FOUNDATION NEURAL-NETWORK QUANTUM STATES FOR MOLECULAR POTENTIAL ENERGY SURFACES IN SECOND QUANTIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Second-quantized neural-network quantum states have achieved accurate molecular energies, but extending them across molecular geometries requires a shared representation of the geometry-dependent wavefunction coefficients. We introduce geometry-conditioned foundation neural-network quantum states for molecular electronic structure in second quantization. A single autoregressive model learns a family of ground states from sparse anchor geometries and provides wavefunctions at untrained geometries without further optimization. Orbital alignment matches orbital identities and transports their phases, establishing an aligned orbital basis across geometries. Frozen energies reach chemical accuracy at every untrained query geometry for N_2, CO, and H_4. On additional molecular paths, the energy-trained wavefunctions yield dipoles, quadrupoles, and natural occupations without property labels. Across three paired N_2 training seeds, orbital alignment lowers the mean absolute energy error over all untrained query geometries from 34–37 mHa to 0.049–0.085 mHa. At approximately 1 mHa mean absolute error, frozen evaluation reduces the per-geometry cost by $986\times$ relative to independent optimization, yielding an estimated $25.8\times$ end-to-end GPU-cost reduction on a 161-point N_2 grid.

1 INTRODUCTION

Molecular potential energy surfaces (PESs) connect electronic structure to vibrations, conformational changes, and chemical reactions. Neural-network quantum states (NQSs) provide an expressive ansatz for many-body wavefunctions (Carleo & Troyer, 2017; Hermann et al., 2023). For molecules in second quantization, NQSs parameterize the wavefunction coefficients in an occupation basis and achieve accurate electronic energies through *ab initio* variational optimization (Choo et al., 2020; Barrett et al., 2022; Zhao et al., 2023; Liu & Clark, 2024; Kan et al., 2025; Shang et al., 2025). Constructing a PES with independently optimized NQSs, however, repeats this optimization at each geometry, leaving the shared electronic structure along the molecular path largely unused (Wu et al., 2025).

Sharing an NQS across a family of Hamiltonians offers a route to amortizing repeated wavefunction optimization. Foundation neural-network quantum states (FNQSs) establish this principle for quantum spin systems by conditioning a shared model on Hamiltonian couplings (Rende et al., 2025). Extending this framework to molecular electronic structure in second quantization requires a compact conditioning representation for the molecular Hamiltonian family. In this work, we introduce a geometry-conditioned FNQS that combines molecular geometry with orbital descriptors to provide global and orbital-specific conditioning. Jointly trained at sparse anchor geometries along a PES, the model provides wavefunctions at untrained geometries with frozen parameters, from which energies and electronic observables are evaluated.

Sharing a second-quantized wavefunction model across molecular geometries also requires a consistent orbital basis (Bensberg & Reiher, 2023; Rath & Booth, 2025). Independently generated molecular orbitals can change order and sign across geometries. The same occupation string can then refer to different orbital identities or phase conventions, introducing artificial discontinuities in the wavefunction coefficients even when the physical state varies smoothly. We align orbital identities

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.

and phases along the path to establish a consistent basis for the shared model. Ablation studies show that this orbital consistency enables accurate predictions at untrained geometries.

Our contributions are:

1. **FNQSs for molecular second quantization.** To our knowledge, this is the first variationally trained, geometry-conditioned NQS for molecular electronic structure in second quantization that predicts wavefunctions at untrained geometries with frozen parameters. Frozen predictions reach chemical accuracy at every untrained query geometry for N_2, CO, and H_4. Energy-trained wavefunctions also yield dipoles, quadrupoles, and natural occupations without property-specific training.
2. **Orbital consistency for accurate parameter sharing.** We combine orbital matching and phase alignment with geometry and per-orbital conditioning. Across three paired N_2 seeds, alignment reduces the mean absolute errors over all non-anchor queries from 34–37 to 0.049–0.085 mHa, demonstrating that orbital consistency enables accurate frozen transfer.
3. **Amortized computation without query-time optimization.** On a 161-point N_2 grid, our method yields estimated end-to-end GPU-cost reductions of $25.8\times$ over independent optimization and $2.53\times$ over shared pretraining with fine-tuning, using fixed budgets with MAEs near 1 mHa. The comparison includes initial training and energy evaluation; a cost model separates this initial investment from the cost of additional geometries.

2 Related work

Shared real-space wavefunctions and potential energy surfaces. PESNet trains a single real-space wavefunction across molecular geometries, reducing repeated variational optimization (Gao & Günnemann, 2022). PlaNet adds an energy surrogate trained on intermediate VMC estimates, enabling energy queries without Monte Carlo sampling at inference (Gao & Günnemann, 2023a). Both methods use real-space formulations, parameterizing electronic wavefunctions over continuous electron coordinates (Pfau et al., 2020; Hermann et al., 2020; von Glehn et al., 2023). Related work extends reuse across molecules through transferable orbital constructions and pretrained wavefunctions (Scherbela et al., 2024; Gao & Günnemann, 2023b; Foster et al., 2026). We instead adopt a second-quantized autoregressive representation, which facilitates comparison with conventional quantum-chemical methods and evaluation of electronic observables, while also enabling direct sampling under particle-number and symmetry constraints (Barrett et al., 2022; Zhao et al., 2023; Malyshev et al., 2023).

Foundation neural-network quantum states. FNQSs condition a shared spin wavefunction on Hamiltonian couplings (Rende et al., 2025). Their embedding concatenates $O(1)$ global couplings to each configuration patch, or pairs patches of $O(N)$ couplings with the corresponding configuration patches, where N is the number of spins. The authors identify second-quantized fermions and molecular systems as directions for extension. Molecular two-electron integrals carry four orbital indices and number $O(M^4)$ for M orbitals, so their full tensor does not directly fit either input construction. Molecular conditioning therefore calls for an appropriate encoding of the Hamiltonian family, alongside orbital alignment to maintain a consistent meaning for the learned coefficients. We use geometry and aligned per-orbital descriptors for this encoding.

Wavefunction interpolation. Eigenvector continuation (EC) interpolates wavefunctions by solving the query Hamiltonian in the span of anchor states (Frame et al., 2018; Mejuto-Zaera & Kemper, 2023). Molecular extensions use a consistent orbital basis to transfer correlated states across geometries and calculate energies and other electronic properties (Rath & Booth, 2025). Applying EC to sampled NQSs requires estimating cross-state overlaps and Hamiltonian matrix elements. Near-linear dependence among anchor states can amplify sampling errors in the resulting generalized eigenvalue problem and produce spurious low-energy solutions; overlap-spectrum truncation and other noise-aware regularization methods address this sensitivity (Epperly et al., 2022; Hicks & Lee, 2023). We instead learn a geometry-conditioned wavefunction whose frozen-query evaluation requires neither cross-state matrix elements nor a generalized eigenvalue solve.

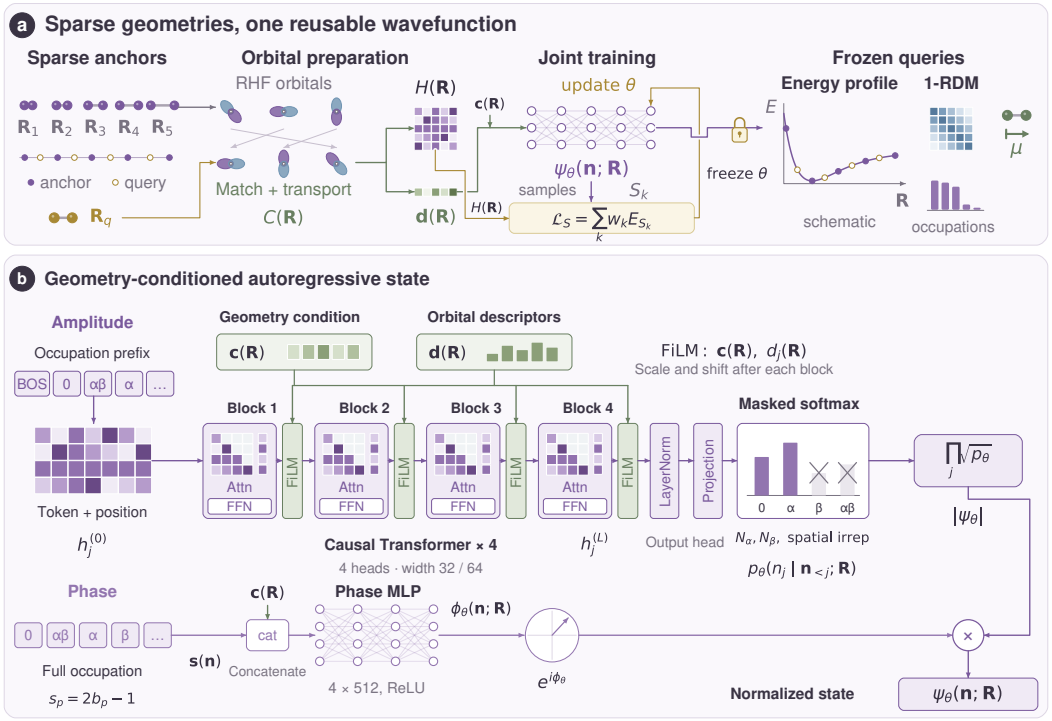


Figure 1: **Geometry-conditioned FNQSs.** (a) Starting from sparse anchor geometries, restricted Hartree–Fock (RHF) orbitals are matched and phase-aligned to construct consistent occupation bases and molecular Hamiltonians. Joint variational Monte Carlo (VMC) training produces a shared wavefunction model. At new geometries, the frozen model yields energies and the one-particle reduced density matrix (1-RDM). (b) Occupation prefixes pass through causal Transformer blocks, with geometry and orbital descriptors supplied through feature-wise linear modulation (FiLM), to produce normalized occupation probabilities and hence the wavefunction amplitude. A separate multilayer perceptron (MLP) maps the full occupation string and geometry to a phase. Amplitude and phase combine into the wavefunction.

3 METHOD

Problem setup. Fix a molecule, an atomic-orbital basis set, and spin populations (N_α, N_β) in M retained spatial orbitals. The electronic Hamiltonian at nuclear geometry $\mathbf{R}$ is

$$H(\mathbf{R}) = E_c(\mathbf{R}) + \sum_{pq} h_{pq}(\mathbf{R}) a_p^\dagger a_q + \frac{1}{2} \sum_{pqrs} v_{pqrs}(\mathbf{R}) a_p^\dagger a_q^\dagger a_s a_r, \tag{1}$$

where p, q, r, s label spin orbitals, h_{pq} and $v_{pqrs} = \langle pq|rs\rangle$ are the one- and two-electron integrals in physicists' notation and E_c collects nuclear repulsion and frozen-core terms. A state is a coefficient function $\psi(\mathbf{n}; \mathbf{R})$ over occupation strings $\mathbf{n} = (n_1, \ldots, n_M)$ with $n_j \in \{0, \alpha, \beta, \alpha\beta\}$. The task is to build a shared model for the ground-state wavefunction of a family $\{H(\mathbf{R})\}$ along a molecular path.

Overview. Fig. 1 summarizes the construction. We first align orbital bases across geometries (Section 3.1), then condition an autoregressive wavefunction on geometry and aligned orbital descriptors (Section 3.2). Joint variational optimization at sparse anchors learns shared parameters, which remain frozen when evaluating energies and electronic observables at new geometries (Section 3.3).

3.1 ORBITAL ALIGNMENT ACROSS GEOMETRIES

The choice of orbital basis affects the expressivity of a restricted variational ansatz (Robledo Moreno et al., 2023). The occupation basis is built from RHF orbitals. In this basis, molecular wavefunctions often concentrate much of their probability mass on a relatively small subset of configurations (Malyshev et al., 2024). Batched autoregressive sampling (BAS) exploits this concentration by aggregating repeated configurations (Barrett et al., 2022).

Sharing coefficients across geometries requires a consistent convention for these orbitals. Sorting orbitals by energy at each geometry does not ensure this consistency: independent RHF calculations can return orbitals with arbitrary signs, and energy crossings can change their order. A sign flip of spin-orbital p multiplies $\psi(\mathbf{n};\mathbf{R})$ by -1 if p is occupied, and an orbital permutation relabels configurations and transforms their coefficients with the corresponding fermionic parity. These changes in the orbital basis can make the coefficients discontinuous even when the physical state varies smoothly, complicating learning with a continuous geometry-conditioned model.

Learning from eigenvectors requires accounting for their nonunique representation: SignNet and BasisNet address sign and degenerate-basis ambiguities in spectral graph learning (Lim et al., 2023), while transferable neural orbitals have incorporated orbital-sign equivariance (Scherbela et al., 2024). In an occupation-basis model, orbital transformations act on the wavefunction coefficients themselves. We handle this dependence by aligning orbital identities and phases before sharing the autoregressive model across geometries.

For orbital coefficient matrices C_A, C_B at neighboring geometries, the overlap is $O_{AB} = C_A^\dagger S_{AB} C_B$, where S_{AB} contains the overlaps between their atomic-orbital (AO) bases. Within compatible occupied/virtual, frozen-core, and symmetry classes, maximum-overlap assignment (Kuhn, 1955) selects the permutation $\pi^* = \arg\max_\pi \sum_j |(O_{AB})_{j,\pi(j)}|$. The target orbitals are reordered by π^* and phase-aligned to make matched overlaps real and nonnegative. Where exact degeneracies leave symmetry labels ambiguous, orbitals are first resolved with respect to a fixed physical symmetry operation, as in the NH_3 construction detailed in Appendix A. One- and two-electron integrals are transformed into the aligned orbital basis, in which orbital descriptors are also evaluated.

3.2 GEOMETRY-CONDITIONED AUTOREGRESSIVE WAVEFUNCTIONS

Using aligned orbitals with coefficient matrix $C(\mathbf{R})$, the model represents the state as

$$
\begin{aligned}
|\Psi_\theta(\mathbf{R})\rangle &= \sum_{\mathbf{n}\in\Omega} \psi_\theta(\mathbf{n};\mathbf{R})|\mathbf{n};C(\mathbf{R})\rangle, \\
\psi_\theta(\mathbf{n};\mathbf{R}) &= e^{i\phi_\theta(\mathbf{n};\mathbf{R})} \prod_{j=1}^{M} \sqrt{p_\theta(n_j \mid \mathbf{n}_{<j};\mathbf{R})}.
\end{aligned}
\tag{2}
$$

Here Ω contains occupation strings with the prescribed spin populations and, when imposed, total spatial irreducible representation (irrep), and $\mathbf{n}_{<j}$ denotes the occupations preceding orbital j. Before each conditional is normalized, a mask excludes choices that cannot be completed to a configuration in Ω, using an exact backward reachability table when spatial symmetry is imposed. The resulting conditionals define a normalized state and permit direct autoregressive sampling (Sharir et al., 2020; Hibat-Allah et al., 2020; Barrett et al., 2022; Malyshev et al., 2023). The amplitude network uses L causal Transformer blocks, each combining attention and a feed-forward network (FFN), to predict each occupation from its prefix (Vaswani et al., 2017; Shang et al., 2025). The shifted token and position embeddings in Fig. 1(b) are $h_j^{(0)} = e_{\mathrm{tok}}(n_{j-1}) + e_{\mathrm{pos}}(j)$, where n_0 is the beginning-of-sequence (BOS) token.

For a fixed molecule, basis set, and electronic sector, geometry parameterizes the Hamiltonian family along the chosen path. We therefore use geometry as a compact global condition, supplemented by aligned orbital descriptors that provide orbital-specific electronic information at each position in the occupation sequence. Geometry enters through a global condition $\mathbf{c}(\mathbf{R})$, such as a bond length or a normalized path coordinate, and a vector $\mathbf{d}(\mathbf{R})$ with one orbital descriptor per aligned orbital. Each component is the aligned Fock diagonal relative to the mean over active occupied orbitals, scaled by one hartree (Ha): $d_j(\mathbf{R}) = \big(F_{jj}^{\mathrm{aligned}}(\mathbf{R}) - \overline{F}_{\mathrm{occ}}(\mathbf{R})\big)/(1\,\mathrm{Ha})$, where F^{aligned} is the Fock matrix

in the aligned orbital basis and $\overline{F}_{\mathrm{occ}}$ is that mean. Under a signed permutation, its diagonal entries equal the canonical orbital energies.

After each Transformer block, a FiLM adapter (Perez et al., 2018) modulates the hidden features using both inputs. For block output $u^{(\ell)} = \mathrm{Transformer}_\ell(h^{(\ell-1)})$, the FiLM update is

$$h_j^{(\ell)} = \big(1 + \gamma_\ell^c(\mathbf{c}) + \gamma_\ell^d(d_j)\big) \odot u_j^{(\ell)} + \beta_\ell^c(\mathbf{c}) + \beta_\ell^d(d_j). \tag{3}$$

The γ and β maps produce feature-wise scales and shifts, and $\odot$ denotes element-wise multiplication. The geometry maps are affine, while the descriptor maps are linear. All four maps are initialized to zero, so each adapter initially acts as the identity. After the final block, layer normalization (LayerNorm) (Ba et al., 2016) and a linear projection produce logits for the masked softmax, yielding $p_\theta(n_j \mid \mathbf{n}_{<j}; \mathbf{R})$ (Fig. 1(b)). Implementation and optimizer settings are given in Appendix B.

A separate phase MLP with rectified linear unit (ReLU) activations takes the concatenation of $\mathbf{c}(\mathbf{R})$ and the full occupation string encoded as $\mathbf{s}(\mathbf{n}) \in \{-1, 1\}^{2M}$, with $s_p = 2b_p - 1$ for spin-orbital occupation $b_p \in \{0, 1\}$. It predicts the angle $\phi_\theta(\mathbf{n}; \mathbf{R})$, whose factor $e^{i\phi_\theta}$ combines with the autoregressive amplitude in Equation (2). Orbital descriptors enter only the amplitude branch.

3.3 Joint variational training and frozen inference

The model is trained using a first-principles VMC method. The geometry-dependent molecular Hamiltonians supply the variational objective directly, without labeled energies or properties. At each optimization update, BAS propagates counts through successive occupation prefixes, drawing configurations at each anchor geometry from the corresponding wavefunction distribution (Barrett et al., 2022). A joint energy estimator constructed from these samples (Wu et al., 2023; Malyshev et al., 2024) is minimized. Let S_k contain the distinct configurations sampled at anchor k, ψ_{S_k} their coefficient vector, and $H_{S_k S_k}$ the corresponding Hamiltonian submatrix. The joint objective is

$$\begin{aligned} E_{S_k}(\theta; \mathbf{R}_k) &= \frac{\psi_{S_k}^\dagger H_{S_k S_k}(\mathbf{R}_k) \psi_{S_k}}{\psi_{S_k}^\dagger \psi_{S_k}}, \\ \mathcal{L}_S(\theta) &= \sum_k w_k E_{S_k}(\theta; \mathbf{R}_k), \end{aligned} \tag{4}$$

where w_k are the prescribed anchor weights. All anchors contribute to every joint update, extending the shared-training infrastructure of Wu et al. (2025). Appendix B gives the explicit gradient and sampling implementation. The energy of the complete network state is evaluated separately using the frozen-query procedure below.

Frozen queries and electronic observables. At a query geometry $\mathbf{R}_q$, orbital preparation supplies the aligned Hamiltonian and the geometry and orbital inputs (Fig. 1(a)). The network parameters remain frozen while fresh configurations are drawn from $|\psi_\theta(\mathbf{n}; \mathbf{R}_q)|^2$. For each sampled configuration, the local energy includes all determinants connected by the Hamiltonian:

$$\begin{aligned} E_{\mathrm{loc},\theta}(\mathbf{n}; \mathbf{R}) &= \sum_{\mathbf{n}'} H_{\mathbf{n}\mathbf{n}'}(\mathbf{R}) \frac{\psi_\theta(\mathbf{n}'; \mathbf{R})}{\psi_\theta(\mathbf{n}; \mathbf{R})}, \\ E_\theta(\mathbf{R}) &= \mathbb{E}_{|\psi_\theta|^2} \operatorname{Re} E_{\mathrm{loc},\theta}. \end{aligned} \tag{5}$$

The expectation is taken over the normalized network distribution at the query geometry. It is a variational upper bound to the ground-state energy in the chosen orbital space and sector; its Monte Carlo estimate carries sampling uncertainty, which is reported.

Replacing H by another operator gives its expectation in the same wavefunction. This yields the spin-summed 1-RDM, $\gamma_{ij} = \sum_\sigma \langle a_{i\sigma}^\dagger a_{j\sigma} \rangle$, where i, j label spatial orbitals. Its eigenvalues are the natural occupations. Molecular dipoles are obtained from the density matrix, including the frozen-core and nuclear contributions.

4 Experiments

Systems and evaluation. The energy benchmarks comprise N_2 and CO bond scans in STO-3G (Hehre et al., 1969) and H_4 rectangular deformation in cc-pVDZ (Dunning, 1989). Property

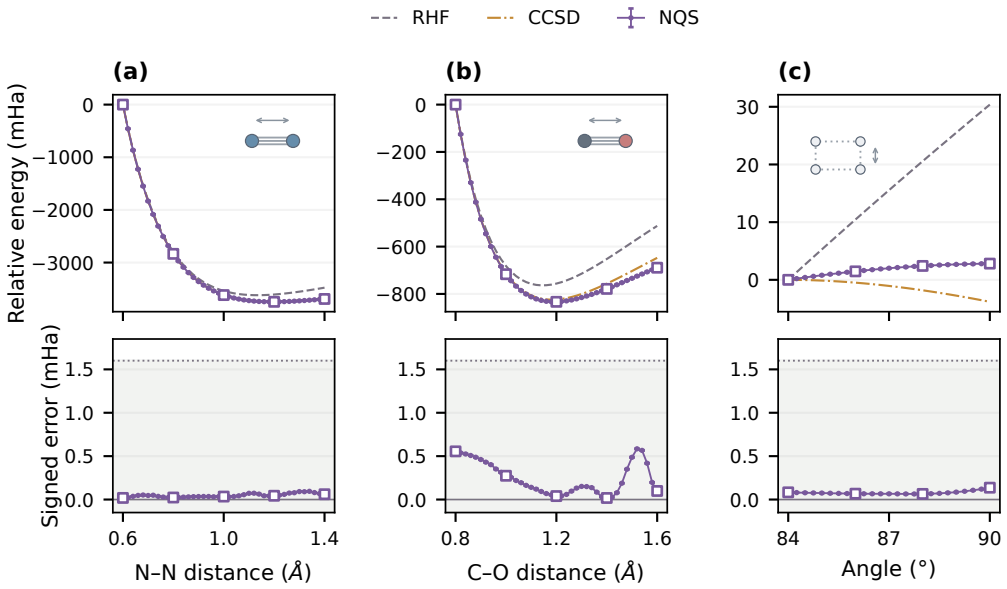


Figure 2: **Potential energy surfaces.** (a) N_2, (b) CO, and (c) H_4. Each panel combines NQS, RHF, and coupled-cluster singles and doubles (CCSD) relative energies (upper) and NQS signed errors against FCI (lower), with the 1.6 mHa band below. Energies are relative to each method's path origin. CCSD uses a restricted reference. Squares mark anchors, dots queries, and bars one sampling standard error.

benchmarks cover NH_3 inversion in 6-31G (Hehre et al., 1972), together with CO_2 symmetric stretching and C_2H_4 torsion in STO-3G. For each system, a model is trained jointly at sparse anchors and evaluated with frozen parameters at intervening query geometries. Appendix C specifies the paths, geometry splits, and retained electronic spaces; Appendix B gives the network and optimization settings.

Full configuration interaction (FCI) calculations provide the reference values in the same basis and frozen-core space, following the validation protocol for second-quantized neural states (Choo et al., 2020; Barrett et al., 2022; Shang et al., 2025). For an evaluation set Q, mean absolute error (MAE) is $|Q|^{-1}\sum_{q\in Q}|E_\theta(\mathbf{R}_q) - E_{\mathrm{FCI}}(\mathbf{R}_q)|$; we also report the maximum absolute error over Q and use 1.6 mHa as the chemical-accuracy threshold. Anchor and query errors are kept separate. The main energy benchmarks report statistics over all non-anchor queries. Neural energies use Equation (5), evaluated using three independent sampling replicates per geometry. Sampling error bars give one standard error of the mean. Property errors use the corresponding FCI observables at the same queries.

4.1 Frozen energy predictions across molecular paths

Sparse-anchor training yields accurate energy profiles at untrained geometries. In Fig. 2, a single frozen model per system resolves the energy profiles of N_2, CO, and H_4, with every anchor and query within 1.6 mHa of FCI. N_2 and CO query MAEs of 0.049 and 0.253 mHa show that anchor accuracy carries across both homonuclear and heteronuclear bond scans without further optimization.

H_4 tests transfer on a finer energy scale: its rise toward the square geometry spans only a few mHa. The model resolves this shallow multireference profile in the larger cc-pVDZ orbital space with a maximum query error of 0.120 mHa. Thus, the shared wavefunction captures both broad bond-stretching curves and small energy differences along deformation. Appendix E.1 gives the full anchor and query statistics.

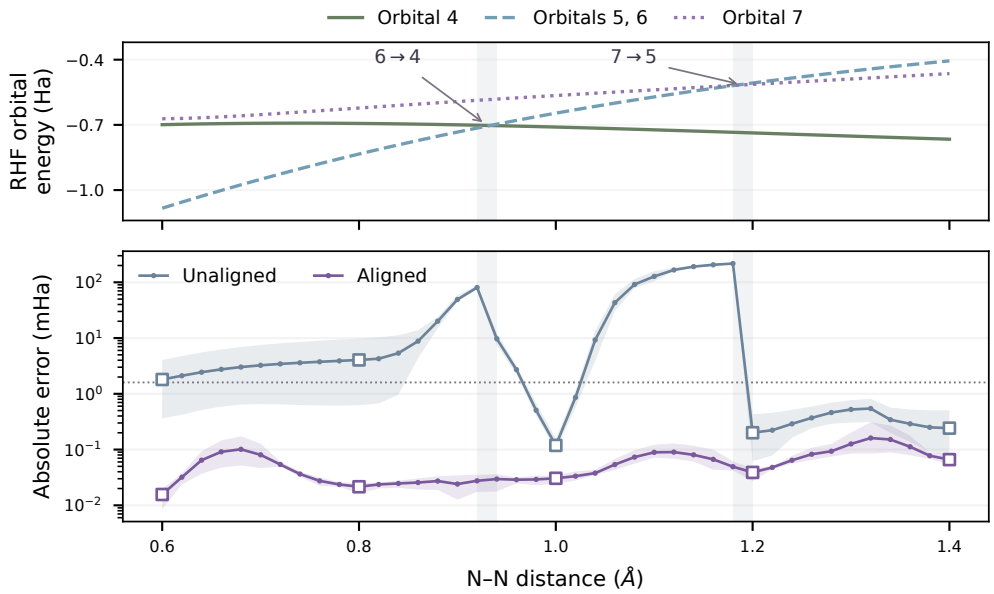


Figure 3: **Orbital alignment.** Alignment enables transfer between N_2 training anchors. The upper plot tracks aligned RHF orbital energies; the degenerate pair 5/6 is drawn as one curve. Arrows label changes in the canonical indices of the indicated continuous branches. Both plots share the N–N distance axis; vertical bands mark the two orbital-ordering exchanges (0.92–0.94 and 1.18–1.20 Å). The lower plot compares absolute energy errors over three paired seeds: lines show means and colored ribbons the minimum–maximum range. Squares mark anchors, dots queries, and the dotted line 1.6 mHa. Architecture, optimization, and sampling budgets are matched within each pair.

4.2 Orbital alignment enables accurate transfer

Fig. 3 compares the errors with unaligned and aligned orbitals along the N_2 scan across three paired training seeds. Within each pair, only the orbital basis changes; the architecture, optimizer, and sampling budget are identical. The aligned models predict every untrained geometry within chemical accuracy (query MAE 0.049–0.085 mHa across seeds). The unaligned models fit most anchors but fail between them, with query MAEs of 34–37 mHa and errors above 100 mHa adjacent to anchors; some unaligned runs also miss chemical accuracy at an anchor. The largest failures occur where canonical orbital ordering changes (upper plot in Fig. 3). Alignment removes these artificial discontinuities in the wavefunction coefficients, allowing accuracy at the anchors to extend to the intervening geometries.

Within the aligned orbital basis, conditioning ablations show smaller gains from FiLM and orbital descriptors (Appendix E.3).

4.3 Electronic observables from the shared wavefunction

The model supplies a complete electronic wavefunction, giving access to observables from the same state used to evaluate energy. Fig. 4 probes charge displacement, spatial charge distribution, and frontier occupations. Each observable is computed from the one-particle reduced density matrix of the frozen, energy-trained state, without property labels or additional fitting.

The NH_3 dipole follows charge redistribution during inversion with query MAE 0.0058 debye (D), comparable to CCSD. Symmetric CO_2 has no permanent dipole; its quadrupole instead reveals the changing charge distribution as the bonds lengthen. The model follows this variation with query MAE 0.0403 ea_0^2, below the CCSD error. Ethylene probes the growing multireference character during torsion: its frontier occupations approach one another toward the twisted geometry, a change that a restricted single determinant cannot describe. Their query MAE, averaged over both displayed occupations, is 0.0009, also below the CCSD error. These complementary observables show that

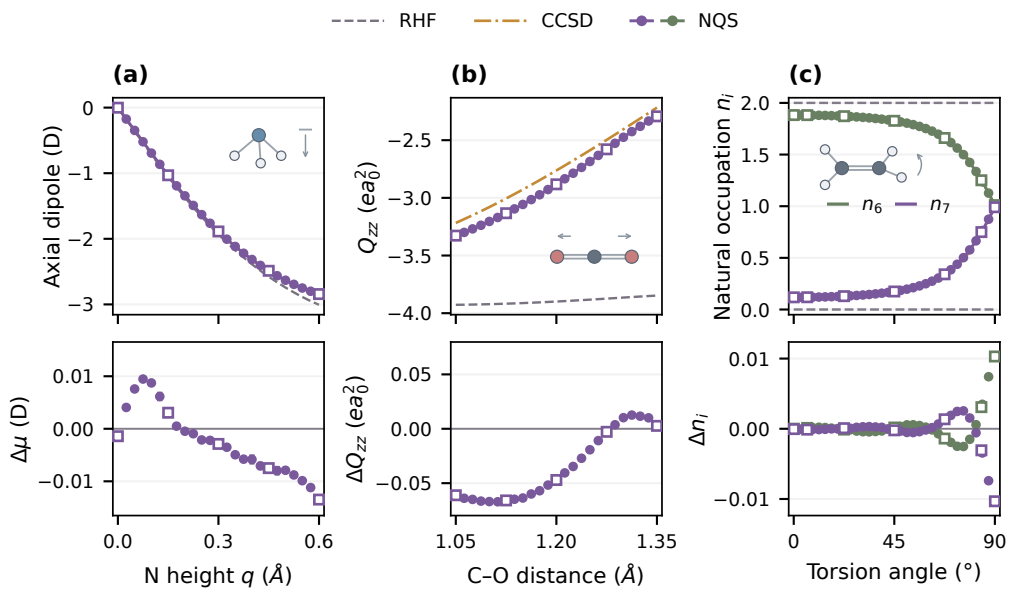


Figure 4: **Electronic properties.** (a) NH_3 dipole, (b) CO_2 quadrupole, and (c) ethylene natural occupations 6/7. Each panel combines the observable (upper) and its NQS-minus-FCI error (lower). Upper curves show NQS estimates connected by solid lines, RHF (dashed), and CCSD (dash-dotted). Circles mark queries, squares anchors, and bars one sampling standard error.

the shared state retains useful electronic-structure information beyond its energy. The lower panels resolve signed errors; Appendix E.2 reports energies from the same checkpoints and RHF/CCSD comparisons, and Appendix C defines the observables.

4.4 Computational efficiency

Sharing a wavefunction model reduces the work required at each new geometry. We quantify this benefit on the N_2 bond scan by comparing four workflows: independent optimization at every geometry, shared pretraining followed by per-geometry fine-tuning (Wu et al., 2025), sequential warm-start optimization from the preceding geometry, and joint training of our geometry-conditioned FNQS followed by frozen evaluation. The shared-pretraining baseline uses the same NQS backbone and aligned orbitals but omits geometry conditioning. It minimizes the average loss over the training anchors before its weights are fine-tuned independently at each query.

The cost comparison uses separate fixed-budget checkpoints from the main accuracy benchmarks, targeting an MAE of approximately 1 mHa; Appendix E.4 reports the budgets, measured errors, and timing details. Following PESNet and PlaNet (Gao & Günnemann, 2022; 2023a), we include both initial training and subsequent queries in the total A100 GPU-hour cost. For each workflow m, the estimated cost of G geometries takes the linear form

$$\widehat{C}_m(G) = a_m + Gb_m. \tag{6}$$

The intercept a_m represents the fixed cost, and the slope b_m the average cost of each additional geometry. Independent optimization has zero intercept: each geometry requires training from scratch and an energy estimate. Shared pretraining contributes a one-time cost, followed by fine-tuning and an energy estimate at each geometry. Sequential warm starts require training from scratch at the first geometry, followed by fine-tuning and an energy estimate at each subsequent geometry. Our geometry-conditioned FNQS incurs a one-time joint-training cost, after which each geometry requires only an energy estimate with frozen parameters (Fig. 5).

Frozen evaluation reduces the cost of each additional geometry by approximately $986\times$ relative to independent optimization, $77\times$ relative to shared pretraining with fine-tuning, and 41–$43\times$ relative to sequential warm starts. On the evaluated 161-point N_2 grid, the total cost is 1.644 GPU-hours, giving

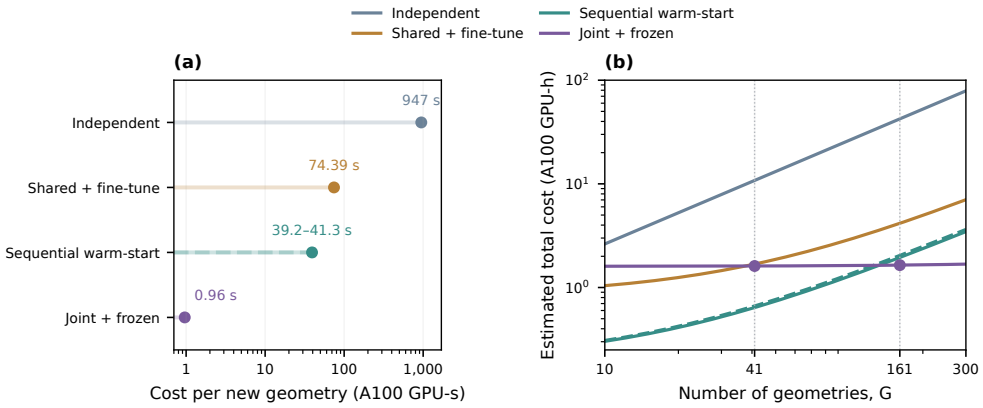


Figure 5: **Computational cost.** The N_2 workflows use fixed training budgets with MAEs near 1 mHa. (a) Average cost per additional geometry, including optimization where applicable and energy evaluation. The warm-start range spans the measured means at 41- and 161-point spacings. (b) Estimated total costs, including initial training, on logarithmic axes. Dots and vertical guides mark frozen evaluation on the 41- and 161-point grids. Independent and shared-adaptation costs are estimated from eight geometries selected by stratified sampling; warm-start dashed and solid curves use the 41- and 161-point spacings, respectively. Each curve holds its measured average per-geometry cost fixed as the query count varies.

estimated end-to-end GPU-cost reductions of $25.8\times$, $2.53\times$, and $1.18\times$, respectively. Denser queries spread the initial training investment over more geometries, bringing the overall speedup closer to the per-geometry ratio. Sequential warm starts impose dependencies along the path; the other workflows permit parallel computation across geometries, including multi-geometry training for the shared and conditioned models. Appendix E.4 reports the budgets, measured MAEs, and cost breakdowns.

5 Discussion and Conclusion

We have established geometry-conditioned FNQSs for molecular electronic structure in second quantization. Joint variational training turns sparse anchor calculations into a shared representation of molecular ground states, supplying energies and electronic observables at untrained geometries with frozen parameters. Orbital alignment establishes an aligned orbital basis across geometries, enabling the shared network to learn the physical variation along a molecular path.

The energy-trained states recover charge redistribution and changes in natural occupations without property-specific fitting. At fixed training budgets, denser queries amortize joint training, while per-geometry optimization and evaluation costs determine the asymptotic speedup. This work establishes an efficient route to computing molecular potential energy surfaces with neural-network quantum states in an occupation basis.

Limitations and outlook. Our benchmarks cover one-dimensional molecular paths with FCI references in specified finite orbital spaces; basis-set errors remain. Extending the method to larger active spaces will require efficient sampling and Hamiltonian evaluation as the number of relevant configurations grows. In multidimensional geometry domains, orbital matching can depend on the transport path, making consistency around closed loops an additional challenge near orbital crossings. A further direction is to extend orbital correspondence across different active spaces and, ultimately, across molecules. Together with architectures that accommodate varying orbital spaces and electron counts, such correspondences could support pretraining and wavefunction reuse across a broader range of chemical systems.

REPRODUCIBILITY STATEMENT

Appendices A, B, and C specify orbital alignment and query attachment, network and sampling settings, and molecular paths and electronic sectors, respectively. The anonymized supplementary archive provides orbital-alignment code, conditioning adapters, selected experiment configurations, pointwise energy and property estimates, and timing records. CPU scripts recompute the reported summaries and regenerate all manuscript figures from these recorded results. The accompanying instructions describe the method-component interfaces, geometry splits, and training schedules.

AI USE STATEMENT

The authors used ChatGPT and Cursor (with Claude Fable 5.1 and Grok 4.7) to assist with methodological development, experimental design, method implementation and debugging, numerical data processing, and interpretation of results. The tools also assisted with literature search, manuscript editing, and figure preparation. The authors reviewed all AI-assisted work, including the text, code, experimental design, results, and references. The authors take responsibility for the final content of this submission.

REFERENCES

Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E. Hinton. Layer normalization. arXiv:1607.06450, 2016. URL https://arxiv.org/abs/1607.06450.

Thomas D. Barrett, Aleksei Malyshev, and A. I. Lvovsky. Autoregressive neural-network wavefunctions for ab initio quantum chemistry. *Nature Machine Intelligence*, 4:351–358, 2022. doi: 10.1038/s42256-022-00461-z. URL https://www.nature.com/articles/s42256-022-00461-z.

Moritz Bensberg and Markus Reiher. Corresponding active orbital spaces along chemical reaction paths. *The Journal of Physical Chemistry Letters*, 14:2112–2118, 2023. doi: 10.1021/acs.jpclett.2c03905. URL https://arxiv.org/abs/2212.12883.

Giuseppe Carleo and Matthias Troyer. Solving the quantum many-body problem with artificial neural networks. *Science*, 355:602–606, 2017. doi: 10.1126/science.aag2302. URL https://arxiv.org/abs/1606.02318.

Kenny Choo, Antonio Mezzacapo, and Giuseppe Carleo. Fermionic neural-network states for ab-initio electronic structure. *Nature Communications*, 11:2368, 2020. doi: 10.1038/s41467-020-15724-9. URL https://www.nature.com/articles/s41467-020-15724-9.

Thom H. Dunning, Jr. Gaussian basis sets for use in correlated molecular calculations. i. the atoms boron through neon and hydrogen. *The Journal of Chemical Physics*, 90:1007–1023, 1989. doi: 10.1063/1.456153.

Ethan N. Epperly, Lin Lin, and Yuji Nakatsukasa. A theory of quantum subspace diagonalization. *SIAM Journal on Matrix Analysis and Applications*, 43(3):1263–1290, 2022. doi: 10.1137/21M145954X. URL https://arxiv.org/abs/2110.07492.

Adam Foster, Zeno Schätzle, P. Bernát Szabó, Lixue Cheng, Jonas Köhler, Gino Cassella, Nicholas Gao, Jiawei Li, Frank Noé, and Jan Hermann. An ab initio foundation model of wavefunctions that accurately describes chemical bond breaking. *Nature Communications*, 17:9976, 2026. doi: 10.1038/s41467-026-76604-2. URL https://www.nature.com/articles/s41467-026-76604-2.

Dillon Frame, Rongzheng He, Ilse Ipsen, Daniel Lee, Dean Lee, and Ermal Rrapaj. Eigenvector continuation with subspace learning. *Physical Review Letters*, 121:032501, 2018. doi: 10.1103/PhysRevLett.121.032501. URL https://arxiv.org/abs/1711.07090.

Nicholas Gao and Stephan Günnemann. Ab-initio potential energy surfaces by pairing GNNs with neural wave functions. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=apv504XsysP.

Nicholas Gao and Stephan Günnemann. Sampling-free inference for ab-initio potential energy surface networks. In *International Conference on Learning Representations*, 2023a. URL https://arxiv.org/abs/2205.14962.

Nicholas Gao and Stephan Günnemann. Generalizing neural wave functions. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *PMLR*, pp. 10708–10726, 2023b. URL https://proceedings.mlr.press/v202/gao23c.html.

Warren J. Hehre, Robert F. Stewart, and John A. Pople. Self-consistent molecular-orbital methods. i. use of gaussian expansions of slater-type atomic orbitals. *The Journal of Chemical Physics*, 51: 2657–2664, 1969. doi: 10.1063/1.1672392.

Warren J. Hehre, Robert Ditchfield, and John A. Pople. Self-consistent molecular orbital methods. xii. further extensions of gaussian-type basis sets for use in molecular orbital studies of organic molecules. *The Journal of Chemical Physics*, 56:2257–2261, 1972. doi: 10.1063/1.1677527.

Jan Hermann, Zeno Schätzle, and Frank Noé. Deep-neural-network solution of the electronic Schrödinger equation. *Nature Chemistry*, 12:891–897, 2020. doi: 10.1038/s41557-020-0544-y. URL https://www.nature.com/articles/s41557-020-0544-y.

Jan Hermann, James Spencer, Kenny Choo, Antonio Mezzacapo, W. M. C. Foulkes, David Pfau, Giuseppe Carleo, and Frank Noé. Ab initio quantum chemistry with neural-network wavefunctions. *Nature Reviews Chemistry*, 7:692–709, 2023. doi: 10.1038/s41570-023-00516-8. URL https://arxiv.org/abs/2208.12590.

Mohamed Hibat-Allah, Martin Ganahl, Lauren E. Hayward, Roger G. Melko, and Juan Carrasquilla. Recurrent neural network wave functions. *Physical Review Research*, 2:023358, 2020. doi: 10.1103/PhysRevResearch.2.023358. URL https://arxiv.org/abs/2002.02973.

Caleb Hicks and Dean Lee. Trimmed sampling algorithm for the noisy generalized eigenvalue problem. *Physical Review Research*, 5:L022001, 2023. doi: 10.1103/PhysRevResearch.5.L022001. URL https://arxiv.org/abs/2209.02083.

Bowen Kan, Yingqi Tian, Yangjun Wu, Yunquan Zhang, and Honghui Shang. Bridging the gap between transformer-based neural networks and tensor networks for quantum chemistry. *Journal of Chemical Theory and Computation*, 21(7):3426–3439, 2025. doi: 10.1021/acs.jctc.4c01703. URL https://pubs.acs.org/doi/10.1021/acs.jctc.4c01703.

Harold W. Kuhn. The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2:83–97, 1955. doi: 10.1002/nav.3800020109.

Derek Lim, Joshua Robinson, Lingxiao Zhao, Tess Smidt, Suvrit Sra, Haggai Maron, and Stefanie Jegelka. Sign and basis invariant networks for spectral graph representation learning. In *International Conference on Learning Representations*, 2023. URL https://arxiv.org/abs/2202.13013.

An-Jun Liu and Bryan K. Clark. Neural network backflow for ab initio quantum chemistry. *Physical Review B*, 110:115137, 2024. doi: 10.1103/PhysRevB.110.115137. URL https://arxiv.org/abs/2403.03286.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019. URL https://arxiv.org/abs/1711.05101.

Aleksei Malyshev, Juan Miguel Arrazola, and A. I. Lvovsky. Autoregressive neural quantum states with quantum number symmetries. arXiv:2310.04166, 2023. URL https://arxiv.org/abs/2310.04166.

Aleksei Malyshev, Markus Schmitt, and A. I. Lvovsky. Neural quantum states and peaked molecular wave functions: Curse or blessing? arXiv:2408.07625, 2024. URL https://arxiv.org/abs/2408.07625.

Carlos Mejuto-Zaera and Alexander F. Kemper. Quantum eigenvector continuation for chemistry applications. *Electronic Structure*, 5:045007, 2023. doi: 10.1088/2516-1075/ad018f. URL https://iopscience.iop.org/article/10.1088/2516-1075/ad018f.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, volume 32, 2019. URL https://arxiv.org/abs/1912.01703.

Ethan Perez, Florian Strub, Harm de Vries, Vincent Dumoulin, and Aaron Courville. FiLM: Visual reasoning with a general conditioning layer. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018. URL https://arxiv.org/abs/1709.07871.

David Pfau, James S. Spencer, Alexander G. de G. Matthews, and W. M. C. Foulkes. Ab initio solution of the many-electron Schrödinger equation with deep neural networks. *Physical Review Research*, 2:033429, 2020. doi: 10.1103/PhysRevResearch.2.033429. URL https://arxiv.org/abs/1909.02487.

George D. Purvis, III and Rodney J. Bartlett. A full coupled-cluster singles and doubles model: The inclusion of disconnected triples. *The Journal of Chemical Physics*, 76:1910–1918, 1982. doi: 10.1063/1.443164.

Yannic Rath and George H. Booth. Interpolating numerically exact many-body wave functions for accelerated molecular dynamics. *Nature Communications*, 16:2005, 2025. doi: 10.1038/s41467-025-57134-9. URL https://www.nature.com/articles/s41467-025-57134-9.

Riccardo Rende, Luciano Loris Viteritti, Federico Becca, Antonello Scardicchio, Alessandro Laio, and Giuseppe Carleo. Foundation neural-networks quantum states as a unified ansatz for multiple Hamiltonians. *Nature Communications*, 16:7213, 2025. doi: 10.1038/s41467-025-62098-x. URL https://arxiv.org/abs/2502.09488.

Javier Robledo Moreno, Jeffrey Cohn, Dries Sels, and Mario Motta. Enhancing the expressivity of variational neural, and hardware-efficient quantum states through orbital rotations. arXiv:2302.11588, 2023. URL https://arxiv.org/abs/2302.11588.

Michael Scherbela, Leon Gerard, and Philipp Grohs. Towards a transferable fermionic neural wavefunction for molecules. *Nature Communications*, 15:120, 2024. doi: 10.1038/s41467-023-44216-9. URL https://www.nature.com/articles/s41467-023-44216-9.

Honghui Shang, Chu Guo, Yangjun Wu, Zhenyu Li, and Jinlong Yang. Solving the many-electron Schrödinger equation with a transformer-based framework. *Nature Communications*, 16:8464, 2025. doi: 10.1038/s41467-025-63219-2. URL https://www.nature.com/articles/s41467-025-63219-2.

Or Sharir, Yoav Levine, Noam Wies, Giuseppe Carleo, and Amnon Shashua. Deep autoregressive models for the efficient variational simulation of many-body quantum systems. *Physical Review Letters*, 124:020503, 2020. doi: 10.1103/PhysRevLett.124.020503. URL https://arxiv.org/abs/1902.04057.

Qiming Sun, Xing Zhang, Samragni Banerjee, Peng Bao, Marc Barbry, et al. Recent developments in the PySCF program package. *The Journal of Chemical Physics*, 153:024109, 2020. doi: 10.1063/5.0006074. URL https://arxiv.org/abs/2002.12531.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://papers.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.

Ingrid von Glehn, James S. Spencer, and David Pfau. A self-attention ansatz for ab-initio quantum chemistry. In *International Conference on Learning Representations*, 2023. URL https://arxiv.org/abs/2211.13672.

Yangjun Wu, Chu Guo, Yi Fan, Pengyu Zhou, and Honghui Shang. NNQS-Transformer: an efficient and scalable neural network quantum states approach for ab initio quantum chemistry. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, pp. 1–13, 2023. doi: 10.1145/3581784.3607061. URL https://arxiv.org/abs/2306.16705.

Yangjun Wu, Wanlu Cao, Jiacheng Zhao, and Honghui Shang. Fast and scalable neural network quantum states method for molecular potential energy surfaces. *IEEE Transactions on Parallel and Distributed Systems*, 36(7):1431–1443, 2025. doi: 10.1109/TPDS.2025.3568360. URL https://ieeexplore.ieee.org/document/11000098.

Tianchen Zhao, James Stokes, and Shravan Veerapaneni. Scalable neural quantum states architecture for quantum chemistry. *Machine Learning: Science and Technology*, 4:025034, 2023. doi: 10.1088/2632-2153/acdb2f. URL https://arxiv.org/abs/2208.05637.

SUMMARY OF THE APPENDIX

The appendix is organized as follows:

- Appendix A describes orbital transformations, alignment protocols, and query attachment.
- Appendix B specifies the network, optimizer, sampling, and energy-evaluation settings.
- Appendix C defines the molecular paths, retained orbital spaces, and electronic observables.
- Appendix D describes the accompanying code and data.
- Appendix E reports detailed energy and property results, per-seed ablations, and computational costs.

A ORBITAL ALIGNMENT AND QUERY ATTACHMENT

A.1 ORBITAL TRANSFORMATIONS

At fixed geometry, $\widetilde{a}_p^\dagger = \sum_q a_q^\dagger U_{qp}$ induces the exterior-power representation of U on determinants. In a basis with separately ordered alpha and beta determinants, its matrix element between occupied sets is

$$\mathcal{U}_{I,J} = \det U_{I_\alpha,J_\alpha} \det U_{I_\beta,J_\beta}. \tag{7}$$

Conversion to interleaved spin-orbital order includes the fermionic ordering signs. Signed permutations therefore relabel determinants with the corresponding parity; general rotations mix them through these minors. For any retained-space unitary, $\mathcal{U}$ is unitary and

$$(\mathcal{U}^\dagger\psi)^\dagger(\mathcal{U}^\dagger H\mathcal{U})(\mathcal{U}^\dagger\psi) = \psi^\dagger H\psi. \tag{8}$$

Rotations are confined to the retained orbital space; frozen-core and excluded orbitals define separate boundaries. Across different geometries, the AO functions themselves move, so cross-AO overlaps enter the matching step.

A.2 ALIGNMENT PROTOCOLS

The N_2/CO construction transports the orbital bases at training anchors from a fixed root, then aligns each query directly to its nearest aligned training anchor. The H_4 endpoint RHF solution is continued from the lower canonical branch in the common C_{2h} subgroup; query attachments are recorded in the supplementary orbital-connection table.

The supplementary file `results/orbital-connections.json` specifies orbital-transport parents and additional RHF support, including query geometries used during orbital preparation.

For NH_3, orbital preparation resolves parity under a fixed physical reflection, $z \mapsto -z$ in the laboratory frame. The reflection matrix is diagonalized within exactly degenerate Fock blocks (energy tolerance 10^{-7} Ha), keeping occupied/virtual and frozen-core boundaries separate. This step rotates degenerate orbitals into a common parity convention before maximum-overlap matching and sign alignment. The orbital bases at the five anchors are transported from the planar root, and each query is attached to its nearest anchor.

B OPTIMIZATION AND SAMPLING DETAILS

B.1 NETWORK AND OPTIMIZER SETTINGS

Electronic-structure calculations use PySCF (Sun et al., 2020), and neural-network models are implemented in PyTorch (Paszke et al., 2019).

Production models use four amplitude blocks, four heads, and four phase-network layers of width 512 with rectified linear unit activations. The amplitude width is 32 except for ethylene, which uses width 64. Occupations are decoded in reverse orbital order; descriptor and irrep arrays follow that same order. The phase branch concatenates the geometry condition $\mathbf{c}(\mathbf{R})$ with the signed spin-occupation vector $\mathbf{s}(\mathbf{n})$. Diatomics use bond lengths in angstroms as geometry features; the other molecular

paths rescale their coordinate to $[-1, 1]$ using fixed domain bounds. Spin populations are fixed in all runs; the N_2 model uses no spatial-irrep mask, while the other five systems impose the prescribed total irrep. Dropout is zero. The FiLM maps in Equation (3) have zero initial weights and biases, with no bias in the descriptor map. The output head applies a final LayerNorm followed by the linear projection. AdamW (Loshchilov & Hutter, 2019) uses a base learning rate 10^{-3}, betas $(0.9, 0.99)$, epsilon 10^{-9}, and zero weight decay. Computation uses float32 model parameters.

The displayed models use seed 11 and 20,000 updates. N_2 uses constant learning rate 10^{-3}. CO, H_4, NH_3, and ethylene use cosine decay from 10^{-3} to 10^{-5} without warmup. CO_2 uses 3×10^{-4} through update 6,000, 10^{-4} through 15,000, and 3×10^{-5} through 20,000. Training uses two A100 GPUs except for ethylene (four A100s) and NH_3, which uses one V100 through update 15,000 and three A100s thereafter. All final models use FiLM after each of the four Transformer blocks. The orbital-alignment and conditioning ablations use the corresponding N_2/CO schedules and seeds 11, 23, and 37.

Table 1: Training budgets and per-geometry distinct-configuration targets.

Study	Steps	Geometries/update	Distinct configurations
N_2	20,000	5	6,000–50,000
CO	20,000	5	600–2,400
H_4/cc-pVDZ	20,000	4	2,000–4,000
NH_3	20,000	5	6,000–20,000
Ethylene	20,000	7	6,000–20,000
CO_2	20,000	5	6,000–20,000

B.2 Sampled-subspace training and full-state evaluation

The native Hamiltonian backend constructs in-set single- and double-excitation connections. Network probabilities, normalized over S_k, weight its energy and gradient; BAS multiplicities determine set membership. Holding S_k fixed during differentiation gives

$$\nabla_\theta E_{S_k} = 2\,\mathrm{Re} \sum_{x \in S_k} \frac{|\psi_\theta(x)|^2}{\sum_{y \in S_k} |\psi_\theta(y)|^2} \left[\frac{(H_{S_k S_k} \psi_{S_k})_x}{\psi_\theta(x)} - E_{S_k} \right] \nabla_\theta \log \psi_\theta(x)^*. \tag{9}$$

The sampled set is refreshed between updates. Related sampled-subspace objectives are described by Wu et al. (2023); Malyshev et al. (2024).

For complete-state reporting, BAS multiplicities c_x retain the frequency of every distinct sampled configuration. The energy estimate is

$$\widehat{E} = \frac{\sum_x c_x \,\mathrm{Re}\, E_{\mathrm{loc}}(x)}{\sum_x c_x}. \tag{10}$$

The local-energy sum includes connected states outside S, with their amplitudes evaluated by the same frozen network. Three independent runs at each geometry give sampling replicates. The standard error of the mean is the larger of the within-replicate propagated error, $\sqrt{\sum_{r=1}^{3} \sigma_r^2}/3$, and the between-replicate sample standard deviation divided by $\sqrt{3}$. Relative-energy uncertainties include the point and reference-origin variances in quadrature.

C Molecular paths and property conventions

Diatomics and H_4. N_2 spans 0.60–1.40 Å and CO spans 0.80–1.60 Å, both at 0.02 Å reporting spacing and with five anchors at 0.20 Å spacing. Each grid contains five anchors and 36 non-anchor queries. H_4 has coordinates $(\pm x, \pm y, 0)$ with $x = R\cos(\theta/2)$, $y = R\sin(\theta/2)$, and $R = 3.2843\,a_0$. Its 25-point grid spans $84°$–$90°$ in $0.25°$ steps, with anchors $84°, 86°, 88°, 90°$, and 21 non-anchor queries. The C_{2h}/A_g convention continues the RHF density from $89.75°$ to the square endpoint.

Table 2: Molecular systems. Electron/orbital counts refer to the retained space.

System	Basis	e^-/orbitals	Coordinate	Anchors
N_2	STO-3G	14/10	bond length	5
CO	STO-3G	14/10	bond length	5
H_4 rectangle	cc-pVDZ	4/20	rectangle angle	4
NH_3	6-31G, frozen core	8/14	umbrella height	5
CO_2	STO-3G, frozen core	16/12	C–O distance	5
C_2H_4	STO-3G, frozen core	12/12	torsion angle	7

NH_3. The coordinate q is the nitrogen height above the H_3 plane at fixed N–H distance 1.02 Å. The five anchors are $q = 0, 0.15, 0.30, 0.45, 0.60$ Å, within a uniform 25-point reporting grid. A common vertical C_s reflection is used at every geometry, including the planar endpoint. N $1s$ is frozen, leaving eight electrons in 14 orbitals. The dipole axis points from the H_3 plane toward positive N height.

CO_2. The molecule is linear with carbon at the origin and oxygen atoms at $z = \pm r$, with $r \in [1.05, 1.35]$ Å. Five anchors at 0.075 Å spacing define the aligned orbital bases; the 25-point reporting grid has 0.0125 Å spacing. Matching preserves D_{2h} irreps and the core, occupied, and virtual orbital classes. Freezing the C and O $1s$ cores leaves 16 electrons in 12 STO-3G orbitals. The reported traceless axial quadrupole is $Q_{zz} = \frac{1}{2}\sum_a q_a(3z_a^2 - r_a^2)$, including electronic, frozen-core, and nuclear contributions, in ea_0^2.

Ethylene. The fixed planar structure has C=C distance 1.339 Å, C–H distance 1.087 Å, and C–C–H angle $121.3°$. Symmetric fragment rotations by $\pm\theta/2$ define $\theta \in [0, 90°]$. Seven Chebyshev–Lobatto anchors are $0, 6.028857, 22.5, 45, 67.5, 83.971143, 90°$. The reporting grid comprises 33 uniformly spaced angles and the two additional anchors at $6.028857°$ and $83.971143°$, giving 35 points and 28 queries. All seven anchors share the orbital labels and phase conventions of the initial five-anchor basis. Both carbon $1s$ cores are frozen, leaving 12 electrons in 12 orbitals in a common C_2 subgroup. The displayed frontier occupations are the sixth and seventh eigenvalues of the active spin-summed 1-RDM, sorted in descending order; their π/π^* character is checked against the reference natural orbitals.

Electronic observables. The spin-summed 1-RDM is Hermitized before diagonalization, and its trace equals the retained electron count. Natural occupations are sorted in descending order. Dipoles include electronic, frozen-core, and nuclear contributions with a fixed origin convention. Property MAEs use all non-anchor reporting points; ethylene metrics average occupations 6 and 7. RHF and CCSD (Purvis & Bartlett, 1982) use the same geometries, retained spaces, origins, and operators as NQS and FCI. CCSD properties are obtained by contracting the relevant operators with the unrelaxed spin-summed one-particle density matrix from converged amplitude and Λ equations. Occupied and virtual orbitals are separately canonicalized while retaining the RHF determinant. All reporting geometries have converged solutions.

D Reproducibility materials

The accompanying materials provide geometry splits and transport records, selected experiment configurations, FCI reference values, pointwise energy and property estimates, and timing records. Scripts recompute reported summaries and regenerate all manuscript figures from these recorded results. Orbital-alignment code and conditioning adapters document the method implementation.

E SUPPORTING RESULTS

E.1 FROZEN ENERGY ACCURACY

Table 3 reports the complete anchor and query errors behind Fig. 2. Table 4 and Fig. 6 give the energies of the same checkpoints used for the three property panels. Every query in Table 3 is within 1.6 mHa of FCI.

Table 3: Frozen molecular predictions, in mHa relative to FCI. Anchor and query metrics use the same final checkpoint.

System	Molecular path	Anchors / queries	Anchor MAE	Query MAE	Query max.
N_2	bond stretch	5 / 36	0.036	0.049	0.095
CO	bond stretch	5 / 36	0.197	0.253	0.585
H_4	rectangle	4 / 21	0.088	0.076	0.120

E.2 MOLECULAR-PROPERTY BENCHMARKS

Table 5 quantifies the property comparisons in Fig. 4. Each metric uses all non-anchor reporting points; occupation errors average over both displayed eigenvalues.

E.3 PER-SEED COMPONENT ABLATIONS

Table 6 gives the individual runs underlying Figs. 3 and 7. Alignment improves N_2 query accuracy for every paired seed. All four CO conditioning variants reach chemical accuracy; their differences are small compared with the orbital-alignment effect. The Transformer and phase-network backbones, sampler settings, and optimization budgets are matched within each comparison, with seeds 11, 23, and 37.

Conditioning architecture in the aligned orbital basis. In the aligned orbital basis, Fig. 7 compares four ways to condition the amplitude. The $\mathbf{c}$ input baseline adds an affine embedding of the geometry condition $\mathbf{c}$ (the bond length for CO) to the token and position embeddings; $\mathbf{c} + \mathbf{d}$ input additionally supplies a linear projection of the orbital descriptor d_j. Their FiLM counterparts apply these signals after each Transformer block, as in Equation (3). Final 20,000-update checkpoints are evaluated on eight CO query geometries over three seeds.

FiLM lowers the mean query MAE from 0.427 to 0.303 mHa with geometry alone. Descriptors have a smaller effect at the input, but combining them with FiLM gives the lowest mean MAE, 0.284 mHa. These differences are comparable to seed variation and smaller than the alignment effect. We use $\mathbf{c} + \mathbf{d}$ FiLM in the final model.

E.4 COMPUTATIONAL COST PROTOCOL AND ESTIMATES

We compare four fixed-budget workflows on the N_2 bond scan. Table 7 gives the budgets, MAEs on the same eight geometries, and total cost estimates. All energies use three independent full-

Table 4: Energy errors for the three molecular-property benchmarks in Fig. 4, in mHa relative to FCI. Anchor and query metrics use the same final checkpoints as the property comparisons; queries include all non-anchor grid points.

System	Molecular path	Anchors / queries	Anchor MAE	Query MAE	Query max.
C_2H_4	torsion	7 / 28	0.919	0.946	1.233
NH_3	inversion	5 / 20	0.872	1.036	1.456
CO_2	symmetric stretch	5 / 20	0.591	0.535	0.895

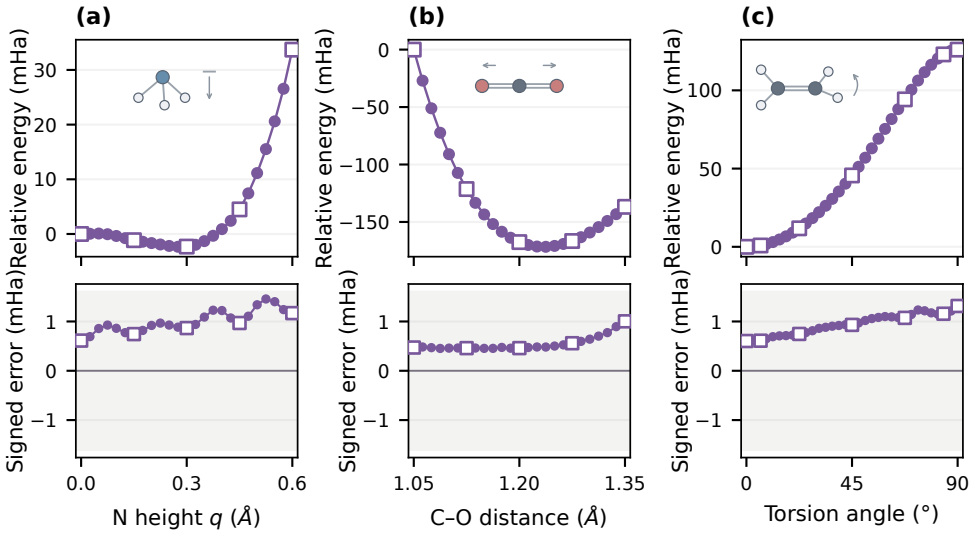


Figure 6: **Frozen energy predictions.** (a) NH_3 inversion, (b) CO_2 symmetric stretching, and (c) ethylene torsion. Each panel combines frozen NQS energies relative to each path origin (upper) and the corresponding signed complete-Hamiltonian errors against FCI (lower), with the ± 1.6 mHa band. Hollow squares mark anchors; dots mark queries. Error bars show one sampling standard error. The corresponding property panels appear in Fig. 4.

Table 5: Query property MAEs against FCI. All methods use identical geometries, retained spaces, and observable conventions.

Observable	RHF	CCSD	NQS
NH_3 dipole (D)	0.06057	0.00575	0.0058
CO_2 quadrupole (ea_0^2)	1.07714	0.06971	0.0403
C_2H_4 occupations	0.27589	0.00485	0.0009

Table 6: Per-seed component ablations (MAE in mHa). N_2 uses all 36 query geometries; CO uses the eight query geometries. All runs use 20,000 updates with post-block FiLM where applicable; N_2 uses a fixed learning rate of 10^{-3} and CO uses cosine decay. Each cell reports the error of a fixed final checkpoint.

Model	Seed 11	Seed 23	Seed 37
N_2 unaligned	36.806	33.609	35.027
N_2 aligned	0.049	0.085	0.053
CO $\mathbf{c}$ input	0.292	0.414	0.575
CO $\mathbf{c}$ FiLM	0.328	0.333	0.247
CO $\mathbf{c} + \mathbf{d}$ input	0.364	0.438	0.443
CO $\mathbf{c} + \mathbf{d}$ FiLM	0.277	0.194	0.382

Hamiltonian evaluation replicates. The cost checkpoints are separate from the 20,000-update models used in the main energy results.

Training budgets. Independent optimization starts each geometry from random initialization and uses AdamW for 4,000 updates, with 100 updates of linear warmup to 3×10^{-4} and cosine decay

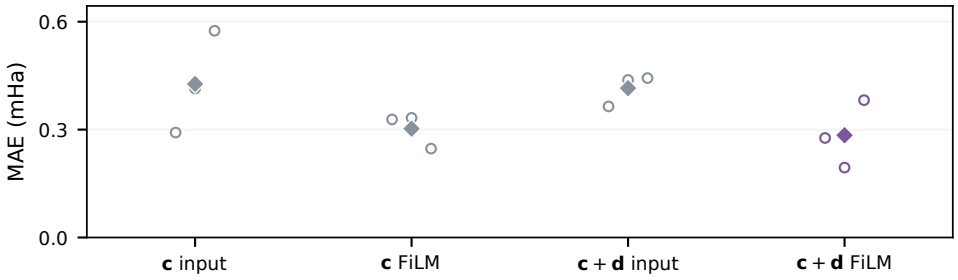


Figure 7: **Conditioning ablations.** All variants use the aligned CO orbital basis. Open circles show query MAEs for three seeds and filled diamonds their means. The geometry condition $\mathbf{c}$ and orbital descriptors $\mathbf{d}$ enter either the input embeddings or post-block FiLM adapters. The backbone, phase network, training budget, and evaluation geometries are matched.

Table 7: Fixed-budget N_2 comparison. MAEs use the common eight geometries. Costs include training and final energy evaluation; independent and shared costs are estimated from the sampled geometries. The two warm-start MAEs correspond to the 41- and 161-point grids, respectively.

Workflow	Updates	MAE (mHa)	41 points (GPU-h)	161 points (GPU-h)
Independent	4,000	1.027	10.785	42.351
Shared + fine-tune	2,000 + 200/point	1.064	1.682	4.162
Sequential warm-start	3,000 + 100/next point	1.173 / 0.988	0.664	1.948
Joint + frozen	4,000	0.793	1.612	1.644

to 3×10^{-5} at update 4,000. Joint training uses five anchors and a constant learning rate of 10^{-3} for 4,000 updates. Shared pretraining removes geometry conditioning and minimizes the equally weighted mean of the five anchor losses for 2,000 updates at 3×10^{-4}. Each query then receives 200 fine-tuning updates with a fresh optimizer. Sequential warm starts optimize the first geometry for 3,000 updates, then transfer model weights to the next geometry with a fresh optimizer for 100 updates. The first-point schedule uses cosine decay from 10^{-3} toward 10^{-5} over 20,000 updates, without warmup. Shared fine-tuning and subsequent warm-start points use 200-update linear warmup toward 10^{-4} and cosine decay toward 10^{-5} over 10,000 updates. All workflows use the same aligned orbital basis and full-Hamiltonian evaluation protocol.

Hardware and accounting. Independent optimization and adaptation use one A100; joint training and shared pretraining use two. GPU-hours sum training, process initialization, checkpoint loading, and final full-Hamiltonian evaluation across allocated GPUs; CPU orbital and integral preparation is outside this GPU-cost accounting. Shared pretraining and joint training cost 0.835 and 1.601 GPU-hours, respectively. The frozen-evaluation slope in Fig. 5 is 0.961 GPU-seconds per geometry, measured over the 161-point grid with three sampling replicates per geometry.

Timing geometries and accuracy. Independent optimization and shared fine-tuning are timed at the same eight geometries selected by stratified sampling (Table 8); per-geometry costs are weighted by the fraction of the grid represented by each stratum. All four workflows report MAE on these common points. Sequential warm starts traverse the 41- and 161-point grids in ascending bond length, with full-grid MAEs of 1.182 and 0.993 mHa; frozen evaluation gives 0.823 and 0.822 mHa, respectively.

Query-count scaling. For each workflow, Equation (6) holds its measured average per-geometry cost fixed. For warm starts, the intercept is the first-point training and evaluation cost minus one subsequent-point cost, so that $G = 1$ recovers the first-point cost exactly. The two warm-start curves

Table 8: Absolute full-Hamiltonian energy errors (mHa) at the eight common N_2 comparison geometries. Warm-start columns correspond to the two complete sequential paths.

R (Å)	Independent	Shared + fine-tune	Warm (41)	Warm (161)	Joint + frozen
0.60	0.284	0.458	0.365	0.365	0.327
0.76	0.663	0.473	0.559	0.458	0.527
0.84	1.054	0.601	0.705	0.566	0.573
0.96	1.097	0.728	0.898	0.787	0.723
1.00	0.709	0.790	0.998	0.872	0.788
1.18	1.868	1.477	1.666	1.392	1.088
1.28	1.755	1.864	2.019	1.669	1.138
1.32	0.782	2.123	2.176	1.795	1.176

use the mean costs measured at their respective grid spacings. Writing the joint-training cost as a_{frozen} and the per-geometry evaluation cost as b_{frozen}, the estimated speedup over joint training with frozen evaluation is

$$S_m(G) = \frac{a_m + G b_m}{a_{\mathrm{frozen}} + G b_{\mathrm{frozen}}}, \qquad \lim_{G\to\infty} S_m(G) = \frac{b_m}{b_{\mathrm{frozen}}}. \tag{11}$$

The slope ratios are 986 for independent optimization, 77.4 for shared fine-tuning, and 40.8–43.0 for sequential warm starts. The 41- and 161-point grids have been evaluated for joint and sequential workflows; curves at other query counts are cost-model projections.