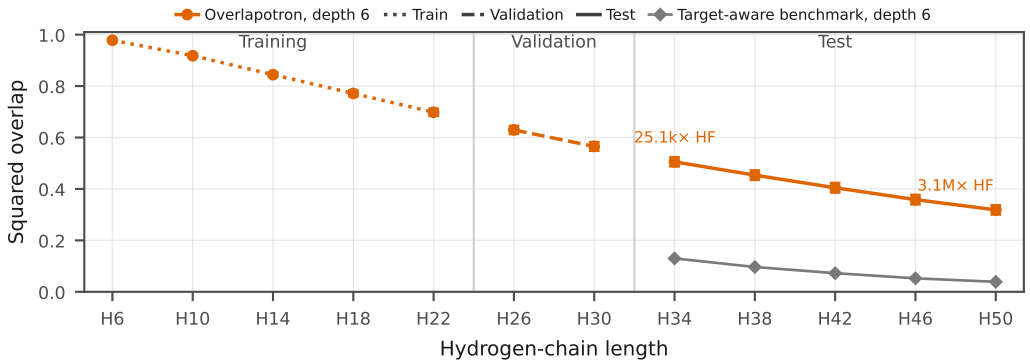


Figure 2: Size extrapolation across the full hydrogen-chain sweep. Overlapotron, the learned brickwall at depth six, is a mean over three training seeds with error bars of one standard error over them. Line type marks the three spans: fitted on H_6–H_{22} (dotted), checkpoint-selected on H_{26} and H_{30} (dashed), strictly held out on H_{34}–H_{50} (solid). The target-aware benchmark at depth six, compiled separately from each target MPS, is drawn on the held-out span. Annotations at its two ends give the improvement factor over the Hartree–Fock reference, itself too small to plot here.

Molecular extrapolation. For chemically diverse extrapolation we train depth-six brickwall on all 25,877 QM7 training states and test on QM9-MPS-100 (Fig. 1b). On paired molecules, brickwall surpasses the four-layer target-aware benchmark by $+0.015$ (95% CI $[0.010, 0.020]$; $[0.008, 0.021]$ with molecules clustered by formula) and an eight-layer version by $+0.009$ ($[0.004, 0.014]$; $[0.002, 0.016]$ clustered). Brickwall uses 43% of the benchmark's T count and 11% of its T depth.

Scaling behavior. Brickwall shows favorable scaling over both circuit depth and training-set size (Fig. 1b and c). Trained on QM7-MPS-7k and evaluated on QM9-MPS-100, increasing brickwall depth from three to six increases improvement over Hartree–Fock from 0.223 ± 0.006 at depth three through 0.245 ± 0.006 and 0.256 ± 0.006 to 0.260 ± 0.006 at depth six. This matches the four-layer target-aware benchmark, with a molecule-paired difference of 0.004 and a 95% confidence interval of $[-0.002, 0.010]$ over the 100 test molecules. Each depth is one training run; two further seeds at depth six give mean improvements of 0.258 and 0.265 around the reported 0.260. Appendix G.2 provides results by molecule, formula, and qubit count.

Brickwall also shows clear scaling with more molecules and more geometries for those molecules (Fig. 1c). We trained depth-six brickwall on increasing size nested subsets of QM7-MPS-7k that

differ in the number of molecules with one geometry each. Each model was then selected using the same 716 validation molecules and evaluated on QM9-MPS-100. The mean gain over Hartree–Fock rises monotonically from 0.183 at 100 molecules to 0.260 at 6,449 molecules, within 0.002 of the eight-layer target-aware benchmark. The improvements show logarithmic scaling, improving by 0.042 overlap for each factor of 10 ($R^2 = 0.98$). Brickwall trained on all 25,877 states, the same molecules with about four geometries each, improves from 0.260 to 0.271 ± 0.006. Gains from more geometries per molecule are smaller than gains from more molecules.

6 Discussion

Overlapotron predicts state-preparation circuit parameters from the molecule alone, for molecules it has not seen and at active-space sizes beyond its training data. On exact state-vector targets, the CEO head reaches the highest overlap, and the depth-four brickwall trails it by 0.11 at 3% of the T count of the dense families. On DMRG targets, one brickwall model trained on QM7 molecules of at most 88 qubits transfers to QM9 molecules of 90 to 112 qubits, and a second model trained on H_6–H_{22} extrapolates to chains up to more than twice the longest training length. In both DMRG settings the predicted circuits exceed the target-aware benchmark at 29% to 43% of its T count and about a tenth of its T depth. Two design choices support this transfer. The equivariant encoder and invariant heads keep the parameter count fixed while the output size follows the molecule, and the brickwall head predicts its angles along the chain of the target MPS, so the loss contracts as a tensor network at any size. The zero-shot overlap grows with the logarithm of the number of training molecules and has not saturated at the size of QM7, so larger training sets are the most direct route to higher overlap. Together these results support a workflow in which target states are computed classically for tractable molecules and Overlapotron is applied beyond that range.

6.1 Limitations

Targets. The DMRG targets have a fixed bond dimension, 250 for the molecules and 500 for the hydrogen chains. We report overlap with these fixed-bond-dimension MPS targets; our results concern transfer to those targets and do not establish convergence to the exact ground states. The state-vector targets are exact within their active space.

Benchmark. The benchmark of Ran (2020) runs at a fixed layer count with no per-target optimization. Rudolph et al. (2024) reach higher fidelity at the same depth by variationally refining its layers; that refinement needs the target state and could equally start from our predicted circuit. We have not run either method with per-target optimization. Loaders with depth logarithmic in the register (Malz et al., 2024) need measurement and feedforward and fall outside our cost model. Our margin over the benchmark also depends on training data: the brickwall matches the four-layer benchmark with 6,449 training molecules and exceeds it only with all 25,877 training states.

Cost model. The T-cost estimates assume all-to-all connectivity and one synthesis budget per circuit, and they exclude routing, error correction, and magic-state distillation. If a compiler synthesizes each one-qubit unitary of the benchmark as a single rotation, its T count on QM9-MPS-100 falls to about $1.2\times$ the brickwall's, while its T depth stays about $5\times$ larger (App. E).

Failures and scope. On the state-vector set, a few geometries receive a circuit below Hartree–Fock, and one $C_5H_8O_2$ geometry reaches an overlap near 10^{-7} under UCCSD and CEO (App. G); a prepared state therefore needs a check, such as its energy, before use. The brickwall gains least where the Hartree–Fock overlap is below 0.20: on the 59 such QM7-StateVec-30k records it gains 0.014 where CEO gains 0.512, and the one such QM9-MPS-100 molecule is the weakest record for every method (App. G). Both DMRG training sets contain such geometries, but only one test molecule probes this regime. All experiments use STO-3G active spaces of C/H/N/O molecules from one automated procedure, and the brickwall head uses a supplied orbital ordering that we do not ablate.

AI use statement

In this work, we used generative AI tools to implement methods. This use includes the model, training, and evaluation code, the target-aware benchmark, and the T-count and T-depth estimates.

We also used these tools to write the code that generates, converts, and reformats our datasets. The dataset targets come from classical CASCI and DMRG calculations, not from generative AI. We also used these tools for feedback on our research methodology and experiments, and to help interpret results. We have not used generative AI tools to develop theoretical models or conceptual frameworks, or to formulate mathematical claims. We have not used them to provide critical ingredients for proofs, to propose or refine hypotheses, or for translation. The writing of proofs and qualitative or thematic data analysis are not applicable to this work. Additionally, we used generative AI tools to suggest experimental parameters and to create and edit analysis code and figures. We also used them to draft parts of the paper, to edit the text for readability, and to identify relevant literature. We have reviewed all AI-assisted work. The authors reviewed and tested all AI-written code. We checked each reported number against the result files and logs. We checked the circuit definitions and derivations by hand and against the implementation. The authors decided which AI feedback on methodology to act on, and we checked each AI-assisted interpretation against the results. Before submission, we checked every citation by hand against its source. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

Ethics statement

This work uses no human subjects and no personal data. The molecular geometries come from the public QM7 (Blum & Reymond, 2009; Rupp et al., 2012) and QM9 (Ruddigkeit et al., 2012; Ramakrishnan et al., 2014) datasets. The hydrogen chains are model systems that we construct. The intended use is the preparation of initial states for the quantum simulation of molecules, with possible benefits for chemistry, drug discovery, and materials science. The method predicts the parameters of circuits that prepare approximate ground states. It does not design molecules, and it does not predict toxicity or synthesis routes. Thus, we see no misuse risk beyond that of standard quantum-chemistry software.

Reproducibility statement

Section 3.1 defines the learning problem and the overlap loss. Section 3.2 and App. C give the model input and the encoder. Section 4.2 and App. C specify the three circuit-prediction heads. Appendix B defines the brickwall Ansatz completely, with its gate order and the sign convention of every angle. Appendix D gives the settings of the target-aware benchmark. Appendix A describes how we build the active spaces, the deformed geometries, and the CASCI and DMRG targets. Appendix H and Tab. 7 list the model sizes, optimizer settings, schedules, and seeds. Tables 1, 2, and 8 give the numbers behind the result figures. On acceptance, we will release the code for the model, training, evaluation, benchmark, and T-cost estimates in a public repository. We will also release the three molecular datasets and the hydrogen-chain targets from the public QM7 and QM9 sets.

References

Dorit Aharonov, Wim van Dam, Julia Kempe, Zeph Landau, Seth Lloyd, and Oded Regev. Adiabatic quantum computation is equivalent to standard quantum computation. *SIAM Journal on Computing*, 37(1):166–194, 2007. doi: 10.1137/S0097539705447323.

Alán Aspuru-Guzik, Anthony D. Dutoi, Peter J. Love, and Martin Head-Gordon. Simulated quantum computation of molecular energies. *Science*, 309(5741):1704–1707, 2005. doi: 10.1126/science.1113479.

G. Barcza, Ö. Legeza, K. H. Marti, and M. Reiher. Quantum-information analysis of electronic states at different molecular structures. *Physical Review A*, 83(1):012508, 2011. doi: 10.1103/PhysRevA.83.012508.

Albert P. Bartók, Mike C. Payne, Risi Kondor, and Gábor Csányi. Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. *Physical Review Letters*, 104 (13):136403, 2010. doi: 10.1103/PhysRevLett.104.136403.

Ilyes Batatia, Dávid Péter Kovács, Gregor N. C. Simm, Christoph Ortner, and Gábor Csányi. MACE: Higher order equivariant message passing neural networks for fast and accurate force fields. In

Advances in Neural Information Processing Systems, volume 35, pp. 11423–11436. Curran Associates, Inc., 2022. doi: 10.52202/068431-0830.

Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P. Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess E. Smidt, and Boris Kozinsky. $E(3)$-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature Communications*, 13(1): 2453, 2022. doi: 10.1038/s41467-022-29939-5.

Jörg Behler and Michele Parrinello. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Physical Review Letters*, 98(14):146401, 2007. doi: 10.1103/PhysRevLett.98.146401.

Davide Bincoletto, Korbinian Stein, Jonas Motyl, and Jakob S. Kottmann. A transferable machine learning approach to predict quantum circuit parameters for electronic structure problems. *Machine Learning: Science and Technology*, 7(3):035055, 2026. doi: 10.1088/2632-2153/ae7532.

Lorenz C. Blum and Jean-Louis Reymond. 970 million druglike small molecules for virtual screening in the chemical universe database GDB-13. *Journal of the American Chemical Society*, 131 (25):8732–8733, 2009. doi: 10.1021/ja902302h.

S. F. Boys. Construction of some molecular orbitals to be approximately invariant for changes from one molecule to another. *Reviews of Modern Physics*, 32(2):296–299, 1960. doi: 10.1103/RevModPhys.32.296.

Garnet Kin-Lic Chan and Martin Head-Gordon. Highly correlated calculations with a polynomial cost algorithm: A study of the density matrix renormalization group. *The Journal of Chemical Physics*, 116(11):4462–4476, 2002. doi: 10.1063/1.1449459.

Jianpeng Chen, Zirui Sheng, Cunxi Gong, and Weitang Li. Automated unitary coupled cluster circuit design via differentiable quantum architecture search. arXiv preprint arXiv:2605.28049, 2026. URL https://arxiv.org/abs/2605.28049.

Edward Farhi, Jeffrey Goldstone, Sam Gutmann, and Michael Sipser. Quantum computation by adiabatic evolution. arXiv preprint arXiv:quant-ph/0001106, 2000. URL https://arxiv.org/abs/quant-ph/0001106.

Stepan Fomichev, Kasra Hejazi, Modjtaba Shokrian Zini, Matthew Kiser, Joana Fraxanet, Pablo Antonio Moreno Casares, Alain Delgado, Joonsuk Huh, Arne-Christian Voigt, Jonathan E. Mueller, and Juan Miguel Arrazola. Initial state preparation for quantum chemistry on quantum computers. *PRX Quantum*, 5(4):040339, 2024. doi: 10.1103/PRXQuantum.5.040339.

Mario Geiger and Tess Smidt. e3nn: Euclidean neural networks. arXiv preprint arXiv:2207.09453, 2022. URL https://arxiv.org/abs/2207.09453.

Gabriel Greene-Diniz, Georgia Prokopiou, David Zsolt Manrique, and David Muñoz Ramo. Quantum state preparation of multiconfigurational states for quantum chemistry. *Digital Discovery*, 5 (1):134–152, 2026. doi: 10.1039/D5DD00350D.

Reza Haghshenas, Johnnie Gray, Andrew C. Potter, and Garnet Kin-Lic Chan. Variational power of quantum circuit tensor networks. *Physical Review X*, 12(1):011047, 2022. doi: 10.1103/PhysRevX.12.011047.

Pascual Jordan and Eugene Wigner. Über das Paulische Äquivalenzverbot. *Zeitschrift für Physik*, 47(9):631–651, 1928. doi: 10.1007/BF01331938.

Amir Karton. Nondynamical correlation effects in multidimensional potential energy curves of prototypical hydrocarbons. *Chemical Physics*, 589:112498, 2025. doi: 10.1016/j.chemphys.2024.112498.

Kimberlee Keithley, Shunsuke Yamamoto, Ryota Kenmoku, Ikko Hamamura, Kouhei Nakaji, Shu Kanno, Takao Kobayashi, Qi Gao, Shumpei Uno, Kohei Oshio, Naoki Watanabe, Takeshi Sato, Naoki Yamamoto, Shunya Minami, Yohichi Suzuki, Yuma Nakamura, Jorge A. Campos-Gonzalez-Angulo, Mohammad Ghazi Vakili, and Alan Aspuru-Guzik. Auger spectroscopy via generative quantum eigensolver: A quantum approach to molecular excitations. arXiv preprint arXiv:2603.12859, 2026. URL https://arxiv.org/abs/2603.12859.

Ryota Kemmoku, Qi Gao, Shu Kanno, Kimberlee Keithley, Ikko Hamamura, Naoki Yamamoto, and Kouhei Nakaji. Generative circuit design for quantum-selected configuration interaction. arXiv preprint arXiv:2604.09756, 2026. URL https://arxiv.org/abs/2604.09756.

Alexei Yu. Kitaev. Quantum measurements and the Abelian stabilizer problem. Technical Report TR96-003, Electronic Colloquium on Computational Complexity (ECCC), 1996. URL https://eccc.weizmann.ac.il/report/1996/003/.

Vadym Kliuchnikov, Kristin Lauter, Romy Minko, Adam Paetznick, and Christophe Petit. Shorter quantum circuits via single-qubit gate approximation. *Quantum*, 7:1208, 2023. doi: 10.22331/q-2023-12-18-1208.

Gerald Knizia. Intrinsic atomic orbitals: An unbiased bridge between quantum theory and chemical concepts. *Journal of Chemical Theory and Computation*, 9(11):4834–4843, 2013. doi: 10.1021/ct400687b.

Alex Koziell-Pipe, Jasmine Brewer, Jem Guhit, Marwa H. Farag, Kripa Panchagnula, Gabriel Laude, Fabian Finger, Carlo Gaggioli, Ludmila Szulakowska, Oliver J. Backhouse, Christos Papalitsas, Jason G. Mustakis, Thomas Soini, David Munoz Ramo, Stephen Clark, Elica Kyoseva, and Enrico Rinaldi. Learning to prepare molecular ground states with transformer models. arXiv preprint arXiv:2607.22468, 2026. URL https://arxiv.org/abs/2607.22468.

Maureen Krumtünger, Alissa Wilms, Paul K. Faehrmann, Jens Eisert, Jakob Kottmann, Paolo Andrea Erdman, and Sumeet Khatri. Reinforcement learning of quantum circuit architectures for molecular potential energy curves. arXiv preprint arXiv:2511.16559, 2025. URL https://arxiv.org/abs/2511.16559.

Akash Kundu and Leopoldo Sarra. Reinforcement learning with learned gadgets to tackle hard quantum problems on real hardware. *Communications Physics*, 9(1):44, 2026. doi: 10.1038/s42005-025-02475-6.

Seunghoon Lee, Joonho Lee, Huanchen Zhai, Yu Tong, Alexander M. Dalzell, Ashutosh Kumar, Phillip Helms, Johnnie Gray, Zhi-Hao Cui, Wenyuan Liu, Michael Kastoryano, Ryan Babbush, John Preskill, David R. Reichman, Earl T. Campbell, Edward F. Valeev, Lin Lin, and Garnet Kin-Lic Chan. Evaluating the evidence for exponential quantum advantage in ground-state quantum chemistry. *Nature Communications*, 14(1):1952, 2023. doi: 10.1038/s41467-023-37587-6.

Örs Legeza, Andor Menczer, Miklós Antal Werner, Sotiris S. Xantheas, Frank Neese, Martin Ganahl, Cole Brower, Samuel Rodriguez Bernabeu, Jeff Hammond, and John Gunnels. Hunting for quantum advantage in electronic structure calculations is a highly non-trivial task. arXiv preprint arXiv:2603.28648, 2026. URL https://arxiv.org/abs/2603.28648.

Shanchuan Li, Daisuke Tsukayama, Jun-ichi Shirakashi, Tetsuo Shibuya, and Hiroshi Imai. Quantum architecture search with neural predictor based on ZX-calculus. *EPJ Quantum Technology*, 12(1):106, 2025. doi: 10.1140/epjqt/s40507-025-00410-w.

Daniel Malz, Georgios Styliaris, Zhi-Yuan Wei, and J. Ignacio Cirac. Preparation of matrix product states with log-depth quantum circuits. *Physical Review Letters*, 132(4):040404, 2024. doi: 10.1103/PhysRevLett.132.040404.

Daniel Marti-Dafcik, Hugh G. A. Burton, and David P. Tew. Spin coupling is all you need: Encoding strong electron correlation in molecules on quantum computers. *Physical Review Research*, 7(1):013191, 2025. doi: 10.1103/PhysRevResearch.7.013191.

Koen Mesman, Yinglu Tang, Matthias Moller, Boyang Chen, and Sebastian Feld. NN-AE-VQE: Neural network parameter prediction on autoencoded variational quantum eigensolvers. arXiv preprint arXiv:2411.15667, 2025. URL https://arxiv.org/abs/2411.15667.

Kouhei Nakaji, Lasse Bjørn Kristensen, Ryota Kemmoku, Jorge A. Campos-Gonzalez-Angulo, Mohammad Ghazi Vakili, Haozhe Huang, Mohsen Bagherimehrab, Christoph Gorgulla, FuTe Wong, Alex McCaskey, Jin-Sung Kim, Thien Nguyen, Pooja Rao, Qi Gao, Michihiko Sugawara, Naoki Yamamoto, and Alán Aspuru-Guzik. The generative quantum eigensolver (GQE) and its application for ground state search. arXiv preprint arXiv:2401.09253, 2024. URL https://arxiv.org/abs/2401.09253.

Roberto Olivares-Amaya, Weifeng Hu, Naoki Nakatani, Sandeep Sharma, Jun Yang, and Garnet Kin-Lic Chan. The ab-initio density matrix renormalization group in practice. *The Journal of Chemical Physics*, 142(3):034102, 2015. doi: 10.1063/1.4905329.

Pauline J. Ollitrault, Cristian L. Cortes, Jérôme F. Gonthier, Robert M. Parrish, Dario Rocca, Gian-Luca Anselmetti, Matthias Degroote, Nikolaj Moll, Raffaele Santagati, and Michael Streif. Enhancing initial state overlap through orbital optimization for faster molecular electronic ground-state energy estimation. *Physical Review Letters*, 133(25):250601, 2024. doi: 10.1103/PhysRevLett.133.250601.

Mateusz Ostaszewski, Lea M. Trenkwalder, Wojciech Masarczyk, Eleanor Scerri, and Vedran Dunjko. Reinforcement learning for optimization of variational quantum circuit architectures. In *Advances in Neural Information Processing Systems*, volume 34, pp. 18182–18194. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper/2021/hash/9724412729185d53a2e3e7f889d9f057-Abstract.html.

Yash J. Patel, Akash Kundu, Mateusz Ostaszewski, Xavier Bonet-Monroig, Vedran Dunjko, and Onur Danaci. Curriculum reinforcement learning for quantum architecture search under hardware errors. In *International Conference on Learning Representations*, pp. 32298–32329, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/88cc8602d5e5a913e1c2a2142ca7377b-Abstract-Conference.html.

Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J. Love, Alán Aspuru-Guzik, and Jeremy L. O'Brien. A variational eigenvalue solver on a photonic quantum processor. *Nature Communications*, 5(1):4213, 2014. doi: 10.1038/ncomms5213.

János Pipek and Paul G. Mezey. A fast intrinsic localization procedure applicable for ab initio and semiempirical linear combination of atomic orbital wave functions. *The Journal of Chemical Physics*, 90(9):4916–4926, 1989. doi: 10.1063/1.456588.

Raghunathan Ramakrishnan, Pavlo O. Dral, Matthias Rupp, and O. Anatole von Lilienfeld. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific Data*, 1(1):140022, 2014. doi: 10.1038/sdata.2014.22.

Mafalda Ramôa, Panagiotis G. Anastasiou, Luis Paulo Santos, Nicholas J. Mayhall, Edwin Barnes, and Sophia E. Economou. Reducing the resources required by ADAPT-VQE using coupled exchange operators and improved subroutines. *npj Quantum Information*, 11(1):86, 2025. doi: 10.1038/s41534-025-01039-4.

Shi-Ju Ran. Encoding of matrix product states into quantum circuits of one- and two-qubit gates. *Physical Review A*, 101(3):032310, 2020. doi: 10.1103/PhysRevA.101.032310.

Jonathan Romero, Ryan Babbush, Jarrod R. McClean, Cornelius Hempel, Peter J. Love, and Alán Aspuru-Guzik. Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz. *Quantum Science and Technology*, 4(1):014008, 2019. doi: 10.1088/2058-9565/aad3e4.

Lars Ruddigkeit, Ruud van Deursen, Lorenz C. Blum, and Jean-Louis Reymond. Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17. *Journal of Chemical Information and Modeling*, 52(11):2864–2875, 2012. doi: 10.1021/ci300415d.

Manuel S. Rudolph, Jing Chen, Jacob Miller, Atithi Acharya, and Alejandro Perdomo-Ortiz. Decomposition of matrix product states into shallow quantum circuits. *Quantum Science and Technology*, 9(1):015012, 2024. doi: 10.1088/2058-9565/ad04e6.

Matthias Rupp, Alexandre Tkatchenko, Klaus-Robert Müller, and O. Anatole von Lilienfeld. Fast and accurate modeling of molecular atomization energies with machine learning. *Physical Review Letters*, 108(5):058301, 2012. doi: 10.1103/PhysRevLett.108.058301.

Felix Rupprecht and Sabine Wölk. Sparse quantum state preparation with improved Toffoli cost. *Quantum*, 10:2208, 2026. doi: 10.22331/q-2026-09-10-2208.

Felix Rupprecht and Sabine Wölk. Faster matrix product state preparation by exploiting symmetry-induced block-sparsity. arXiv preprint arXiv:2605.28489, 2026. URL https://arxiv.org/abs/2605.28489.

Elvira R. Sayfutyarova, Qiming Sun, Garnet Kin-Lic Chan, and Gerald Knizia. Automated construction of molecular active spaces from atomic valence orbitals. *Journal of Chemical Theory and Computation*, 13(9):4063–4078, 2017. doi: 10.1021/acs.jctc.7b00128.

Sandeep Sharma and Garnet Kin-Lic Chan. Spin-adapted density matrix renormalization group algorithms for quantum chemistry. *The Journal of Chemical Physics*, 136(12):124121, 2012. doi: 10.1063/1.3695642.

Kenji Sugisaki, Kazuo Toyota, Kazunobu Sato, Daisuke Shiomi, and Takeji Takui. Adiabatic state preparation of correlated wave functions with nonlinear scheduling functions and broken-symmetry wave functions. *Communications Chemistry*, 5(1):84, 2022. doi: 10.1038/s42004-022-00701-8.

Yanxian Tao, Xiongzhi Zeng, Yi Fan, Jie Liu, Zhenyu Li, and Jinlong Yang. Exploring accurate potential energy surfaces via integrating variational quantum eigensolver with machine learning. *The Journal of Physical Chemistry Letters*, 13(28):6420–6426, 2022. doi: 10.1021/acs.jpclett.2c01738.

Luca Thiede, Abdulrahman Aldossary, Andreas Burger, Jorge Arturo Campos-Gonzalez-Angulo, Ning Wang, Alexander Zook, Melisa Alkan, Kouhei Nakaji, Taylor Lee Patti, Jérôme Florian Gonthier, Mohammad Ghazi Vakili, and Alán Aspuru-Guzik. Coupled cluster con MōLe: Molecular orbital learning for neural wavefunctions. arXiv preprint arXiv:2602.20232, 2026. URL https://arxiv.org/abs/2602.20232.

Norm M. Tubman, Carlos Mejuto-Zaera, Jeffrey M. Epstein, Diptarka Hait, Daniel S. Levine, William Huggins, Zhang Jiang, Jarrod R. McClean, Ryan Babbush, Martin Head-Gordon, and K. Birgitta Whaley. Postponing the orthogonality catastrophe: efficient state preparation for electronic structure simulations on quantum devices. arXiv preprint arXiv:1809.05523, 2018. URL https://arxiv.org/abs/1809.05523.

Javad Vahedi and Hadi H. Arefi. Graph neural networks for fast operator selection in adaptive VQE. arXiv preprint arXiv:2606.08794, 2026. URL https://arxiv.org/abs/2606.08794.

Steven R. White. Density matrix formulation for quantum renormalization groups. *Physical Review Letters*, 69(19):2863–2866, 1992. doi: 10.1103/PhysRevLett.69.2863.

James D. Whitfield, Jacob Biamonte, and Alán Aspuru-Guzik. Simulation of electronic structure Hamiltonians using quantum computers. *Molecular Physics*, 109(5):735–750, 2011. doi: 10.1080/00268976.2011.552441.

Yordan S. Yordanov, V. Armaos, Crispin H. W. Barnes, and David R. M. Arvidsson-Shukur. Qubit-excitation-based adaptive variational quantum eigensolver. *Communications Physics*, 4(1):228, 2021. doi: 10.1038/s42005-021-00730-0.

Shi-Xin Zhang, Chang-Yu Hsieh, Shengyu Zhang, and Hong Yao. Neural predictor based quantum architecture search. *Machine Learning: Science and Technology*, 2(4):045027, 2021. doi: 10.1088/2632-2153/ac28dd.

Shiwei Zhang and Henry Krakauer. Quantum Monte Carlo method using phase-free random walks with Slater determinants. *Physical Review Letters*, 90(13):136401, 2003. doi: 10.1103/PhysRevLett.90.136401.

Hang Zou, Martin Rahm, Anton Frisk Kockum, and Simon Olsson. Generative flow-based warm start of the variational quantum eigensolver. *npj Quantum Information*, 12(1):5, 2025. doi: 10.1038/s41534-025-01159-x.

A Dataset generation

A.1 State vector target states

For the QM7-StateVec-30k dataset construction, the target active space sizes are limited to 20 qubits, since the state vector simulations are expensive and training requires the overlap evaluation to be fast enough. In addition, we would like the dataset to contain a large number of molecules of the same family, and for the systems to be strongly correlated. Indeed, strongly correlated systems would be the primary targets for quantum computing applications.

Therefore, we chose to start from the QM7 dataset of organic molecules (Blum & Reymond, 2009; Rupp et al., 2012). We then stretched all the bonds in each molecule by factors of 1.2, 1.4, 1.6, 1.8 and 2.0 to introduce strong correlation. Given the size of these molecules, the 20 qubit limitation implied that we have to define an *active space*, a subset of orbitals to treat on the quantum computer. Moreover, this choice should be done in a consistent way across molecules, be automated so it can scale, and result in a set of localized orbitals that facilitate transferability.

Although the AVAS method (Sayfutyarova et al., 2017) is automated, in our tests it did not result in well-localized orbitals, especially in the small active space sizes chosen. Therefore, we defined our own procedure for active space selection and construction. First of all, we identify an arbitrary bond near the center of the molecule that will be the focus of our active space. We also define a target active space size N_{act} of either 16, 18 or 20 qubits. Then, atoms are selected starting from the stretched bond and extending outwards until the number of electrons and expected valence virtuals is high enough. Once these atoms are selected, active space construction proceeds as follows:

- Run a closed-shell RHF/STO-3G calculation. Perform stability analysis and follow any imaginary modes until the calculation converges to a local minima as verified by second derivatives.
- Localize the orbitals using the IBO method (Knizia, 2013), separately for occupied and virtuals
- Compute the IAO population of the occupied orbitals on the selected atoms.
- Select the $N_{act}/2$ spin-orbitals with the highest populations. If there is a tie (10^{-3} threshold), also include the tied spin-orbitals
- Rank the occupied orbitals by increasing pseudo-energy (expectation value of the Fock operator)
- Assign each virtual orbital to the occupied orbital with which it has the highest exchange integral
- For each occupied orbital in the active space, starting with the highest energy and going down:
  1. Include the virtual assigned to it with the largest exchange integral
  2. If the next virtuals' exchange integrals are within 10^{-3} Ha, also include them

  until at least $N_{act}/2$ virtual spin-orbitals have been included.

The virtual orbitals are then stored in that order. There are several points to note:

1. Since we include orbitals when ties happen, the final active space may be larger than the target. In practice, we generate active spaces for slightly more data points than we need and post-select. The advantage is that active spaces are chosen more consistently, and degenerate orbital are not arbitrarily split between the active and inactive space.
2. Although we specify a fixed ordering for the occupied and virtual active orbitals, this is immaterial for the chosen MōLe encoder architecture since it is equivariant under orbital permutation.

Once the active space is selected, we need to obtain the corresponding target state vector. In these highly stretched molecules, the true ground state is not always a singlet, but to avoid introducing more complexity in the model and the dataset, we aimed to obtain the lowest singlet ground state. In PySCF, CASCI calculations are performed in the determinant basis and only enforce S_z and not

S^2, meaning higher spin states than singlets can be obtained. To ensure the calculation converged on the intended state, we added a penalty term (as implemented in PySCF) to enforce $S^2 = 0$, and we also requested convergence of 3 roots as the Davidson procedure can sometimes select the wrong eigentstate.

After the dataset generation, we discovered that for highly symmetric molecule, an excited singlet state could still be obtained if the initial guess was in the wrong irreducible representation. This can be alleviated by adding a small amount of noise to the initial guess. The problematic data point was removed and we verified that remaining points included in QM7-StateVec-30k were unaffected.

A.2 DMRG target states

Three datasets carry DMRG targets, and every DMRG optimization below uses an energy convergence tolerance of 10^{-4}. Two of them place the orbitals along the MPS chain in Fiedler order, the ordering obtained from the orbital-interaction matrix $\widetilde{K}$ of App. C by forming the graph Laplacian $L = D - \widetilde{K}$ with $D_{ii} = \sum_j \widetilde{K}_{ij}$ and sorting the entries of its second-lowest eigenvector, the Fiedler vector (Barcza et al., 2011; Olivares-Amaya et al., 2015).

For the hydrogen chain dataset construction, the geometries are equally spaced linear STO-3G chains at a nearest-neighbor separation of 2.5 Å, with every MO kept in the active space and no deformation applied: the separation alone makes them strongly correlated. Their targets are optimized at bond dimension 500 in a fixed chain ordering, each localized bonding orbital followed by its antibonding partner. Fiedler ordering is not used here because it is degenerate on a chain: the two orbitals of one bond couple to the rest almost identically, so their Fiedler components fall within 10^{-5} of each other and their relative order is decided by the last bits of the SCF solution, which is not reproducible past about thirty sites; since an MPS is stored in site order, a dataset whose ordering is not reproducible cannot read its own states back.

For the QM7-MPS-7k dataset construction, a molecule at its equilibrium geometry is usually dominated by its Hartree–Fock determinant and is therefore trivial for any Ansatz, so the geometries are deformed: each is uniformly stretched and swept until the Hartree–Fock determinant stops dominating the DMRG ground state. The sweep targeted a Hartree–Fock overlap amplitude between 0.20 and 0.50, a squared overlap of 0.04 to 0.25, which 5,173 of the 7,165 molecules reached; for each remaining molecule we retain the closest successfully swept geometry rather than dropping it, so the benchmark is not filtered by how easy its Hartree–Fock reference happens to be. The targets are optimized at bond dimension 250 in Fiedler orbital ordering.

For the 25,877-state QM7 training set construction, the sweep is the same one; what changes is how much of it is kept. QM7-MPS-7k retains one geometry per molecule and discards the other points the sweep had already computed, whereas this set keeps every swept point of the same fitting molecules with a Hartree–Fock overlap amplitude of at least 0.20, a squared overlap of at least 0.04. That gives 25,877 states in the same active spaces of 16 to 88 qubits and at the same bond dimension 250 in Fiedler ordering. The retained uniform stretch runs from 1.00 to 2.24 with a median of 1.20.

For the QM9-MPS-100 dataset construction, the same procedure is applied to QM9, with only the selection changed so that every test active space exceeds every training one. Molecules are restricted to C, H, N and O and to at least 90 qubits under the same frozen-core STO-3G counting, which admits only the eight- and nine-heavy-atom end of QM9. Twenty formulas survive that filter, and the hundred test molecules are four to six isomers of each, so that no single composition dominates the sample, spanning 90 to 112 qubits against QM7-MPS-7k's ceiling of 88. Their geometries are swept the same way and their targets optimized at the same bond dimension 250 in Fiedler ordering, and the retained uniform stretch runs from 1.09 to 1.40 with a median of 1.28. Three of the hundred land inside the target window and the remaining ninety-seven are the nearest swept points outside it, all on the high side, so the realized squared overlap spans 0.102 to 0.641 with a median of 0.354.

B Brickwall Ansatz

This appendix gives the complete mathematical definition of the paper-specific brickwall family summarized in Sec. 4.1, including its generator ordering, locality, and conserved quantities.

We apply the circuit illustrated in Fig. 3 to the Hartree–Fock product state and arrange the localized spatial orbitals in the same one-dimensional order as the MPS. Each spatial orbital is one site carrying the four-valued occupation index of Sec. 3.3, written $|0\rangle$, $|\alpha\rangle$, $|\beta\rangle$ and $|\alpha\beta\rangle$ for empty, one spin-up electron, one spin-down electron, and double occupation, and indexed in that order; a two-site label $|a, b\rangle$ puts site i first. The symbol $|0\rangle$ therefore denotes an empty spatial orbital, not the initial state of every site. The circuit consists only of two-site gates between neighboring spatial orbitals. This matching between the circuit geometry and the MPS ordering allows the circuit and target state to be contracted as a one-dimensional tensor-network strip at fixed circuit depth (Haghshenas et al., 2022).

On each neighboring pair (i, j) we use four real rotations, indexed $g = 1, \ldots, 4$: the α- and β-spin Givens rotations, pair hopping, and spin exchange (Fig. 3b). We number the spin orbitals with the α mode of a site before its β mode, so site i occupies modes $2i$ and $2i + 1$ and two adjacent spatial orbitals occupy four contiguous modes. Writing $a^\dagger_{i\sigma}$ for creation in the spin orbital (i, σ) and $n_{i\sigma}$ for its occupation, the bare excitations behind them are

$$\begin{aligned} E_1 &= a^\dagger_{j\alpha} a_{i\alpha}, & E_2 &= a^\dagger_{j\beta} a_{i\beta}, \\ E_3 &= a^\dagger_{j\alpha} a^\dagger_{j\beta} a_{i\beta} a_{i\alpha}, & E_4 &= a^\dagger_{i\beta} a^\dagger_{j\alpha} a_{j\beta} a_{i\alpha}, \end{aligned} \tag{6}$$

and the generators the implementation rotates with are

$$A_1 = (-1)^{n_{i\beta}} \left(E_1 - E_1^\dagger \right), \quad A_2 = (-1)^{n_{j\alpha}} \left(E_2 - E_2^\dagger \right), \quad A_g = E_g - E_g^\dagger \ \ (g = 3, 4). \tag{7}$$

Each rotation is then

$$R_g(\theta_g) = I + (\cos\theta_g - 1) P_g + \sin\theta_g A_g = \exp(\theta_g A_g), \qquad P_g = -A_g^2. \tag{8}$$

Each A_g is real and skew-symmetric, so R_g is a real orthogonal matrix and is the identity at $\theta_g = 0$, and P_g projects onto its support, which is what the closed form requires. The implementation applies generators $g = 1, 2, 3, 4$ in that order, so the bond-gate matrix is $R_4 R_3 R_2 R_1$. Since no A_g connects local basis states of different electron number and spin projection $(N_\mathrm{e}, 2S_z)$, the full circuit conserves particle number and spin-z projection; it does *not* in general conserve S^2.

A fermionic excitation between two sites is in general nonlocal, its Jordan–Wigner string covering every mode in between (Jordan & Wigner, 1928). Here the two sites are adjacent, so their four modes are contiguous and every string stays inside them; each bond gate is therefore exactly a two-site operator, which is what the strip contraction needs.

The gates are arranged in a brickwall pattern. Even layers act on $(0, 1), (2, 3), \ldots$, and odd layers act on $(1, 2), (3, 4), \ldots$. The bonds within one layer are disjoint and can be applied in parallel. Defining

$$U_\ell = \prod_{(i,j) \in \mathcal{B}_\ell} [R_4 R_3 R_2 R_1]_{ij,\ell}, \qquad U_\mathrm{BW} = U_{d-1} \cdots U_1 U_0, \tag{9}$$

where $\mathcal{B}_\ell$ denotes the parity-compatible bonds, gives the depth-d circuit, applied in increasing layer order.

C Model architecture details

C.1 Input representation

Each molecular orbital of Sec. 3.2 is expanded in the atom-centered atomic-orbital basis as

$$\psi_p(\mathbf{r}) = \sum_A \sum_{k \in \mathcal{K}_A} \sum_{\ell \in \mathcal{L}_{A,k}} \sum_{m=-\ell}^{\ell} C_{pA,k\ell m} \phi^m_{k\ell}(\mathbf{r} - \mathbf{R}_A), \tag{10}$$

where p indexes the MO, A indexes atoms, and k, ℓ, m label the radial shell and the angular-momentum channel. The coefficients $C_{pA,k\ell m}$ are entries of the Hartree–Fock coefficient matrix in this basis, so they specify how much the basis function $\phi^m_{k\ell}$ on atom A contributes to the orbital ψ_p.

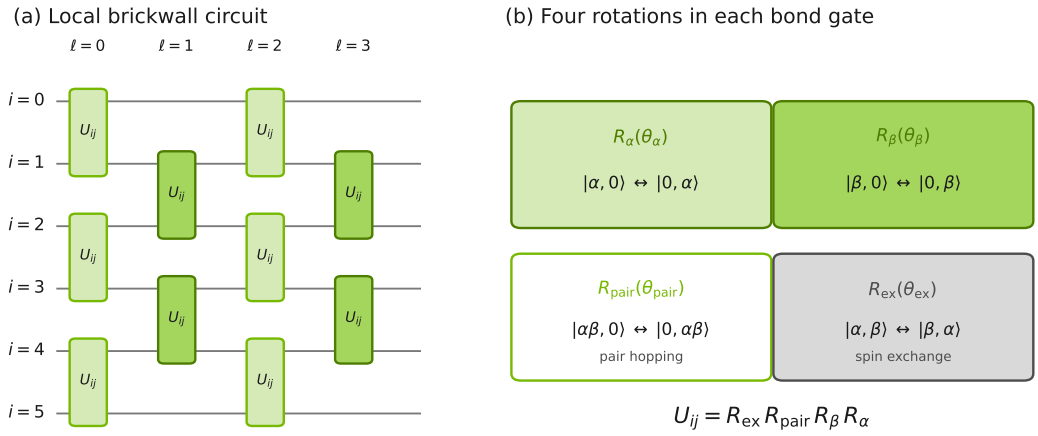


Figure 3: Brickwall Ansatz applied to the Hartree–Fock product state. (a) Neighboring spatial-orbital sites are coupled in alternating even and odd layers; all bond gates within one layer act on disjoint sites and can be executed in parallel. (b) Each bond gate is the ordered product of spin-up and spin-down Givens rotations, pair hopping, and spin exchange. The arrows show the coupled local occupation sectors; for a Givens rotation, the unshown opposite-spin occupations are unchanged spectators.

Different atom types carry different numbers of basis functions, so the raw coefficient blocks do not have the same number of radial channels for every atom. Following MōLe, we zero-pad them to a common set of irreducible-representation channels for each ℓ, producing the standardized input tensor $C^{\text{padded}}_{pA,k\ell m}$. After padding, the index range of (k, ℓ, m) is the same for every atom A, and entries corresponding to basis functions absent on a particular atom are zero.

C.2 Shared encoder and invariances

Following MōLe (Thiede et al., 2026), an equivariant linear layer first projects these padded MO coefficients, after which the encoder alternates two kinds of mixing. *Orbital-wise attention* exchanges information along the MO axis through equivariant query, key, and value maps whose parameters are independent of the number of orbitals. *Geometric message passing* uses MACE interactions and an equivariant product basis to exchange information over the atom graph (Batatia et al., 2022; Geiger & Smidt, 2022). The resulting representation contains one embedding per (orbital, atom) pair.

The model output is not a geometric object: it is a collection of scalar angles indexed by molecular orbitals. It must therefore be invariant, rather than equivariant, to a change of molecular frame or atom ordering. Translation invariance follows from using only relative atomic positions; equivariant message passing, permutation-symmetric aggregation, and scalar head outputs give the corresponding rotation, reflection, and atom-permutation invariances.

A separate input ambiguity comes from the fact that each molecular orbital is defined only up to sign, and the encoder reads the coefficient matrix directly, so the architecture does not enforce invariance to it. Every record therefore carries its coefficients, its integrals, and its target in one and the same orbital basis, the one its source produced, and no rephasing is applied on import.

C.3 Head details

Fixed-order UCCSD and CEO heads. The single-excitation readout of Sec. 4.2 reads an occupied–virtual pair (i, a) with an equivariant projection P followed by a contraction over atoms,

$$s_{ia} = \sum_A \langle P\mathbf{h}_{iA},\, P\mathbf{h}_{aA} \rangle \,, \tag{11}$$

and a residual multilayer perceptron maps the invariant pair feature to the output channels. Before the projection, each orbital embedding is normalized by its own norm over the joint atom and fea-

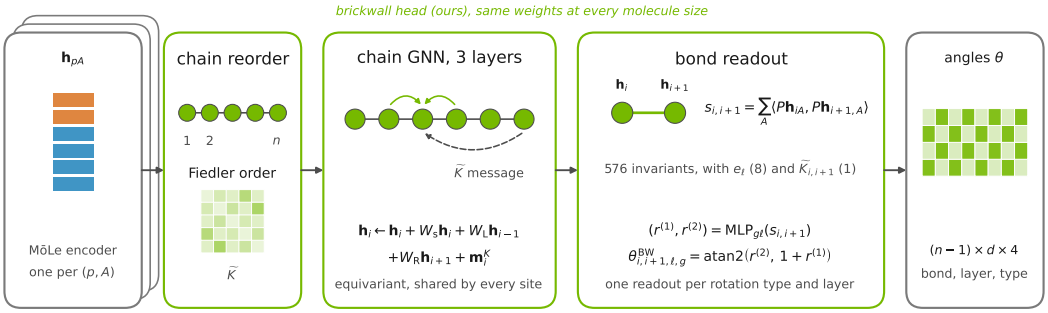


Figure 4: The brickwall head of Sec. 4.2, which turns the inherited encoder output into the angles of Eq. (2): one equivariant embedding per orbital and atom is reordered onto the Fiedler chain, propagated along it by Eq. (3), and read out per bond by Eq. (4) and the circular map of Eq. (5). Channel counts are those of the depth-six molecular models. No learned weight depends on molecule size; only the emitted tensor, of shape $(N-1) \times d \times 4$ for N orbitals and d layers, grows with it.

ture axes, so that the feature measures the alignment of two orbital embeddings rather than their magnitudes.

Double excitations use two independent equivariant occupied–virtual products, whose pair features are normalized the same way over the joint atom and feature axes, contracted into a quadruple feature s_{iajb}, and symmetrized under $(i,a) \leftrightarrow (j,b)$, matching the exchange symmetry of the double-excitation generator itself. For n_o occupied and n_v virtual orbitals the UCCSD head emits

$$\theta^{\mathrm{UCCSD}} \in \mathbb{R}^{2n_o n_v} \oplus \mathbb{R}^{(n_o n_v)^2} \oplus \mathbb{R}^{\binom{n_o}{2}\binom{n_v}{2}} \oplus \mathbb{R}^{\binom{n_o}{2}\binom{n_v}{2}}, \tag{12}$$

the singles first, with the α and β sectors of one pair (i,a) adjacent rather than in separate blocks, then $T_2^{\alpha\beta}$, $T_2^{\alpha\alpha}$ and $T_2^{\beta\beta}$, the same-spin blocks running over $i<j$ and $a<b$. The CEO head reads the same orbital keys and widens the readout to the channels the CEO pool requires, one per single, two A/B channels per mixed-spin double, and six per same-spin quadruple, three pairings times the A/B variants, gathered into the canonical CEO angle vector.

Both contractions are indexed by spatial orbitals while the pools are indexed by spin orbitals, and the gap is closed by a bias learned per spin sector and shared by every orbital tuple. The UCCSD head uses two for the singles and three for the doubles, reading s_{iajb} at the quadruples with $i<j$ and $a<b$ for the same-spin blocks. The CEO head uses two for the singles and two for the same-spin doubles; its mixed block needs none, because it carries a single spin type and the per-channel bias of its own readout already covers it. Two sectors of one orbital tuple therefore differ by a single learned number, and since the biases are zero-initialized along with the output projections, the head starts spin symmetric as well as at the identity.

Brickwall head. In the circular parameterization of Eq. (5) the offset 1 in the cosine channel keeps the zero-initialized readout away from the branch cut of $\mathrm{atan2}$, while gradients still reach θ through the sine channel.

The exchange conditioning uses the orbital-interaction matrix employed for the Fiedler ordering,

$$K_{ij} = |(ij|ji)| + 10^{-7}\left|h_{ij}^{(1)}\right|, \qquad \widetilde{K}_{ij} = \frac{K_{ij}}{\max_{p\neq q} K_{pq}}, \quad \widetilde{K}_{ii} = 0, \tag{13}$$

built from the exchange integrals with a small one-electron tie breaker, and normalized so that its entries are comparable across molecules of different size. The molecular brickwall head consumes it on two channels. The bond readout of Eq. (4) takes the adjacent weight as a further argument, $f_{g\ell}(\mathbf{h}_i, \mathbf{h}_{i+1}, e_\ell, \widetilde{K}_{i,i+1})$, and the chain network of Eq. (3) gains the nonlocal message

$$\mathbf{m}_{iA}^K = \sum_j \frac{\widetilde{K}_{ij}}{\sum_r \widetilde{K}_{ir}} W_K \mathbf{h}_{jA}^{(t)}, \tag{14}$$

added on the right-hand side, which reaches strongly coupled orbitals that the two chain neighbours miss. Neither channel adds a parameter to the Ansatz. The hydrogen-chain models switch both off and so realize Eqs. (3) and (4) as written in the main text.

D THE TARGET-AWARE NON-VARIATIONAL BENCHMARK

Settings. The sequential matrix-product disentangler of Sec. 5.3 splits each four-state spatial-orbital site exactly into interleaved α/β qubit sites. It then extracts successive approximate disentangling layers and compresses the residual MPS with an SVD cutoff of 10^{-12} and a maximum bond dimension $\chi_{\max} = 64$. This $\chi_{\max}$ is a truncation parameter of the benchmark, fixed for every molecule; it is not the bond dimension 250 of the target states. We measure the squared overlap by contracting the compiled circuit with the target MPS.

Truncation. Doubling $\chi_{\max}$ to 128 does not change the benchmark. On twenty molecules sampled from QM9-MPS-100, the molecule-paired difference at four layers is -0.0010 ± 0.0012 (SE), with ten of the twenty molecules improving; at two and three layers it is $+0.0005 \pm 0.0013$ and -0.0006 ± 0.0013. The 95% interval at four layers is $[-0.0035, +0.0016]$, well below the 0.0147 margin in Tab. 6.

Randomness. The compiler completes each isometry to a unitary with an unseeded uniform draw for its null columns, so two compilations of the same molecule differ. The first layer acts on the all-zero state and does not depend on the draw; the two $\chi_{\max}$ runs agree there to every digit. Each later layer is extracted from a residual that depends on the draw. From the paired spread of 0.0055 between the two runs, one compilation varies by at most about 0.0039 per molecule and by 0.0004 in a mean over 100 molecules. The per-molecule spread of the brickwall's paired lead over the benchmark is 0.0248, six times larger. The margins of 0.0147 and 0.0090 in Tab. 6 are an order of magnitude above 0.0004.

Layer count. The main text reports the benchmark at four layers so that Fig. 1 carries one benchmark curve. A reference chosen by where our model stops leading would make the comparison circular, so we report the full sweep here. Table 2 reports one to eight layers, and Fig. 5 plots one, two, four, and six. The benchmark is not converged at eight layers. Its mean gain over Hartree–Fock rises by 0.042 from one layer to two, by 0.011 to three, by 0.004 to four, and then by 0.002, 0.0015, 0.0011, and 0.0009 through eight. Each step is significant in a molecule-paired test, because consecutive layers are scored on the same molecules; eight layers exceed four by 0.0057 with a standard error of 0.0002.

Ceiling. The increments decay geometrically with ratio 0.68, so the extrapolated ceiling is 0.632, against 0.639 for the depth-six brickwall trained on all 25,877 training states. That model leads the eight-layer benchmark, and therefore every shallower one, by 0.0090 (95% CI $[0.004, 0.014]$, $p = 0.0007$; Tab. 6).

Cost per layer. Each added layer costs $58{,}000$ to $64{,}000$ T gates but only about $1{,}000$ to $1{,}800$ stages on a critical path of $29{,}819$ stages at one layer, so the sweep runs almost horizontally in Fig. 5a and almost vertically in Fig. 5b. The brickwall T depth of $3{,}796$ is about an eighth of the one-layer benchmark, and its T count of $108{,}402$ lies between the one- and two-layer benchmark (Tab. 2). At eight layers the benchmark T count is $502{,}423$, more than four times the brickwall, at a lower overlap.

Hydrogen chains. On the chains we compiled the benchmark at four, five, and six layers and compare at six, the depth of the brickwall. At six layers it reaches 0.130 on H_{34} and 0.039 on H_{50}, against brickwall seed means of 0.506 and 0.319 (Tab. 8). The step from five to six layers adds $+0.0022$ on H_{34} and $+0.0009$ on H_{50}. At that rate, about 170 and 310 further layers would close the gaps, each with its T cost. At six layers the benchmark already needs $7.0\times$ and $10.0\times$ the brickwall T depth on those chains, and that depth grows with both the chain length and the layer count. We quote the increment from five to six layers because the four-layer circuits come from a separate compilation and a separate random draw.

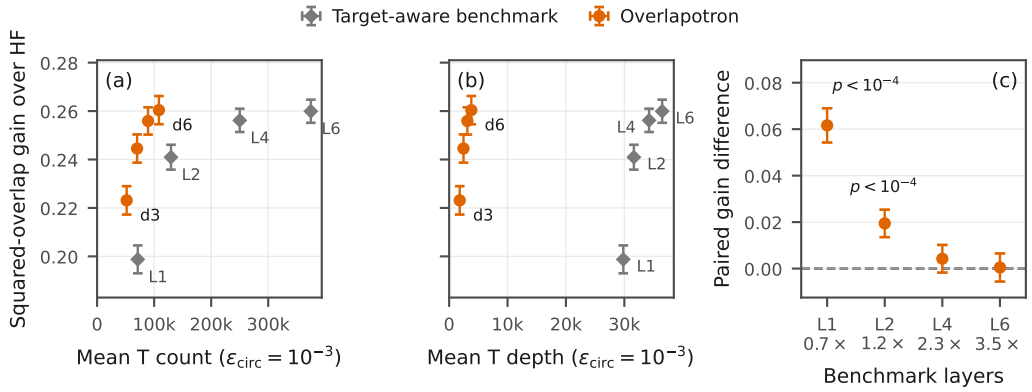


Figure 5: The target-aware benchmark on QM9-MPS-100. (a) T count and (b) T depth against squared overlap for the benchmark at one, two, four, and six layers (L1 to L6) and for the brickwall trained on QM7-MPS-7K at depths three to six (d3 to d6), under one synthesis budget of 10^{-3} per circuit (App. E). Error bars are standard errors over the 100 molecules. (c) Molecule-paired difference between the depth-six brickwall and the benchmark at each layer count, with 95% confidence intervals. The lower tick row gives the benchmark T count relative to the brickwall.

Table 2: Squared overlap, gain over Hartree–Fock, and T costs on QM9-MPS-100 for every benchmark layer count and every brickwall model. The three blocks give the benchmark at one to eight layers, the brickwall at depths three to six, and the depth-six brickwall trained on nested subsets of the QM7 training set. Circuit cost does not depend on the training set, so the costs in the last block repeat. Entries are means over the 100 molecules ± one standard error. The gain over Hartree–Fock has a standard error of about 0.006 in every learned row, smaller than the overlap error, because the Hartree–Fock overlap varies across molecules and moves with the model overlap. T costs use one synthesis budget of 10^{-3} per circuit, and the benchmark costs come from a gate-count formula that bounds the compiled cost from above (App. E). Figure 5 and Fig. 1b–c plot these values.

Method	Squared overlap	Gain over HF	T count	T depth
Hartree–Fock	0.3683 ± 0.0101	0	0	0
Target-aware benchmark, by layer count				
L1	0.5671 ± 0.0133	$+0.1988$	$71{,}308 \pm 534$	$29{,}819 \pm 223$
L2	0.6092 ± 0.0124	$+0.2410$	$129{,}548 \pm 916$	$31{,}632 \pm 218$
L3	0.6204 ± 0.0120	$+0.2521$	$189{,}543 \pm 1{,}414$	$33{,}067 \pm 237$
L4	0.6245 ± 0.0119	$+0.2562$	$250{,}101 \pm 1{,}864$	$34{,}237 \pm 240$
L5	0.6267 ± 0.0119	$+0.2584$	$312{,}389 \pm 2{,}326$	$35{,}427 \pm 244$
L6	0.6282 ± 0.0119	$+0.2599$	$374{,}771 \pm 2{,}811$	$36{,}477 \pm 249$
L7	0.6293 ± 0.0119	$+0.2610$	$438{,}486 \pm 3{,}098$	$37{,}553 \pm 235$
L8	0.6302 ± 0.0119	$+0.2619$	$502{,}423 \pm 3{,}734$	$38{,}562 \pm 252$
Brickwall, by circuit depth, at 6,449 training states				
d3	0.5914 ± 0.0126	$+0.2231$	$51{,}849 \pm 400$	$1{,}808 \pm 1$
d4	0.6128 ± 0.0124	$+0.2445$	$70{,}158 \pm 537$	$2{,}456 \pm 2$
d5	0.6242 ± 0.0121	$+0.2559$	$89{,}350 \pm 688$	$3{,}120.5 \pm 2.5$
d6	0.6287 ± 0.0121	$+0.2604$	$108{,}402 \pm 824$	$3{,}796 \pm 3$
Brickwall at depth 6, by training states				
100	0.5513 ± 0.0128	$+0.1830$	$108{,}402 \pm 824$	$3{,}796 \pm 3$
500	0.5908 ± 0.0125	$+0.2225$	$108{,}402 \pm 824$	$3{,}796 \pm 3$
1,500	0.6092 ± 0.0122	$+0.2410$	$108{,}402 \pm 824$	$3{,}796 \pm 3$
3,000	0.6165 ± 0.0123	$+0.2482$	$108{,}402 \pm 824$	$3{,}796 \pm 3$
25,877	0.6392 ± 0.0123	$+0.2709$	$108{,}402 \pm 824$	$3{,}796 \pm 3$

E RESOURCE ESTIMATION

All T costs in the paper use one Clifford+T model. The model assumes all-to-all connectivity. It excludes routing, error correction, magic-state factory limits, and measurement repetitions.

Synthesis budget. Each circuit receives one synthesis budget $\epsilon_{\mathrm{circ}} = 10^{-3}$, shared equally over its N_{arb} arbitrary rotations, so each rotation is synthesized to precision $\delta = \epsilon_{\mathrm{circ}}/N_{\mathrm{arb}}$. A rotation costs

$$t(\delta) = \mathrm{round}(3.02 \log_2(1/\delta) + 0.77) \tag{15}$$

T gates, an empirical fit to rotation-synthesis costs rather than a certified bound. This fit was obtained from the PyLIQTR source code (`https://doi.org/10.5281/zenodo.18154991`), where it is mentioned it was obtained by decomposing 1000 random angles to varying precisions. The T count is $N_{\mathrm{arb}}\, t(\delta)$. The T depth is the number of serial rotation stages times $t(\delta)$. A scheduler finds the stages: it keeps one clock per qubit, and each CNOT synchronizes only its two qubits. Because $t(\delta)$ grows with $\log N_{\mathrm{arb}}$, the brickwall T depth grows slowly with molecule size, and its standard error in Tab. 2 is not zero.

Brickwall bond gate. In the qubit encoding, the two Givens generators of Eq. (9) are sums of two Pauli words each, on disjoint qubits, so they run in parallel. The pair-hopping and spin-exchange generators are sums of the same eight commuting weight-four Pauli words, with orthogonal coefficient vectors. Therefore $e^{\theta_3 A_3} e^{\theta_4 A_4} = e^{\theta_3 A_3 + \theta_4 A_4}$ exactly, and the two rotations compile together into eight Pauli rotations. One bond gate thus costs 12 arbitrary rotations in 10 serial stages, where a separate compilation of the four rotations costs 20 rotations in 18 stages. We checked the fused compilation numerically. Over 1,000 random angle sets, the largest operator-norm difference between the fused and separate bond gates is 1.4×10^{-14}. Over twelve whole circuits of 8 to 10 orbitals at depths three to six, the largest difference between the prepared states is 3×10^{-13}. Every reported brickwall cost uses the fused compilation.

UCCSD and CEO. Both families compile to eight Pauli rotations on each four-qubit support that carries a double excitation. The additional CEO channels of App. C change only the coefficients of these fixed rotations, not their number. The two families therefore have equal T counts. CEO has the lower T depth because its short excitation circuits allow more parallel work.

Target-aware benchmark. On QM7-STATEVEC-30K, we transpile each compiled benchmark circuit at Qiskit optimization level 3 to the gate set $\{\mathrm{CX}, R_Z, \sqrt{X}, X\}$, synthesize each maximal run of one-qubit gates as one rotation (see below), and cost the result. On QM9-MPS-100 and the hydrogen chains, no compiled circuits were saved, so we cost the benchmark with the gate-count formula

$$N_{\mathrm{arb}} = (12q - 9) + (L - 1)(9q - 9), \qquad \text{stages} = 5q - 2 + 10(L - 1), \tag{16}$$

for $q = 2N$ qubits and L layers. The formula overstates the compiled cost. On the 300 QM7-STATEVEC-30K records, the compiled circuits have 0.67, 0.87, 0.91, and 0.93 of the formula's arbitrary rotations at one to four layers, and 0.82 to 0.97 of its stages. At four layers, the compiled scale would raise the brickwall-to-benchmark T-count ratio on QM9-MPS-100 from 0.43 to about 0.47. The T-depth ratio would change by a similar small amount.

Benchmark with direct one-qubit synthesis. The QM9-MPS-100 and hydrogen-chain benchmark costs above charge each generic one-qubit unitary as three R_Z rotations. A compiler can instead synthesize each maximal run of one-qubit gates as one arbitrary rotation (Kliuchnikov et al., 2023). We treat that run as costing one R_Z and a run that is exactly Clifford as free. This is the most favorable benchmark cost under our budget, not a certified count. The same merge leaves the fused brickwall unchanged, because each of its rotations already stands alone between two CNOTs. Table 3 compares both benchmark costs. On QM7-STATEVEC-30K they come from recompiled circuits. On QM9-MPS-100 and the hydrogen chains they come from the closed forms $N_{\mathrm{arb}} = (6q-5)+5(q-1)(L-1)$ and $3q-2+6(L-1)$ stages, which match 125 of 128 generic test circuits exactly. With direct synthesis, the benchmark at four layers needs about $1.2\times$ the brickwall T count on QM9-MPS-100, not $2.3\times$. Its T depth stays about $5\times$ larger. At one and two layers, it needs fewer T gates than the depth-six brickwall, but it also reaches a lower overlap (0.567 and 0.609, against 0.629 for the depth-six brickwall trained on 6,449 states).

Table 3: Benchmark T costs with each one-qubit unitary charged as three R_Z rotations or synthesized as one rotation. The main text uses the second column on QM7-StateVec-30k and the first on QM9-MPS-100. All entries are means over the molecules or chains, under one synthesis budget of 10^{-3} per circuit. The brickwall costs do not change.

		Three R_Z per unitary		One rotation per unitary	
Targets	Circuit	T count	T depth	T count	T depth
QM7-StateVec-30k	benchmark, L4	36,696	6,802	18,807	3,880
	brickwall, depth 4	10,442	2,148	10,442	2,148
QM9-MPS-100	benchmark, L1	71,308	29,819	33,889	16,974
	benchmark, L4	250,101	34,237	129,420	19,696
	benchmark, L8	502,423	38,562	264,446	22,228
	brickwall, depth 6	108,402	3,796	108,402	3,796
H_{34}–H_{50}	benchmark, L6	321,421	31,770	167,508	18,219
	brickwall, depth 6	92,484	3,756	92,484	3,756

F Why the brickwall angles need a new head

The brickwall head of Sec. 4.2 adds three components to the inherited pair readout of Eq. (11): the chain propagation of Eq. (3), the orbital-interaction message $\mathbf{m}^K$ of Eq. (14) carried inside that loop, and the adjacent weight $\widetilde{K}_{i,i+1}$ appended to the pair in the readout (App. C). With all three switched off, Eq. (4) reduces to Eq. (11) evaluated on the chain-adjacent pair $(i, i+1)$, the construction that the UCCSD and CEO heads use. We call this the inherited readout and compare the two heads under one protocol. The inherited readout keeps the layer embedding e_ℓ, because a brickwall emits one angle per bond per circuit layer, and without it the d layers would receive the same angle. We remove the three components together: without message-passing layers nothing carries $\mathbf{m}^K$, and $\widetilde{K}_{i,i+1}$ enters the readout, so it would survive removal of the loop alone. The three components hold $147{,}840$ of the head's $1{,}264{,}200$ parameters, or 12%.

Both heads are trained at depth six on the same 500 QM7 molecules and selected on the same 716-molecule holdout, with the same optimizer, plateau schedule, seed, epoch cap, early stopping, global batch size of four, and 125 optimizer steps per epoch. The comparison run splits that batch over two GPUs as two per device, which leaves the averaged gradient unchanged. The inherited readout reaches a mean validation overlap of 0.2356, against 0.3801 for the brickwall head; the gains over Hartree–Fock are 0.0126 and 0.1572. On the training molecules it reaches 0.2682 against 0.4495 (Fig. 6). The result supports the chain propagation and the orbital-interaction features together and does not isolate their individual effects.

G Per-molecule analysis

G.1 State-vector targets

Table 1 reports means over all records. This appendix breaks the same results down by record, by parent molecule, by qubit count, and by Hartree–Fock overlap. The set is the 300 QM7-StateVec-30k records of Tab. 1, drawn from 68 parent molecules and 36 molecular formulas. These records are the split used for checkpoint selection, so they are not an independent test set. The analysis uses the saved evaluations and does no new training or inference.

For record j and training seed s, let p_{js} be the squared overlap. The gain of record j is $g_j = \mathrm{mean}_s(p_{js} - p_j^{\mathrm{HF}})$ over the five seeds, and every record has equal weight. We report two standard errors of the mean gain. The record SE treats the 300 records as independent. The parent SE treats the geometries of one parent molecule as one cluster,

$$\mathrm{SE}_{\mathrm{parent}} = \frac{1}{N}\sqrt{\frac{G}{G-1}\sum_{c=1}^{G}\Big[\sum_{j\in c}(g_j-\bar{g})\Big]^2}, \qquad N = 300,\ G = 68. \tag{17}$$

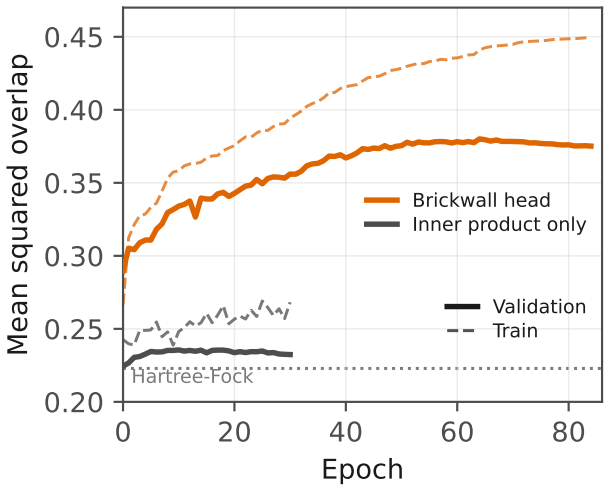


Figure 6: The brickwall head against the inherited pair readout, both at depth six on the same 500 QM7 molecules. "Inner product only" is Eq. (11) on chain-adjacent pairs, conditioned on the circuit layer. The dotted line is the mean Hartree–Fock squared overlap on the validation split, where every generated circuit starts because the final projection is zero-initialized.

Isomers have different parents, so they are separate clusters. Both SEs describe uncertainty over the dataset for the trained models. They do not include the variation between training runs, which Tab. 1 reports separately.

Table 4: Per-record statistics on the 300 QM7-StateVec-30k records. A record is below HF when its seed-mean gain is negative; every such record also has $g_j < -10^{-6}$. After averaging over its geometries, no parent molecule has a negative gain. The mean $1/p$ averages $\mathrm{mean}_s(1/p_{js})$ over the records.

Method	Mean gain	SE, parent	SE, record	Records below HF	Mean $1/p$
brickwall, depth 4	0.1471	0.0097	0.0078	4	6.6×10^5
UCCSD	0.2208	0.0065	0.0081	1	3.6×10^3
CEO	0.2607	0.0102	0.0117	1	1.7×10^3
Target-aware benchmark, L4	0.2228	0.0099	0.0105	0	1.33
Hartree–Fock	0	—	—	—	1.3×10^6

Table 5: Mean gain over Hartree–Fock by Hartree–Fock overlap. The gain is bounded above by $1 - p_j^{\mathrm{HF}}$, so part of its fall at high HF overlap follows from the definition of the metric.

HF overlap	Records	Parents	brickwall	UCCSD	CEO	Benchmark, L4
$[0.0, 0.2)$	59	34	0.014	0.312	0.512	0.486
$[0.2, 0.4)$	13	13	0.305	0.363	0.383	0.334
$[0.4, 0.6)$	39	31	0.342	0.385	0.388	0.310
$[0.6, 0.8)$	76	53	0.219	0.229	0.223	0.186
$[0.8, 1.0]$	113	63	0.083	0.095	0.097	0.067

Correlation strength. Table 5 and Fig. 7 show where the methods differ. For HF overlap above 0.2, the depth-four brickwall gains 80 to 98% of what CEO gains in each bin, and it gains more than the benchmark for HF overlap above 0.4. For the 59 records with HF overlap below 0.2, the brickwall gains only 0.014, while CEO gains 0.512 and the benchmark gains 0.486. All four brickwall records below HF fall in this bin. Thus, this bin accounts for 86% of the mean-gain gap between the brickwall and CEO in Tab. 1.

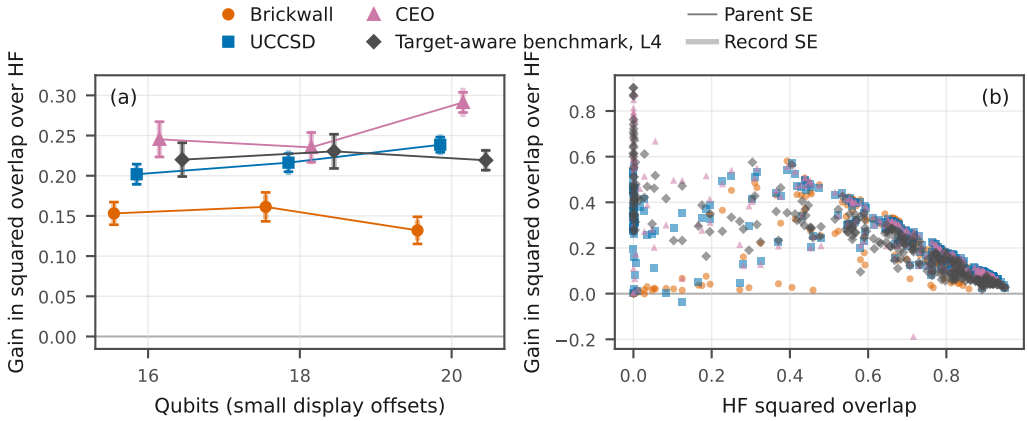


Figure 7: Gain over Hartree–Fock on the 300 QM7-StateVec-30k records. (a) Mean gain by qubit count, with the parent-cluster SE (thin bar) and the record SE (thick bar). (b) One point per record against its Hartree–Fock overlap. Each learned value is the mean over five training seeds.

Qubit count. The mean gain changes little with active-space size. From 16 to 18 to 20 qubits (91, 89, and 120 records), the gain is 0.153, 0.161, and 0.132 for the brickwall, 0.202, 0.216, and 0.239 for UCCSD, 0.245, 0.235, and 0.291 for CEO, and 0.220, 0.230, and 0.219 for the benchmark. The three slices contain different molecules, so these differences do not isolate an effect of size.

Rare failures. The expected number of phase-estimation repetitions scales with $1/p$, and a few records with very small overlap dominate its mean. One geometry, a $C_5H_8O_2$ molecule at stretch factor 1.6, contributes 99.9% of the UCCSD sum and 99.8% of the CEO sum; its lowest seed overlaps are 1.8×10^{-7} and 3.9×10^{-7}. The same parent at stretch factor 2.0 contributes 64% of the Hartree–Fock sum and 65% of the brickwall sum. The benchmark, which receives each target state, has no such failure, and its mean $1/p$ is 1.33. The worst single records are a C_6H_5N geometry for the brickwall (gain -0.001) and for UCCSD (-0.038), and a $C_6H_{12}O$ geometry for CEO (-0.188). A learned initial state therefore needs a check before use, for example the energy of the prepared state, because a rare geometry can receive a circuit worse than Hartree–Fock.

Scope. The brickwall is scored against an exact MPS factorization of the CASCI vector, and UCCSD, CEO, and the benchmark against the dense vector itself. The two forms represent the same state: their mean Hartree–Fock overlaps agree to 7×10^{-7}. The analysis is descriptive. We selected no hypothesis test from these slices.

G.2 Zero-shot QM9 targets

The same analysis on the 100 QM9-MPS-100 molecules uses one training run per model, so its standard errors describe the spread over molecules only. Each test molecule is its own parent, so the parent-cluster SE equals the record SE. The 100 molecules are four to six isomers of each of 20 formulas, and isomers of one formula can be correlated. We therefore also cluster by formula ($G = 20$). Table 6 gives the paired differences that the comparison of Sec. 5.3 rests on. Formula clustering widens their standard errors by a factor of 1.3 to 1.6. The depth-six brickwall trained on more geometries stays ahead of the benchmark at four and at eight layers. The depth-six brickwall on one geometry per molecule is not distinguishable from the benchmark at three or four layers under formula clustering. Every step of the training-set sweep stays significant.

No failures below Hartree–Fock. No model and no benchmark condition falls below Hartree–Fock on any QM9-MPS-100 molecule. The mean $1/p$ is 2.93 for Hartree–Fock, 1.68 for the depth-six brickwall trained on 6,449 states, 1.64 for the one trained on all 25,877 states, and 1.67 and 1.65 for the benchmark at four and eight layers. Unlike the QM7-StateVec-30k set, this set has no rare near-zero overlap, so the mean $1/p$ tracks the mean overlap.

Table 6: Molecule-paired differences in gain over Hartree–Fock on the 100 QM9-MPS-100 molecules, with 95% intervals from the record SE and from the SE clustered by molecular formula ($G = 20$); intervals are mean ± 1.96 SE. Each model is one training run.

Paired difference in gain	Mean	95% CI	95% CI, formula
d6, more geometries − benchmark L4	+0.0147	[+0.0098, +0.0196]	[+0.0082, +0.0212]
d6, more geometries − benchmark L8	+0.0090	[+0.0039, +0.0141]	[+0.0023, +0.0157]
d6, 6,449 states − benchmark L3	+0.0083	[+0.0024, +0.0142]	[−0.0011, +0.0177]
d6, 6,449 states − benchmark L4	+0.0042	[−0.0017, +0.0101]	[−0.0052, +0.0136]

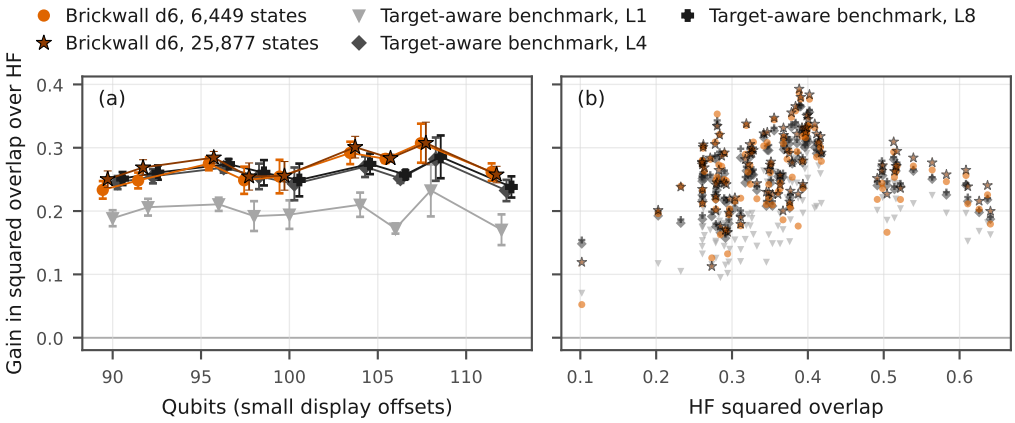


Figure 8: Gain over Hartree–Fock on the 100 QM9-MPS-100 molecules. (a) Mean gain by qubit count, with one standard error over molecules. (b) One point per molecule against its Hartree–Fock overlap. The benchmark is shown at one, four, and eight layers. Each learned model is one training run.

Qubit count and correlation strength. The gain does not fall with active-space size (Fig. 8). Over the qubit counts of the test set (90 to 112), the mean gain of the depth-six brickwall stays between 0.23 and 0.31, and it exceeds that of the benchmark at four layers in six of the nine qubit counts. The test set spans a narrow range of Hartree–Fock overlap: 75 of the 100 molecules lie between 0.2 and 0.4. The one molecule below 0.2, a $C_4H_9N_5$ isomer with Hartree–Fock overlap 0.10, is the lowest-overlap record for every method. There the depth-six brickwall reaches an overlap of 0.15, the model trained on more geometries 0.22, and the benchmark at eight layers 0.26. This case agrees with the state-vector result above: the brickwall gains least where the Hartree–Fock state is weakest.

H Training details

All reported models use the four-layer equivariant MO encoder with representations $64{\times}0e{\oplus}64{\times}1o{\oplus}64{\times}2e$, a $576{\times}0e$ scalar hidden representation, attention latent representations $32{\times}0e{\oplus}32{\times}1o{\oplus}32{\times}2e$, a radial cutoff of 4.0 Å, and eight average neighbors. The dense UCCSD and CEO heads use the same $32{\times}0e \oplus 32{\times}1o \oplus 32{\times}2e$ latent pair representations. The molecular brickwall head uses three chain graph-neural-network layers, a 16-unit residual multilayer perceptron, an eight-dimensional layer embedding, untied circuit layers, and circular angle parameterization. We train matched molecular models at brickwall depths three, four, five, and six. The hydrogen-chain brickwall models use one chain graph-neural-network layer and the same depths three through six; only depth six is repeated across three optimization seeds, and only that depth is reported.

We minimize $1 - \langle \mathcal{O}_\theta \rangle$ with AdamW in double-precision complex simulation. The run configurations are summarized in Tab. 7; the epoch counts are upper bounds because validation-based early stopping is used where indicated. On QM7-StateVec-30k, one epoch is 1,024 batches; the learning rate is halved after 36 validation epochs without an absolute loss improvement of 10^{-4}, and training stops after 84 such epochs. For the QM7 DMRG-MPS depth comparison, the learning rate

is halved after five validation epochs without an absolute loss improvement of 10^{-4}, down to 10^{-6}, and training stops after 15 such epochs. The hydrogen-chain scheduler uses the same factor and threshold with patience 40, and early stopping patience 150. Reported seed studies vary only the optimization seed while holding the data split fixed.

Table 7: Optimization hyperparameters for the reported experiment families. Every run uses a reduce-on-plateau schedule with early stopping on the validation loss; the validation column gives the held-out records or molecules used for checkpoint selection.

Setting	Batch	Initial LR	Epoch cap	Validation
QM7-StateVec-30k, UCCSD, CEO, brickwall d4	4	10^{-3}	700	300 records
Hydrogen-chain brickwall	1	10^{-2}	3000	H_{26}, H_{30}
QM7-MPS-7k, depths 3–6	4	10^{-3}	80	716
QM7-MPS-7k subsets, 100–3,000 molecules	4	10^{-3}	150–300	716
All 25,877 states, depth 6	4	10^{-3}	80	716

Table 8 lists the per-chain squared overlaps behind Fig. 2.

Table 8: Squared overlap with the bond-dimension-500 DMRG target on the strictly held-out hydrogen chains, at brickwall depth six. The three seeds differ only in optimization; the data split is fixed, and the standard error of the mean (SEM) is taken over the three seed-level values. Overlapotron checkpoints are selected only by mean overlap on H_{26} and H_{30}, while the target-aware benchmark is compiled from each target MPS at $\chi_{\max} = 64$ and at the same depth. The last column is the ratio of the seed mean to the benchmark.

		Overlapotron, $d = 6$					
Chain	Hartree–Fock	seed 42	seed 43	seed 44	mean ± SEM	Benchmark	Ratio
H_{34}	2.01×10^{-5}	0.5398	0.4816	0.4952	0.5055 ± 0.0176	0.1299	3.9×
H_{38}	5.43×10^{-6}	0.4887	0.4293	0.4429	0.4536 ± 0.0180	0.0965	4.7×
H_{42}	1.45×10^{-6}	0.4403	0.3801	0.3938	0.4047 ± 0.0182	0.0725	5.6×
H_{46}	3.82×10^{-7}	0.3939	0.3342	0.3476	0.3586 ± 0.0181	0.0525	6.8×
H_{50}	1.01×10^{-7}	0.3533	0.2947	0.3077	0.3186 ± 0.0178	0.0389	8.2×
Mean	—	0.4432	0.3840	0.3974	0.4082 ± 0.0179	0.0781	5.2×