MōLe-Λ: Learning the Coupled-Cluster Response State for Energies, Gradients, and Properties

Andreas Burger [1 2 3] **Luca Thiede** [1 2 3] **Abdulrahman Aldossary** [3] **Jorge Arturo Campos-Gonzalez-Angulo** [1 2]
Alexander Zook [3] **Jérôme Florian Gonthier** [3] **Alán Aspuru-Guzik** [1 2 4 5]

Abstract

Coupled-cluster (CC) theory is often considered the gold standard of quantum chemistry, but its high computational cost limits routine access to accurate energies, forces and response properties. While the right-hand T-amplitudes determine the correlated wavefunction, many practically important observables additionally require the left-hand Λ-amplitudes. We introduce MōLe-Λ, an extension of Molecular Orbital Learning (MōLe) that predicts the full ground-state coupled-cluster singles and doubles (CCSD) response state by jointly learning right-hand amplitudes (T_1, T_2) and left-hand amplitudes (Λ_1, Λ_2) from localized Hartree–Fock molecular orbitals. Architecturally, MōLe-Λ extends MōLe with Λ_1 and Λ_2 readouts that mirror the symmetry constraints of the T_1 and T_2 heads, while preserving the original equivariant orbital encoder, odd sign-equivariant decoding, locality and size-extensivity. The resulting model yields accurate CC-quality energies and forces, while simultaneously recovering dipoles, quadrupoles, polarizabilities, the electron density, and 2-electron observables such as the pair density. We show that MōLe-Λ further extends the speed advantage of MōLe over full CCSD while substantially expanding the accessible properties, providing a route to wavefunction-level surrogate models for correlated quantum chemistry.

[1]University of Toronto, Canada [2]Vector Institute for Artificial Intelligence, Toronto, Canada [3]NVIDIA, Toronto, Canada [4]Acceleration Consortium, Toronto, Canada [5] Senior Fellow, Canadian Institute for Advanced Research (CIFAR), Toronto, Canada . Correspondence to: Andreas Burger <me@andreas-burger.com>.

Proceedings of the AI4Physics Workshop at the 43rd International Conference on Machine Learning (AI4Physics@ICML 2026), Seoul, South Korea. 2026. Copyright 2026 by the author(s).

1. Introduction

The ability to predict molecular properties accurately and efficiently is central to computational chemistry. Density functional theory (DFT) provides an attractive compromise between cost and accuracy and therefore underlies much of modern atomistic simulation (Kohn & Sham, 1965; Mardirossian & Head-Gordon, 2017). For applications that require substantially higher fidelity, however, correlated wavefunction methods remain essential (Szabo & Ostlund, 2012; Helgaker et al., 2013). In particular, coupled-cluster theory is the standard route to chemically accurate ground-state properties, but its steep computational scaling of $O(N^6)$ limits routine use beyond small systems (Purvis III & Bartlett, 1982; Raghavachari et al., 1989; Bartlett & Musiał, 2007).

Related work Recent machine learning work has narrowed this accuracy–efficiency gap from several directions. Machine-learned interatomic potentials can predict energies and forces at very low cost (Smith et al., 2020b; Mehdizadeh & Schindler, 2025; Martyka et al., 2025; Burger et al., 2025; Chen & Dral, 2025; Banchode et al., 2025). A separate line of work learns densities, density matrices, Hamiltonians, or Fock-like intermediates to improve one-particle observables or accelerate self-consistent electronic-structure calculations (Brockherde et al., 2017; Shao et al., 2023; Hazra et al., 2024; Febrer et al., 2025; Elsborg et al., 2025; Jørgensen & Bhowmik, 2022; Yu et al., 2023; Luo et al., 2025; Kim et al., 2025b; Li et al., 2025b). More recent work has explored machine learning targets beyond standard DFT observables, including coupled-cluster properties, Green's functions, and wavefunction-inspired representations (Townsend & Vogiatzis, 2019; 2020; Venturella et al., 2024; 2025; Shang et al., 2025; Gao & Günnemann, 2023; Scherbela et al., 2024).

These approaches have been highly effective for their intended tasks, but they do not directly parameterize the coupled-cluster response state. Learning energies and forces is attractive for molecular dynamics and screening, but does not recover the underlying correlated electronic state. Learning densities or density matrices provides access to one-

particle observables, while learning Hamiltonians can accelerate self-consistent field procedures. Yet many chemically relevant quantities, including response densities, multipole moments, polarizabilities, and energy gradients, ultimately depend on the correlated wavefunction.

A particularly promising abstraction level is therefore the set of coupled-cluster amplitudes. Molecular Orbital Learning (MoLe) (Thiede et al., 2026), showed that learning the right-hand T-amplitudes directly from Hartree–Fock molecular orbitals yields accurate CC energies, accurate one-particle densities, and useful warm starts for CC solvers, especially in the low-data regime where expensive CC labels are scarce. However, the right-hand amplitudes alone do not fully characterize the relaxed response properties of the CC state (Eriksen et al., 2014; Helgaker et al., 2013). In Lagrangian formulations of CC theory, these properties additionally require the left-hand de-excitation amplitudes Λ. This observation motivates the central idea of the present work. Once the T and Λ amplitudes are available, one can recover not only energies but also analytic gradients, multipoles, polarizabilities, coupled-cluster reduced density matrices, natural occupations, and two-particle observables from a single learned electronic object.

Present work In this work, we introduce MōLe-Λ, an extension of MōLe that predicts both T and Λ amplitudes from localized Hartree–Fock molecular orbitals. Architecturally, MōLe-Λ preserves the symmetry-aware orbital representation of MōLe and augments it with Λ_1 and Λ_2 readouts that mirror the T_1 and T_2 heads. Empirically, we show that MōLe-Λ substantially broadens the practical utility of amplitude learning. The model yields accurate correlated energies and forces competitive with strong MLIP baselines, while simultaneously recovering higher-order molecular properties that standard energy models cannot access directly. In particular, we demonstrate accurate dipoles, quadrupoles, polarizabilities, Λ-state 1-RDMs and electron densities, as well as the 2-RDM-derived pair density. We further show that MōLe-Λ obtains the T and Λ amplitudes over two orders of magnitude faster than full CCSD + Λ-equations, thereby moving beyond property-specific surrogates toward a unified machine-learned representation of correlated wavefunction response.

2. Background and Preliminaries

Hartree–Fock We begin from restricted Hartree–Fock molecular orbitals (MOs) $\psi_p(\mathbf{r} \mid \{\mathbf{R}_A\})$. Each spatial orbital is expanded in an atomic-orbital basis as

$$\psi_p(\mathbf{r}) = \sum_A \sum_{k\in\mathcal{K}_A} \sum_{\ell\in\mathcal{L}_{A,k}} \sum_{m=-\ell}^{\ell} C_{pA,k\ell m}\, \phi^m_{k\ell}(\mathbf{r}-\mathbf{R}_A), \tag{1}$$

where A labels atoms with positions $\mathbf{R}_A$, k, ℓ, m denote the principal, azimuthal, and magnetic quantum numbers, and $C_{pA,k\ell m}$ are elements of the MO coefficient matrix $\mathbf{C}$. Throughout the paper, p, q index all orbitals, i, j index occupied orbitals, and a, b index virtual orbitals. The MO coefficients are obtained by solving the Roothaan–Hall equations

$$\mathbf{F}(\mathbf{C})\mathbf{C} = \mathbf{S}\mathbf{C}\boldsymbol{\varepsilon} \tag{2}$$

with $\mathbf{F}(\mathbf{C})$ the Fock matrix, $\mathbf{S}$ the atomic-orbital overlap matrix, $\mathbf{C}$ the MO coefficient matrix, and $\boldsymbol{\varepsilon}$ the diagonal matrix of orbital energies. Since $\mathbf{F}$ depends on $\mathbf{C}$, this equation is solved iteratively.

Localized molecular orbitals Canonical molecular orbitals are generally delocalized over the full molecule. We therefore work in a separately localized occupied and virtual gauge,

$$\tilde{\psi}_i = \sum_{j\in\mathrm{occ}} (U_{\mathrm{occ}})_{ji}\psi_j, \qquad \tilde{\psi}_a = \sum_{b\in\mathrm{virt}} (U_{\mathrm{virt}})_{ba}\psi_b, \tag{3}$$

where U_{occ} and U_{virt} are unitary transformations within the occupied and virtual subspaces, respectively. This transformation leaves all observables invariant, while exposing the fragment structure of the amplitudes and thereby providing an important inductive bias for transfer across molecular size and chemistry. Unless noted otherwise, all subsequent orbital-indexed quantities are expressed in this localized MO basis.

Coupled-cluster right and left states In coupled-cluster theory, the correlated right-hand wavefunction is written as

$$|\Psi_{\mathrm{CC}}\rangle = e^{\hat{T}}\,|\Phi_{\mathrm{HF}}\rangle\,, \qquad \hat{T} = \hat{T}_1 + \hat{T}_2, \tag{4}$$

where truncation to singles and doubles defines CCSD. Using standard spin-orbital notation for the second-quantized operators, with the restricted closed-shell implementation represented by the corresponding spin-adapted spatial-orbital tensors, the cluster operators are

$$\hat{T}_1 = \sum_{ia} t_i^a\, \hat{a}_a^\dagger \hat{a}_i, \quad \hat{T}_2 = \frac{1}{4}\sum_{ijab} t_{ij}^{ab}\, \hat{a}_a^\dagger \hat{a}_b^\dagger \hat{a}_j \hat{a}_i. \tag{5}$$

Here, t_i^a and t_{ij}^{ab} are the singles and doubles amplitudes. They are obtained from a nonlinear system of equations, and solving the CCSD equations formally scales as $\mathcal{O}(N^6)$ (Purvis III & Bartlett, 1982; Szabo & Ostlund, 2012). The corresponding correlation energy can be computed from only T_1 and T_2. For relaxed observables and response properties, the T amplitudes alone are insufficient. Unlike variational methods, CC theory is not stationary with respect to the right amplitudes alone, so the total derivative $dE_{\mathrm{CC}}/d\xi$ contains extra terms $\partial t_\mu/\partial\xi$ for each perturbation ξ and

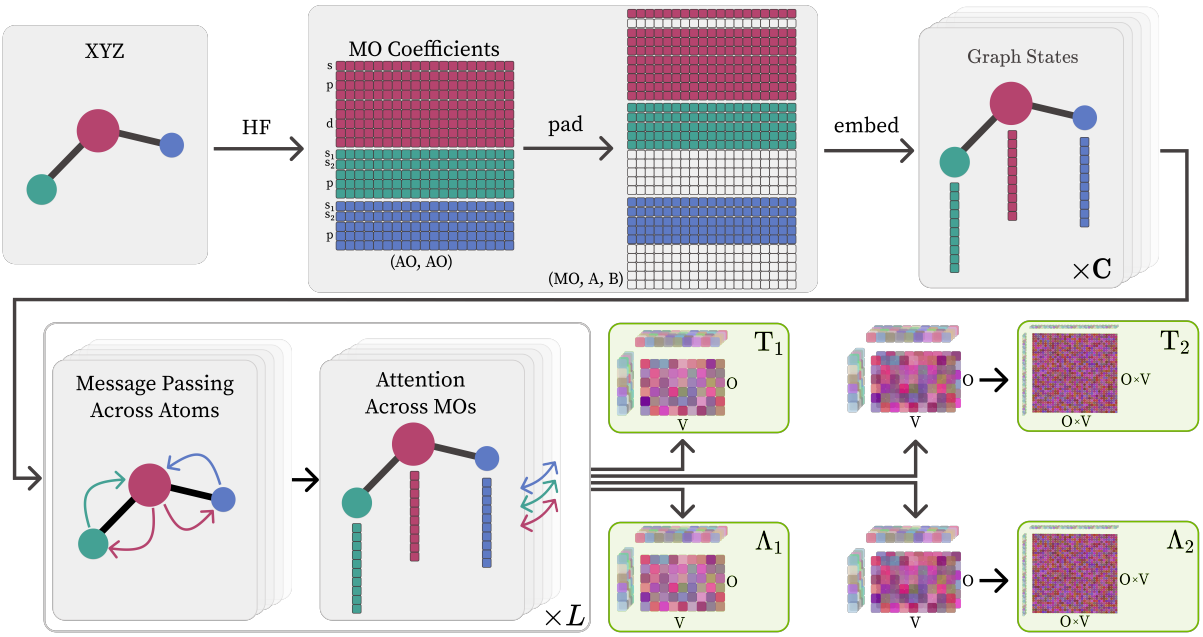

Figure 1. Given a molecule, a Hartree-Fock calculation provides the molecular orbitals (MOs) represented by their coefficient matrix $\mathbf{C}$. The coefficient vector is padded for each atom to ensure that all atoms have the same number of basis coefficients, and embedded in an equivariant neural network. The model then alternates (a) message passing to mix information within the MOs and (b) attention layers to mix information between MOs. Finally, the embeddings are read out with specialized readout heads for the $(T_1, T_2, \Lambda_1, \Lambda_2)$ amplitudes.

excitation μ that would need to be computed. The coupled-cluster Lagrangian avoids this by introducing Λ as an adjoint variable (Koch & Jørgensen, 1990)

$$\mathcal{L}(T, \Lambda) = \langle \Phi_{\text{HF}} | \, (1 + \hat{\Lambda}) e^{-\hat{T}} \hat{H} e^{\hat{T}} \, | \Phi_{\text{HF}} \rangle \,, \tag{6}$$

with the adjoint left state $\langle \widetilde{\Psi}_{\text{CC}} |$, and the de-excitation operators $\hat{\Lambda}_1, \hat{\Lambda}_2$

$$\langle \widetilde{\Psi}_{\text{CC}} | = \langle \Phi_{\text{HF}} | \, (1 + \hat{\Lambda}) e^{-\hat{T}}, \qquad \hat{\Lambda} = \hat{\Lambda}_1 + \hat{\Lambda}_2 \tag{7}$$

$$\hat{\Lambda}_1 = \sum_{ia} \lambda_a^i \, \hat{a}_i^\dagger \hat{a}_a, \quad \hat{\Lambda}_2 = \frac{1}{4} \sum_{ijab} \lambda_{ab}^{ij} \, \hat{a}_i^\dagger \hat{a}_j^\dagger \hat{a}_b \hat{a}_a \tag{8}$$

The introduction of Λ simplifies the derivative of a non-variational CC energy into the partial derivative of a stationary Lagrangian (Bartlett & Musiał, 2007)

$$\frac{dE}{d\xi} = \frac{\partial \mathcal{L}(T, \Lambda; \xi)}{\partial \xi}, \tag{9}$$

where $T = \{t_i^a, t_{ij}^{ab}\}$ and $\Lambda = \{\lambda_a^i, \lambda_{ab}^{ij}\}$ denote the CCSD amplitude tensors, while $\hat{T}$ and $\hat{\Lambda}$ denote the corresponding second-quantized operators. In conventional CCSD, the T amplitudes have to be solved first, after which the Λ amplitudes can be obtained from the linear adjoint equations

$$\frac{\partial \mathcal{L}}{\partial t_\mu} = \frac{\partial E_{\text{CC}}}{\partial t_\mu} + \sum_\nu \lambda_\nu \frac{\partial R_\nu}{\partial t_\mu} = 0. \tag{10}$$

Unfortunately, these CCSD Λ-equations are dominated by the same type of four-virtual/two-occupied contractions that dominate CCSD, with a scaling of $\mathcal{O}(N^6)$. If the amplitudes (T, Λ) can be made accessible cheaply, however, they form the CCSD response-state required for coupled-cluster density matrices, analytic gradients, and response properties, making them an attractive learning target.

3. MōLe-Λ

We now present MōLe-Λ, which extends Molecular Orbital Learning (MōLe) (Thiede et al., 2026) to predict both the right-hand coupled-cluster amplitudes $T = (T_1, T_2)$ and the left-hand adjoint amplitudes $\Lambda = (\Lambda_1, \Lambda_2)$ directly from localized Hartree–Fock molecular orbitals. Our architecture design follows the same philosophy as MōLe, now applied to the full response state. The model should satisfy four constraints:

1. Molecular-orbital coefficients are rotation equivariant
2. Predicted amplitudes are rotationally invariant
3. Predicted amplitudes are sign equivariant with respect to orbital phase flips
4. Amplitudes coupling localized orbitals on non-interacting fragments should vanish

	Validation		Size extrapolation				Out-of-equilibrium					
	QM7		Amino acids		PubChem		Diels-Alder		Dihedral scan		Chair-to-boat	
	E	F	E	F	E	F	E	F	E	F	E	F
MP2	57.32	1.50	60.49	1.33	82.55	1.32	69.33	1.18	60.99	0.57	75.03	0.54
Mace	0.79	1.20	9.03	9.99	19.45	9.44	11.25	7.99	2.92	1.03	3.19	1.87
Mace+MP2	0.16	0.23	0.51	1.90	2.07	2.49	1.61	1.43	0.34	0.90	0.41	0.83
eSEN	0.74	0.89	11.26	6.70	31.53	7.54	16.39	12.08	2.14	7.07	2.89	5.43
eSEN+MP2	0.15	0.17	3.20	0.69	8.12	1.81	1.81	1.94	0.63	0.17	0.68	0.19
MōLe-Λ	**0.10**	**0.12**	**0.37**	**0.27**	**0.63**	**0.26**	**1.09**	**0.24**	**0.29**	**0.12**	**0.40**	**0.13**

Table 1. Mean absolute error of energy and force predictions in mHa and mHa/Bohr.

With these constraints in mind, MōLe-Λ first runs a fast HF calculation to obtain MOs, performs orbital localization, passes the localized MOs into a shared equivariant encoder and backbone, and finally predicts $(T_1, T_2, \Lambda_1, \Lambda_2)$ via separate readout heads, as depicted in Fig. 1. The output amplitudes are subsequently passed to conventional coupled-cluster post-processing routines to obtain energies, forces, and response observables.

Shared equivariant encoder We reuse the orbital-centric representation of MōLe. Each localized orbital is embedded into an equivariant latent representation, one graph per MO, and processed by the same backbone of molecular-orbital attention, Odd-MACE updates, and equivariant normalization (Qin et al., 2022; Batatia et al., 2022; Thiede et al., 2026). The encoder is shared across occupied and virtual orbitals and across all four readout heads.

Amplitude readout The original MōLe readout produces singles amplitudes from pairwise invariant features $\mathbf{y}_{ia}$ and doubles amplitudes from four-orbital invariant features $\mathbf{y}_{ijab}$. We retain that structure for the T amplitudes, and mirror it for the left state,

$$t_i^a = \mathrm{OddReadout}_{T_1}(\mathbf{y}_{ia}), \tag{11}$$

$$t_{ij}^{ab} = \mathrm{OddReadout}_{T_2}(\mathbf{y}_{ijab}) \tag{12}$$

$$\lambda_a^i = \mathrm{OddReadout}_{\Lambda_1}(\mathbf{y}_{ia}), \tag{13}$$

$$\lambda_{ab}^{ij} = \mathrm{OddReadout}_{\Lambda_2}(\mathbf{y}_{ijab}) \tag{14}$$

The odd readout ensures sign equivariance, while the invariant features guarantee rotational invariance of the predicted amplitudes.

Direct and residual targets In the present restricted closed-shell systems with real amplitudes, the lowest-order perturbative structure provides a natural prior. For the right-hand doubles amplitudes T_2, the natural low-order baseline is the canonical-orbital MP2 amplitude

$$t_{ij,\mathrm{MP2}}^{ab} = \frac{\langle ij||ab\rangle}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b}, \tag{15}$$

with $T_1^{\mathrm{MP2}} = 0$. For the left-hand amplitudes, restricted closed-shell systems with real amplitudes admit an analogous perturbative simplification:

$$\Lambda_1^{\mathrm{MP2}} = 0, \qquad \Lambda_2^{\mathrm{MP2}} = T_2^{\mathrm{MP2}*} = T_2^{\mathrm{MP2}} \tag{16}$$

For residual training, these MP2 amplitudes are computed in the canonical HF basis and then transformed with the same occupied and virtual localization matrices as the CCSD amplitudes. After this transformation, we use the same symbols for the localized-gauge tensors. This gives two natural training strategies for the Λ heads. In direct mode, we predict t_{ij}^{ab} and λ_{ab}^{ij} outright. In residual mode, we instead learn the corrections

$$\Delta t_{ij}^{ab} = t_{ij,\mathrm{CCSD}}^{ab} - t_{ij,\mathrm{MP2}}^{ab}, \tag{17}$$

$$\Delta\lambda_{ab}^{ij} = \lambda_{ab,\mathrm{CCSD}}^{ij} - \lambda_{ab,\mathrm{MP2}}^{ij}, \quad \lambda_{ab,\mathrm{MP2}}^{ij} \equiv t_{ij,\mathrm{MP2}}^{ab}. \tag{18}$$

Residualization is especially attractive in the low-data regime, because it offloads the leading-order correlation structure to perturbation theory and asks the network to learn only the physically smaller, but chemically crucial, correction.

Training objective The training objective is an amplitude reconstruction loss implemented as a per-molecule sum of elementwise squared amplitude errors, averaged over the batch,

$$\mathcal{J}_{\mathrm{amp}} = \frac{1}{B}\sum_{b=1}^{B} \sum_{\substack{X\in \\ \{T_1,T_2,\Lambda_1,\Lambda_2\}}} w_X \sum_{n=1}^{N_X^{(b)}} \left(\hat{X}_{b,n} - X_{b,n}^{\mathrm{ref}}\right)^2, \tag{19}$$

where B is the batch size and $N_X^{(b)}$ denotes the number of supervised elements in tensor X for molecule b. When residual targets are used, the same loss is applied to the residual tensors instead of the raw amplitudes. The loss weights w_X can be tuned for optimal downstream performance, and we set all active amplitude loss weights to one for simplicity.

An important feature of this training objective is that we do not supervise individual downstream properties with sepa-

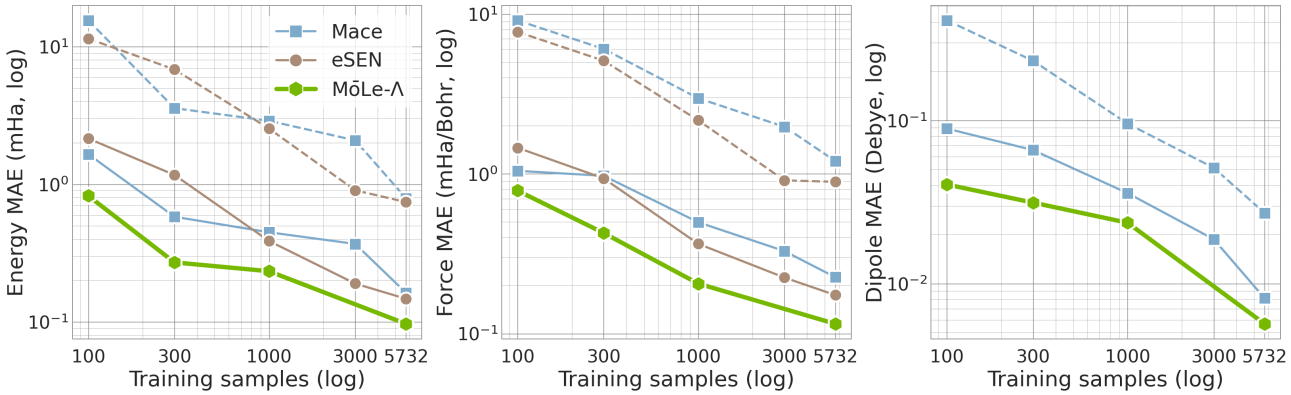

Figure 2. Scaling of the energy, force, and dipole MAE of MōLe-Λ and MLIPs with increasing data. Dashed lines predict CCSD labels, solid lines predict the delta between CCSD and MP2.

rate prediction heads. All reported energies, forces, multipoles, densities, and two-electron observables are obtained from the predicted amplitudes. Instead of learning each property independently, the model learns the underlying coupled-cluster representation.

Coupled-cluster property reconstruction The CCSD correlation energy depends only on the right-hand amplitudes and is obtained from the predicted T_1 and T_2 amplitude tensors (Bartlett & Musiał, 2007)

$$E_{\text{corr}} = \sum_{ijab} \left(\frac{1}{4} t_{ij}^{ab} + \frac{1}{2} t_i^a t_j^b \right) \langle ij||ab\rangle, \tag{20}$$

$$E_{CC} = E_{HF} + E_{corr} \tag{21}$$

To access relaxed properties we require the coupled-cluster Lagrangian in eq. 6, which allows us to compute analytical derivatives with respect to an external perturbation ξ via 9. Forces are computed from the derivative with respect to nuclear coordinates

$$\mathbf{F}_A = -\frac{\partial \mathcal{L}(T, \Lambda)}{\partial \mathbf{R}_A} \tag{22}$$

To account for the response of the molecular orbitals themselves under nuclear perturbations, one constructs CCSD density intermediates from (T, Λ) and augments them with orbital-response contributions obtained by solving a coupled perturbed Hartree–Fock (CPHF) equation (Adamowicz et al., 1984; Handy & Schaefer III, 1984).

A large class of fixed-orbital observables can be computed directly as expectation values of the Λ-state one- and two-particle reduced density matrices (1-/2-RDMs) (Stanton & Bartlett, 1993; Bartlett & Musiał, 2007). From the predicted amplitudes, we reconstruct the fixed-orbital CCSD 1-RDM and 2-RDM via

$$\gamma_{pq} = \langle \Phi_0 | (1 + \hat{\Lambda}) e^{-\hat{T}} a_p^\dagger a_q e^{\hat{T}} | \Phi_0 \rangle , \tag{23}$$

$$\Gamma_{pq,rs} = \langle \Phi_0 | (1 + \hat{\Lambda}) e^{-\hat{T}} a_p^\dagger a_q^\dagger a_s a_r e^{\hat{T}} | \Phi_0 \rangle \tag{24}$$

This convention defines γ_{pq} and $\Gamma_{pq,rs}$ with the same operator ordering that appears in the contractions below. For any one-body operator $\hat{O}^{(1)}$, its expectation value can be obtained from the 1-RDM

$$\hat{O}^{(1)} = \sum_{pq} o_{pq} a_p^\dagger a_q, \quad \langle \hat{O}^{(1)} \rangle = \sum_{pq} o_{pq} \gamma_{pq} \tag{25}$$

This includes the electron density, dipole moment, quadrupole moment, and, in principle, higher multipoles. For any two-body operator $\hat{O}^{(2)}$ the expectation value uses the 2-RDM

$$\hat{O}^{(2)} = \frac{1}{4} \sum_{pqrs} v_{pq,rs} a_p^\dagger a_q^\dagger a_s a_r, \tag{26}$$

$$\langle \hat{O}^{(2)} \rangle = \frac{1}{4} \sum_{pqrs} v_{pq,rs} \Gamma_{pq,rs} \tag{27}$$

In real space, the 2-RDM determines quantities such as the pair density. Field-response quantities such as polarizabilities are obtained by evaluating the reconstructed reduced density matrices under weak external perturbations and forming the corresponding derivatives (see Sec. A.3).

4. Experiments

We evaluate whether a learned coupled-cluster response state can replace the explicit solution of the CCSD and Λ equations for downstream property reconstruction. The central object is the set of four amplitude tensors $(T_1, T_2, \Lambda_1, \Lambda_2)$. Once these amplitudes are predicted, all reported energies, gradients, multipoles, density matrices, densities, and pair-density observables are obtained by standard coupled-cluster post-processing rather than training property-specific prediction heads.

To train MōLe-Λ we recalculate the QM7 dataset using coupled-cluster. QM7 (Rupp et al., 2012; Blum & Reymond, 2009) contains 7165 small organic molecules composed of C, N, O, S, and H, which we divide into a 80/20

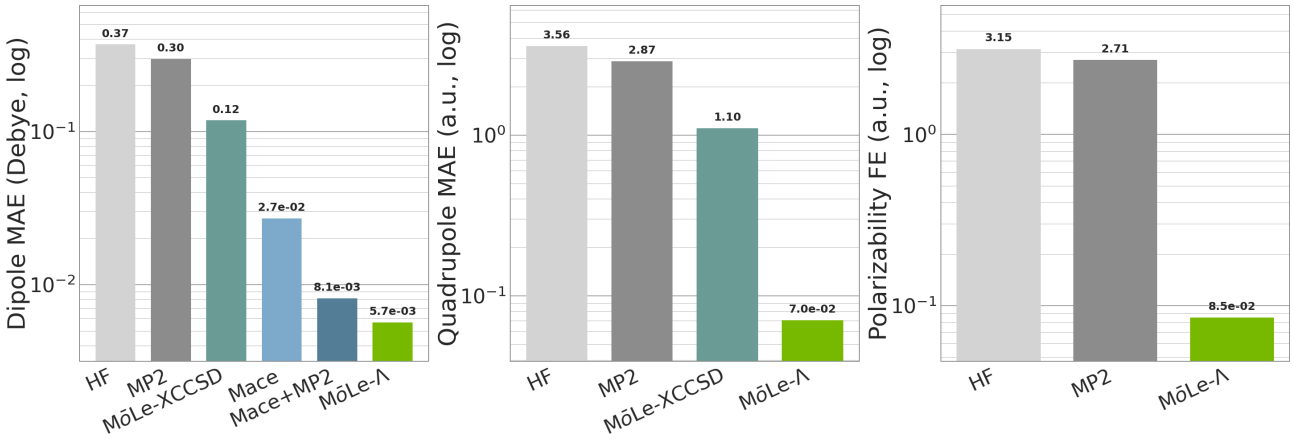

Figure 3. Mean absolute error for dipoles, quadrupoles, and polarizability on the QM7 test set. Plotted on a log scale.

train–test split. For each geometry in the dataset, we compute restricted Hartree–Fock orbitals and CCSD response amplitudes $(T_1, T_2, \Lambda_1, \Lambda_2)$ at the CCSD/def2-SVP level of theory (Weigend & Ahlrichs, 2005). All amplitudes are transformed to the separately localized occupied and virtual orbital gauge used by the model. Training details and hyperparameters are given in Appendix B.1.

Baselines We compare against Hartree–Fock, MP2, and, where appropriate, MōLe with XCCSD-style reconstructions from predicted T amplitudes only. HF provides the uncorrelated mean-field reference. MP2 provides a perturbative correlated baseline and, in residual mode, the low-order doubles amplitudes used by MōLe-Λ. However, MP2 does not provide the full relaxed CCSD response state. Thus, improvements over MP2 for Λ-dependent observables measure whether the model has learned corrections beyond leading-order perturbation theory. For energies and forces, we additionally compare against machine-learned interatomic potentials trained on the same data, as described in Appendix B.2.

Coupled-cluster energies and gradients We first assess whether the predicted amplitudes reproduce the CCSD right and left states. On the QM7 test set, MōLe-Λ achieves elementwise mean absolute errors of 2.6–2.7 $\times 10^{-5}$ for T_1 and Λ_1, and 5.3×10^{-7} for both T_2 and Λ_2. These amplitude errors translate into small downstream errors in the CCSD correlation energy and analytic forces, with MAEs of 0.10 mHa and 0.12 mHa/Bohr, respectively. As shown in Table 1, MōLe-Λ improves over both direct CCSD MLIP baselines and Δ-MP2 MLIP baselines.

We next test transfer to larger molecules. The first size-extrapolation set contains 18 structurally diverse amino acids with up to 15 heavy atoms, roughly twice the size of the largest QM7 training molecules. The second contains 100 PubChem molecules with 14 heavy atoms, spanning diverse scaffolds, functional groups, ring systems, and heterocycles. On both sets, MōLe-Λ retains low CCSD energy and force errors, whereas the geometry-only MLIP baselines degrade more strongly. This supports the hypothesis that localized orbital amplitudes provide a size-transferable representation of correlation.

Finally, we probe out-of-equilibrium generalization on three chemically distinct scans: the Diels–Alder reaction between ethylene and 1,3-butadiene, a butane central-bond dihedral scan, and the chair-to-boat conformational change of cyclohexane. Across all three scans, MōLe-Λ gives the lowest or tied-lowest energy errors and the lowest force errors in Table 1, indicating that the orbital response-state surrogate extrapolates more robustly than property-specific MLIPs.

Data efficiency and scaling Because CCSD labels are expensive, the surrogate must be accurate in the low-data regime. We therefore train MōLe-Λ and the MLIP baselines on increasingly large subsets of QM7 and evaluate the resulting energy, force, and dipole errors. Fig. 2 shows that MōLe-Λ consistently improves over MLIPs at all tested data regimes, suggesting that supervising the structured amplitude tensors $(T_1, T_2, \Lambda_1, \Lambda_2)$ is more data-efficient than directly fitting molecular properties.

Multipoles and static response We next evaluate observables that require the coupled-cluster response state, namely dipoles, quadrupoles, and polarizability. As shown in Fig. 3, MōLe-Λ substantially improves over HF, MP2, and right-state-only MōLe-XCCSD reconstructions. This indicates that the predicted Λ amplitudes provide the adjoint information needed for accurate one-particle and response observables.

Figure 4 reports the force MAE along the central C–C dihedral scan of butane (top) and the dipole MAE of methanol stretched along the C–O bond (bottom). The Mace model trained directly on CCSD data gives wildly inconsistent

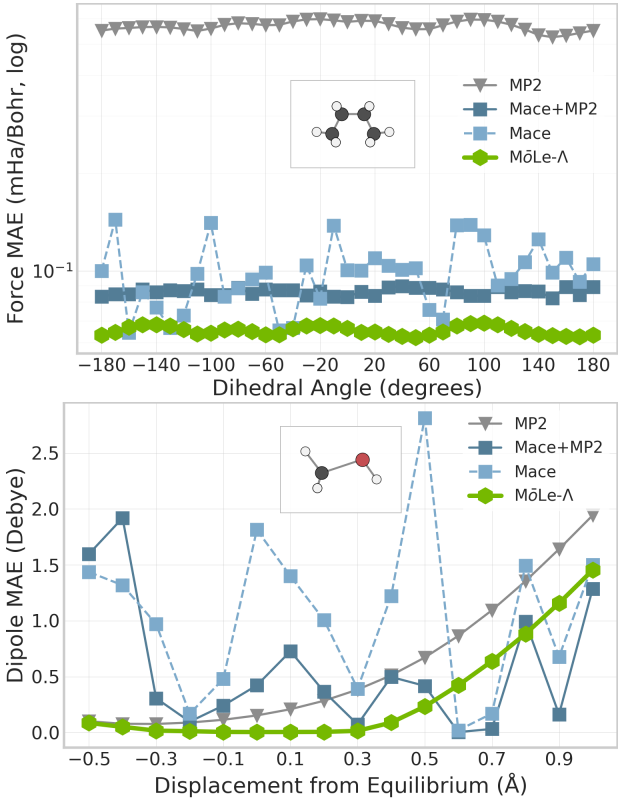

Figure 4. Out-of-equilibrium tests of the force MAE along a dihedral scan of the centre bond in a butane molecule (top) and the dipole MAE of methanol stretched along the C–O bond (bottom).

predictions. The Δ-MP2 Mace performs better on the dihedral scan, but still worse than our method, and shows erratic dipole errors on the bond stretch. MōLe-Λ consistently improves over MP2 and both baselines. This suggests that learning the response state in an orbital representation provides a more stable extrapolation mechanism than fitting properties alone.

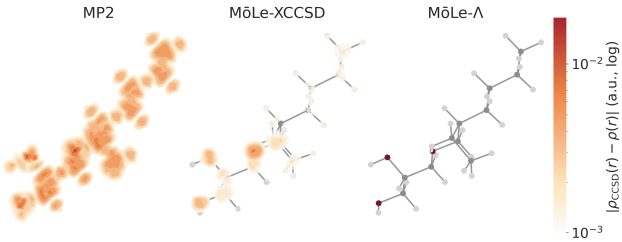

Figure 5. Difference in the electron density $\rho(\mathbf{r})$ of MP2, MōLe-XCCSD, and MōLe-Λ compared to CCSD on ethylhexylglycerin, PubChem CID 9859093, $C_{11}H_{24}O_3$. Oxygen is coloured red, nitrogen blue, carbon dark grey, and hydrogen light grey.

Electron density via the 1-RDM We then evaluate the real-space electron density reconstructed from the Λ-state 1-RDM. Figure 5 compares the density error for MP2, right-state-only MōLe with XCCSD-style reconstruction, and

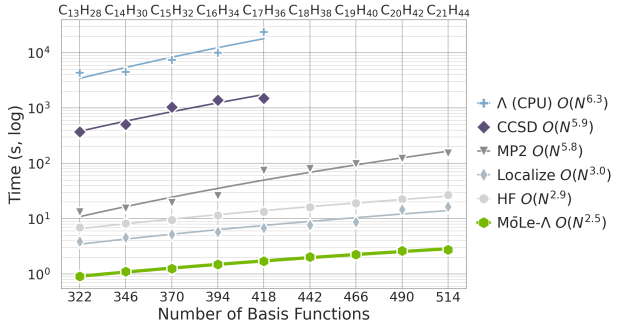

Figure 6. Empirical scaling timed on a H100 for alkane chains of increasing size. CCSD, MP2, and HF use GPU4PySCF (Li et al., 2025a). The Λ-equations and localization, are timed on CPU.

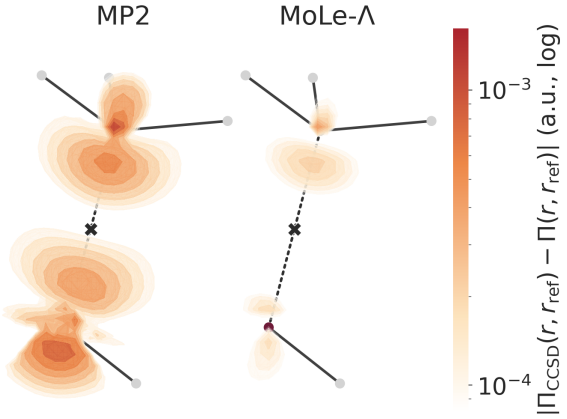

Figure 7. Difference in the pair-density $\Pi(\mathbf{r}, \mathbf{r}_{\text{ref}})$, obtained from the 2-RDM, of MP2 and MōLe-Λ relative to CCSD on a methanol geometry stretched along the C–O bond.

MōLe-Λ. MP2 produces a broad and diffuse error over the molecular volume, indicating an extended deviation from the CCSD density. MōLe significantly reduces both the magnitude and spatial extent of the residual relative to CCSD. MōLe-Λ further suppresses off-atom and bond-adjacent density errors. This improvement demonstrates the practical value of learning the adjoint amplitudes rather than the right state alone.

Pair density via the 2-RDM Unlike energies, forces, or one-body properties, the pair density $\Pi(\mathbf{r}, \mathbf{r}_{\text{ref}})$ depends directly on the Λ-state 2-RDM and therefore probes whether MōLe-Λ captures correlated electron-pair structure. Figure 7 compares the pair-density error for stretched methanol along the C–O bond, with the reference point $\mathbf{r}_{\text{ref}}$ placed in the internuclear region. Relative to MP2, MōLe-Λ yields a smaller and more localized error. The broad MP2 error lobes near the bond are strongly suppressed, indicating that the learned response state better captures the redistribution of electron-pair correlation in the stretched polar bond.

Computational scaling Finally, we compare wall-clock scaling against conventional CCSD+Λ calculations. Figure 6 reports timings for increasing alkane chains using GPU4PySCF (Li et al., 2025a); orbital localization and the Λ equations are only available on CPU. Conventional CCSD calculations include HF, localization, MP2, CCSD, and the Λ solve, whereas our method replaces the CCSD+Λ steps by optional MP2 calculation and a MōLe-Λ forward pass. For MōLe-Λ, the dominant outputs are the dense doubles tensors T_2 and Λ_2, each of size $\mathcal{O}(N_{\text{occ}}^2 N_{\text{virt}}^2)$. The empirical scaling of MōLe-Λ is lower than $\mathcal{O}(N^3)$ in the tested regime, likely because the dominant tensor operations are efficiently parallelized on GPU. At $C_{17}H_{36}$, CCSD exhausts H100 memory, while MōLe-Λ scales far beyond C_{21}. At the largest common system size, MōLe-Λ obtains both left- and right-hand amplitudes more than two orders of magnitude faster than solving the CCSD and Λ equations. The dense T_2 and Λ_2 outputs are not yet the practical bottleneck in the present experiments, but sparse, local, or compressed doubles representations provide a natural path to larger systems. Further gains should follow from GPU-native localization and MP2.

Limitations and future work The present study is restricted to the def2-SVP basis and to the elements represented in QM7. Larger basis sets, heavier elements, open-shell systems, and more diverse chemical environments are natural next steps. The current implementation also treats HF, localization, and optional MP2 residual construction as external preprocessing steps; these are not yet optimized to the same degree as the neural network evaluation. Finally, future work should explore sparse representations for T_2 and Λ_2 tensors to scale to substantially larger systems.

5. Conclusions

MōLe-Λ extends orbital-based amplitude learning from the CCSD right state to the full coupled-cluster ground-state response state. The core claim is that if a model predicts $(T_1, T_2, \Lambda_1, \Lambda_2)$ from localized Hartree–Fock molecular orbitals, then one learned object can support energies, forces, response properties, and reduced-density-matrix observables within a common framework. This reframes machine learning in correlated chemistry from approximating isolated observables to approximating the underlying electronic-structure state.

Impact Statement

This paper presents work whose goal is to advance machine learning methods for correlated quantum chemistry. By making coupled-cluster-level response properties more accessible, MōLe-Λ could reduce the computational cost of molecular simulation and support downstream research in materials, catalysis, energy, and molecular design. As with other learned surrogates for scientific computing, possible negative impacts arise primarily from misuse or overinterpretation outside the training domain, where inaccurate energies, forces, or response properties could lead to misleading scientific conclusions. The broader ethical and societal implications are otherwise those associated with advancing machine learning and computational chemistry; we do not identify additional impacts requiring specific emphasis.

References

Adamowicz, L., Laidig, W., and Bartlett, R. Analytical gradients for the coupled-cluster method. *International Journal of Quantum Chemistry*, 26(S18):245–254, 1984. 3

Banchode, R., Das, S., Raghunathan, S., and Ramakrishnan, R. Machine-Learned Potentials for Solvation Modeling, October 2025. URL http://arxiv.org/abs/2505.22402. arXiv:2505.22402 [physics]. 1

Bartlett, R. J. and Musiał, M. Coupled-cluster theory in quantum chemistry. *Reviews of Modern Physics*, 79(1): 291–352, 2007. 1, 2, 3, 3

Batatia, I., Kovacs, D. P., Simm, G., Ortner, C., and Csanyi, G. MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields. *Advances in Neural Information Processing Systems*, 35: 11423–11436, December 2022. 3

Blum, L. C. and Reymond, J.-L. 970 million druglike small molecules for virtual screening in the chemical universe database GDB-13. *J. Am. Chem. Soc.*, 131:8732, 2009. 4

Brockherde, F., Vogt, L., Li, L., Tuckerman, M. E., Burke, K., and Müller, K.-R. Bypassing the kohn-sham equations with machine learning. *Nature communications*, 8(1):872, 2017. 1

Burger, A., Thiede, L., Rønne, N., Bernales, V., Vijaykumar, N., Vegge, T., Bhowmik, A., and Aspuru-Guzik, A. Shoot from the hip: Hessian interatomic potentials without derivatives. *arXiv preprint arXiv:2509.21624*, 2025. 1

Chen, Y. and Dral, P. O. One to Rule Them All: A Universal Interatomic Potential Learning across Quantum Chemical Levels. *Journal of Chemical Theory and Computation*, 21(18):8762–8772, September 2025. ISSN 1549-9618, 1549-9626. doi: 10.1021/acs.jctc.5c00858. URL https://pubs.acs.org/doi/10.1021/acs.jctc.5c00858. 1

Elsborg, J., Thiede, L., Aspuru-Guzik, A., Vegge, T., and Bhowmik, A. Electra: A cartesian network for 3d charge

density prediction with floating orbitals. *arXiv preprint arXiv:2503.08305*, 2025. 1

Eriksen, J. J., Kristensen, K., Kjærgaard, T., Jørgensen, P., and Gauss, J. A Lagrangian framework for deriving triples and quadruples corrections to the CCSD energy. *The Journal of Chemical Physics*, 140(6):064108, February 2014. ISSN 0021-9606. doi: 10.1063/1.4862501. URL https://doi.org/10.1063/1.4862501. 1

Febrer, P., Jørgensen, P. B., Pruneda, M., García, A., Ordejón, P., and Bhowmik, A. Graph2mat: universal graph to matrix conversion for electron density prediction. *Machine Learning: Science and Technology*, 6(2):025013, 2025. 1

Gao, N. and Günnemann, S. Generalizing Neural Wave Functions. In *Proceedings of the 40th International Conference on Machine Learning*, pp. 10708–10726. PMLR, July 2023. URL https://proceedings.mlr.press/v202/gao23c.html. ISSN: 2640-3498. 1

Handy, N. C. and Schaefer III, H. F. On the evaluation of analytic energy derivatives for correlated wave functions. *The Journal of chemical physics*, 81(11):5031–5033, 1984. 3

Hazra, S., Patil, U., and Sanvito, S. Predicting the one-particle density matrix with machine learning. *Journal of Chemical Theory and Computation*, 20(11):4569–4578, 2024. 1

Helgaker, T., Jorgensen, P., and Olsen, J. *Molecular electronic-structure theory*. John Wiley & Sons, 2013. 1

Jørgensen, P. B. and Bhowmik, A. Equivariant graph neural networks for fast electron density estimation of molecules, liquids, and solids. *npj Computational Materials*, 8(1): 183, 2022. 1

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B. A., Thiessen, P. A., Yu, B., et al. Pubchem 2025 update. *Nucleic acids research*, 53(D1): D1516–D1525, 2025a. B.3

Kim, S., Kim, N., Kim, D., and Ahn, S. High-order equivariant flow matching for density functional theory hamiltonian prediction. *arXiv preprint arXiv:2505.18817*, 2025b. 1

Koch, H. and Jørgensen, P. Coupled cluster response functions. *Journal of Chemical Physics*, 93(5):3333–3344, 1990. 2

Kohn, W. and Sham, L. J. Self-consistent equations including exchange and correlation effects. *Physical review*, 140(4A):A1133, 1965. 1

Korona, T. and Jeziorski, B. One-electron properties and electrostatic interaction energies from the expectation value expression and wave function of singles and doubles coupled cluster theory. *The Journal of Chemical Physics*, 125(18):184109, November 2006. ISSN 0021-9606. doi: 10.1063/1.2364489. URL https://doi.org/10.1063/1.2364489. A.2

Li, R., Sun, Q., Zhang, X., and Chan, G. K.-L. Introducing gpu acceleration into the python-based simulations of chemistry framework. *The Journal of Physical Chemistry A*, 129(5):1459–1468, 2025a. 6, 4

Li, Y., Xia, Z., Huang, L., Wei, X., Yang, H., Harshe, S., Wang, Z., Liu, C., Zhang, J., Shao, B., et al. Enhancing the scalability and applicability of kohn-sham hamiltonians for molecular systems. *arXiv preprint arXiv:2502.19227*, 2025b. 1

Luo, E., Wei, X., Huang, L., Li, Y., Yang, H., Xia, Z., Wang, Z., Liu, C., Shao, B., and Zhang, J. Efficient and scalable density functional theory hamiltonian prediction through adaptive sparsity. *arXiv preprint arXiv:2502.01171*, 2025. 1

Mardirossian, N. and Head-Gordon, M. Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals. *Molecular physics*, 115(19):2315–2372, 2017. 1

Martyka, M., Tong, X.-Y., Jankowska, J., and Dral, P. O. OMNI-P2x: A Universal Neural Network Potential for Excited-State Simulations, October 2025. URL https://chemrxiv.org/engage/chemrxiv/article-details/68fdbb0daec32c65683cdfb3. 1

Mehdizadeh, A. and Schindler, P. Surface Stability Modeling with Universal Machine Learning Interatomic Potentials: A Comprehensive Cleavage Energy Benchmarking Study. *AI for Science*, 1(2):025002, December 2025. ISSN 3050-287X. doi: 10.1088/3050-287X/ae1408. URL http://arxiv.org/abs/2508.21663. arXiv:2508.21663 [cond-mat]. 1

Purvis III, G. D. and Bartlett, R. J. A full coupled-cluster singles and doubles model: The inclusion of disconnected triples. *The Journal of chemical physics*, 76(4):1910–1918, 1982. 1, 2

Qin, Z., Han, X., Sun, W., Li, D., Kong, L., Barnes, N., and Zhong, Y. The Devil in Linear Transformer. *arXiv.org*, October 2022. URL https://arxiv.org/abs/2210.10340v1. 3

Raghavachari, K., Trucks, G. W., Pople, J. A., and Head-Gordon, M. A fifth-order perturbation comparison of electron correlation theories. *Chemical Physics Letters*, 157(6):479–483, 1989. 1

Rupp, M., Tkatchenko, A., Müller, K.-R., and von Lilienfeld, O. A. Fast and accurate modeling of molecular atomization energies with machine learning. *Physical Review Letters*, 108:058301, 2012. 4

Scherbela, M., Gerard, L., and Grohs, P. Towards a transferable fermionic neural wavefunction for molecules. *Nature Communications*, 15(1):120, January 2024. ISSN 2041-1723. doi: 10.1038/s41467-023-44216-9. URL https://www.nature.com/articles/s41467-023-44216-9. Publisher: Nature Publishing Group. 1

Shang, H., Guo, C., Wu, Y., Li, Z., and Yang, J. Solving the many-electron Schrödinger equation with a transformer-based framework. *Nature Communications*, 16(1):8464, September 2025. ISSN 2041-1723. doi: 10.1038/s41467-025-63219-2. URL https://www.nature.com/articles/s41467-025-63219-2. Publisher: Nature Publishing Group. 1

Shao, X., Paetow, L., Tuckerman, M. E., and Pavanello, M. Machine learning electronic structure methods based on the one-electron reduced density matrix. *Nature communications*, 14(1):6281, 2023. 1

Smith, D. G., Burns, L. A., Simmonett, A. C., Parrish, R. M., Schieber, M. C., Galvelis, R., Kraus, P., Kruse, H., Di Remigio, R., Alenaizan, A., et al. Psi4 1.4: Open-source software for high-throughput quantum chemistry. *The Journal of chemical physics*, 152(18), 2020a. A.3

Smith, J. S., Zubatyuk, R., Nebgen, B., Lubbers, N., Barros, K., Roitberg, A. E., Isayev, O., and Tretiak, S. The ani-1ccx and ani-1x data sets, coupled-cluster and density functional theory properties for molecules. *Scientific data*, 7(1):134, 2020b. 1

Stanton, J. F. and Bartlett, R. J. The equation of motion coupled-cluster method. a systematic biorthogonal approach to molecular excitation energies, transition probabilities, and excited state properties. *The Journal of chemical physics*, 98(9):7029–7039, 1993. 3

Szabo, A. and Ostlund, N. S. *Modern quantum chemistry: introduction to advanced electronic structure theory*. Courier Corporation, 2012. 1, 2

Thiede, L., Aldossary, A., Burger, A., Campos-Gonzalez-Angulo, J. A., Wang, N., Zook, A., Alkan, M., Nakaji, K., Patti, T. L., Gonthier, J. F., Vakili, M. G., and Aspuru-Guzik, A. Coupled cluster con mole: Molecular orbital learning for neural wavefunctions. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=i5QhkUCzJh. 1, 3, 3, B.1, B.1

Townsend, J. and Vogiatzis, K. D. Data-driven acceleration of the coupled-cluster singles and doubles iterative solver. *The journal of physical chemistry letters*, 10(14):4129–4135, 2019. 1

Townsend, J. and Vogiatzis, K. D. Transferable mp2-based machine learning for accurate coupled-cluster energies. *Journal of Chemical Theory and Computation*, 16(12):7453–7461, 2020. 1

Venturella, C., Hillenbrand, C., Li, J., and Zhu, T. Machine Learning Many-Body Green's Functions for Molecular Excitation Spectra. *Journal of Chemical Theory and Computation*, 20(1):143–154, January 2024. ISSN 1549-9618. doi: 10.1021/acs.jctc.3c01146. URL https://doi.org/10.1021/acs.jctc.3c01146. Publisher: American Chemical Society. 1

Venturella, C., Li, J., Hillenbrand, C., Leyva Peralta, X., Liu, J., and Zhu, T. Unified deep learning framework for many-body quantum chemistry via Green's functions. *Nature Computational Science*, 5(6):502–513, June 2025. ISSN 2662-8457. doi: 10.1038/s43588-025-00810-z. URL https://www.nature.com/articles/s43588-025-00810-z. Publisher: Nature Publishing Group. 1

Weigend, F. and Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for h to rn: Design and assessment of accuracy. *Physical Chemistry Chemical Physics*, 7(18):3297–3305, 2005. 4

Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., et al. Drugbank 5.0: a major update to the drugbank database for 2018. *Nucleic acids research*, 46(D1):D1074–D1082, 2018. B.3

Yu, H., Xu, Z., Qian, X., Qian, X., and Ji, S. Efficient and equivariant graph networks for predicting quantum hamiltonian. In *International Conference on Machine Learning*, pp. 40412–40424. PMLR, 2023. 1

A. Additional Background

A.1. Λ amplitudes

In conventional CCSD, the right amplitudes are obtained first from the projected residual equations $R_\mu(T) = 0$, where μ indexes single and double excitations. For fixed converged T, the Λ amplitudes are then obtained from the linear adjoint equations

$$\frac{\partial \mathcal{L}}{\partial t_\mu} = \frac{\partial E_{\mathrm{CC}}}{\partial t_\mu} + \sum_\nu \lambda_\nu \frac{\partial R_\nu}{\partial t_\mu} = 0. \tag{28}$$

Λ acts as the Lagrange multiplier enforcing stationarity of the CC residuals, or equivalently the left eigenvector-like response state associated with the CC Jacobian.

The reason Λ is useful can be seen from differentiating the CC energy with respect to an external perturbation ξ. Since the converged amplitudes depend on the perturbation, $T = T(\xi)$, the total derivative contains an explicit term and an implicit amplitude-response term,

$$\frac{dE_{\mathrm{CC}}}{d\xi} = \frac{\partial E_{\mathrm{CC}}}{\partial \xi} + \sum_\mu \frac{\partial E_{\mathrm{CC}}}{\partial t_\mu} \frac{\partial t_\mu}{\partial \xi}, \tag{29}$$

where μ indexes the single and double excitation amplitudes. Unlike variational methods, CC theory is not stationary with respect to the right amplitudes alone, so the second term does not vanish in general. Computing it directly would require solving for the response amplitudes $\partial t_\mu / \partial \xi$ for each perturbation.

The Lagrangian avoids this by introducing Λ as an adjoint variable. Writing the projected CC residual equations as $R_\mu(T; \xi) = 0$, the Lagrangian can be written schematically as

$$\mathcal{L}(T, \Lambda; \xi) = E_{\mathrm{CC}}(T; \xi) + \sum_\mu \lambda_\mu R_\mu(T; \xi). \tag{30}$$

The Λ amplitudes are chosen such that

$$\frac{\partial \mathcal{L}}{\partial t_\mu} = 0 \quad \text{for all } \mu. \tag{31}$$

Therefore, when the total derivative is taken, the terms proportional to $\partial t_\mu / \partial \xi$ cancel,

$$\frac{dE}{d\xi} = \frac{d\mathcal{L}}{d\xi} = \frac{\partial \mathcal{L}}{\partial \xi} + \sum_\mu \frac{\partial \mathcal{L}}{\partial t_\mu} \frac{\partial t_\mu}{\partial \xi} + \sum_\mu \frac{\partial \mathcal{L}}{\partial \lambda_\mu} \frac{\partial \lambda_\mu}{\partial \xi} = \frac{\partial \mathcal{L}}{\partial \xi}. \tag{32}$$

In this sense, Λ absorbs the implicit response of the right amplitudes, and converts the derivative of a non-variational CC energy into the partial derivative of a stationary Lagrangian.

A.2. Expectation-value CCSD and one-particle properties

While the CCSD energy depends only on the excitation amplitudes t_i^a and t_{ij}^{ab}, general molecular observables require expectation values of operators. In particular, one-particle properties such as the dipole moment and the electron density are determined by the one-particle reduced density matrix (1-RDM). In this work, we use expectation-value coupled cluster with singles and doubles (XCCSD) (Korona & Jeziorski, 2006) evaluated from the predicted T amplitudes as a natural baseline for such properties. This provides a simple route from T_1 and T_2 to approximate correlated observables, against which we compare the more accurate properties obtained when both T- and Λ-amplitudes are available.

In XCC, the normalized expectation value of a one-body operator $\hat{X}$ is written as

$$\langle \hat{X} \rangle = \frac{\langle \Phi_{\mathrm{HF}} | \, e^{\hat{T}^\dagger} \hat{X} e^{\hat{T}} \, | \Phi_{\mathrm{HF}} \rangle}{\langle \Phi_{\mathrm{HF}} | \, e^{\hat{T}^\dagger} e^{\hat{T}} \, | \Phi_{\mathrm{HF}} \rangle}, \tag{33}$$

with $\hat{T} = \hat{T}_1 + \hat{T}_2$ in CCSD. This expression can be recast into a connected form through an auxiliary operator $\hat{S}$ constructed from $\hat{T}$ by a commutator expansion. In practice, this expansion is truncated, yielding an approximate but size-extensive expectation-value formalism. Here, we use this low-order XCCSD treatment as a baseline method for extracting one-particle properties from predicted T amplitudes alone.

The most general one-particle operator has the second-quantized form

$$\hat{X} = \sum_{pq} X_{pq}\, \hat{a}_p^\dagger \hat{a}_q, \tag{34}$$

where X_{pq} are matrix elements in the molecular-orbital basis. Its expectation value is fully determined by the 1-RDM,

$$\gamma_{pq} = \langle \hat{a}_p^\dagger \hat{a}_q \rangle, \tag{35}$$

since

$$\langle \hat{X} \rangle = \sum_{pq} X_{pq}\, \gamma_{pq}. \tag{36}$$

Thus, once an approximate XCCSD 1-RDM has been constructed from the amplitudes, any one-particle observable follows by contraction with the corresponding operator matrix.

Two important examples are the dipole moment and the real-space electron density. The electronic contribution to the dipole moment is obtained from the dipole-integral matrix $\mu_{pq}^{(\alpha)}$ for Cartesian component $\alpha \in \{x, y, z\}$ as

$$\mu_\alpha^{\mathrm{el}} = -\sum_{pq} \mu_{pq}^{(\alpha)} \gamma_{pq}, \tag{37}$$

to which the nuclear contribution is added in the usual way,

$$\mu_\alpha = \sum_A Z_A R_{A\alpha} + \mu_\alpha^{\mathrm{el}}. \tag{38}$$

Likewise, the electron density is obtained from the diagonal of the real-space 1-RDM,

$$\rho(\mathbf{r}) = \sum_{pq} \gamma_{pq}\, \psi_p(\mathbf{r}) \psi_q(\mathbf{r}). \tag{39}$$

Therefore, predicted T amplitudes alone already provide access to approximate dipoles, densities, and other one-particle observables through XCCSD.

Because XCCSD relies only on the right-hand amplitudes and a truncated expectation-value construction, its accuracy is limited relative to formulations that also include the left-hand coupled-cluster information. This motivates our extension to models that predict the Λ amplitudes in addition to T, enabling substantially more accurate one-particle properties.

A.3. Polarizability

We compute the static frozen-orbital polarizability analytically from the coupled-cluster response state. The implementation follows Psi4 (Smith et al., 2020a), but is implemented in our PySCF-compatible GPU codebase. In this setting, the Hartree–Fock orbitals are held fixed under the external electric field, and the polarizability measures the correlated response of the CCSD amplitudes and Λ-state density matrices in that fixed molecular-orbital basis. Following the Psi4 documentation recommendation, we use this frozen-orbital convention for the coupled-cluster response comparison reported here.

Let $\mathbf{F} = (F_x, F_y, F_z)$ denote a homogeneous electric field. The field-dependent Hamiltonian is

$$\hat{H}(\mathbf{F}) = \hat{H}(0) - \sum_{\alpha \in \{x,y,z\}} F_\alpha \hat{\mu}_\alpha, \tag{40}$$

where $\hat{\mu}_\alpha$ is the α component of the total dipole operator. The frozen-orbital polarizability tensor is

$$\alpha_{\alpha\beta}^{\mathrm{FO}} = -\left.\frac{\partial^2 E(\mathbf{F})}{\partial F_\alpha \partial F_\beta}\right|_{\mathbf{F}=0} = \left.\frac{\partial \mu_\alpha}{\partial F_\beta}\right|_{\mathbf{F}=0}. \tag{41}$$

Because the orbitals are fixed, no coupled-perturbed Hartree–Fock orbital-response terms are included.

For CCSD, the derivative is evaluated through the coupled-cluster Lagrangian,

$$\mathcal{L}(T, \Lambda; \mathbf{F}) = \langle \Phi_0 | \, (1 + \hat{\Lambda}) e^{-\hat{T}} \hat{H}(\mathbf{F}) e^{\hat{T}} \, | \Phi_0 \rangle \, . \tag{42}$$

At stationarity with respect to T and Λ, field derivatives of the energy can be written as partial derivatives of the Lagrangian. The first derivative gives the relaxed dipole moment,

$$\mu_\alpha = -\frac{\partial \mathcal{L}}{\partial F_\alpha} = \sum_A Z_A R_{A\alpha} - \sum_{pq} \mu_{pq}^{(\alpha)} \gamma_{pq}, \tag{43}$$

where γ is the fixed-orbital CCSD one-particle density matrix. The polarizability is then obtained from the analytic derivative of this density with respect to the perturbing field,

$$\alpha_{\alpha\beta}^{\mathrm{FO}} = -\sum_{pq} \mu_{pq}^{(\alpha)} \frac{\partial \gamma_{pq}}{\partial F_\beta}. \tag{44}$$

The required density response is computed by solving the fixed-orbital CCSD response equations for each Cartesian perturbation direction β and contracting the resulting response intermediates with the dipole-integral matrices. For the HF and MP2 baselines, we follow Psi4 and compute polarizabilities by finite differences rather than by an analytic response implementation. In the main text we report the Frobenius error

$$|\alpha_{\text{pred}} - \alpha_{\text{CCSD}}|_F = \sqrt{\sum_{i,j} (\alpha_{ij}^{\text{pred}} - \alpha_{ij}^{\text{CCSD}})^2} \tag{45}$$

B. Experimental details

B.1. Hyperparameters for joint T- and Λ-amplitude training

We summarize the hyperparameters for MōLe-Λ in tab. 2. Compared to (Thiede et al., 2026), we increase the equivariant transformer width and readout capacity and train jointly on the Λ_1 and Λ_2 amplitude targets in addition to the T-amplitude readout configuration. We opt for a wider latent representation, using 256 channels for each of the $\ell = 0, 1, 2$ hidden irreps, a larger attention bottleneck, and higher-capacity single-to-pair readouts. The pair-to-quadruple stage is changed from an MLP readout to a dot-product readout, which directly contracts pair features over atoms and feature channels to produce scalar quadruple amplitudes. We switch from post-norm to pre-norm transformer blocks, applying equivariant normalization before the attention and message-passing updates to improve training stability. We increased the radial basis from 10 to 16 Bessel functions and the cutoff from 4.0 to 6.0, as we found both improve generalization to out-of-equilibrium geometries. We also update optimization to use Muon with mixed bfloat16 precision and cosine learning-rate decay with warmup. Our model is trained for up to a week on a single H100.

L_2 and T_2 dot-product readout Compared to MōLe (Thiede et al., 2026), the pair-to-quadruple stage is changed from an MLP readout to a simpler dot-product readout. Both variants take pair features $f_{ia}, f_{jb} \in \mathbb{R}^{B \times N_o \times N_v \times N_{\text{atom}} \times F}$ and return scalar quadruple amplitudes $p_{ijab} \in \mathbb{R}^{B \times N_o \times N_o \times N_v \times N_v}$. The dot-product readout computes this scalar directly as

$$p_{ijab} = \frac{1}{\sqrt{F}} \sum_{\alpha=1}^{N_{\text{atom}}} \sum_{c=1}^{F} f_{ia\alpha c} f_{jb\alpha c},$$

i.e. it contracts over the atom axis and the full pair-feature channel axis. In the MLP variant, the pair features are first normalized over the combined (N_{atom}, F) axis and projected into a hidden equivariant representation, e.g. $32 \times 0e + 16 \times 1e + 8 \times 2e$ with total feature dimension $32 + 16 \cdot 3 + 8 \cdot 5 = 120$. An invariant contraction is then performed blockwise over atoms and spherical components, producing one scalar per multiplicity channel, $32 + 16 + 8 = 56$ invariant features per (i, j, a, b) quadruple. These 56 features are passed through a small MLP, for example $56 \to 16 \to 1$, to produce the final scalar amplitude. Empirically we find that the dot-product readout reaches almost the same accuracy, at 30% faster training and inference.

B.2. MLIP baselines

For energy and force prediction we compare against state-of-the-art machine learning interatomic potentials (MLIPs) that are trained on the same five QM7 training-set sizes (100, 300, 1000, 3000, and full 5732) for both direct CCSD targets and residual targets $\Delta = \text{CCSD} - \text{MP2}$. For the CCSD target prediction we use fixed atomic reference energies; for the delta

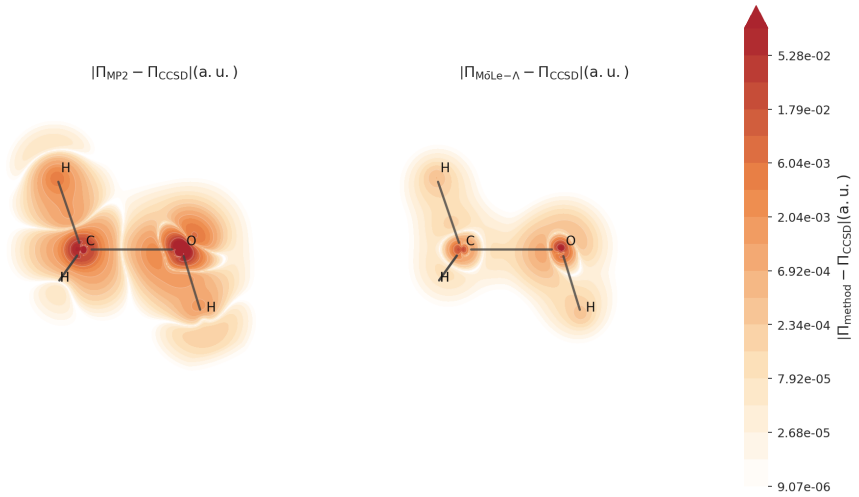

Figure 8. Error of the on-top pair-density Π, obtained from the 2-RDM, of MP2 and MōLe-Λ compared to CCSD.

target we set the atomic references to zero. All baseline runs are trained on a single NVIDIA A100 80GB GPU for 3000 epochs and follow the standard train/validation splits used throughout our QM7 experiments.

MACE: For MACE, we use three interaction blocks, hidden irreps `128x0e + 128x1o`, cutoff radius $r_{\max} = 6.0$ Å, and spherical harmonics up to $\ell = 3$. Optimization uses AdamW with batch size 32, learning rate 10^{-3}, weight decay 10^{-3}, cosine annealing to 10^{-6}, and 3000 epochs. The loss weights are $(1, 100, 1)$ for energy, forces, and dipoles. Across the 10 runs, the total training time is approximately 150 A100-hours.

eSEN: For eSEN, we use the UMA `K4L2` backbone with 128 sphere channels, $l_{\max} = 2$, $m_{\max} = 2$, cutoff radius 6 Å, and at most 30 neighbors. Training uses AdamW with learning rate 5×10^{-4}, gradient clipping of 10, zero weight decay, bf16 precision, and max-atoms batching. We supervise direct energies and forces with coefficient 10 for both terms. The ten runs account for roughly 110 A100-hours.

B.3. Evaluation set selection

PubChem. The PubChem size-extrapolation set was drawn from the PubChem Compound DrugBank (Kim et al., 2025a; Wishart et al., 2018). We downloaded the DrugBank information-source subset in JSON format with coordinate type set to 3D, then kept even-electron molecules containing only H, C, N, and O, with available conformer coordinates and exactly 14 heavy atoms. The random selection used NumPy seed 42. The final 100 CIDs are 62787, 46937113, 5289331, 604519, 5192, 2482, 11229234, 680935, 5288483, 1318, 54695768, 9859093, 7184, 6199, 26757, 445245, 122282, 25429, 736186, 4099, 62529, 161358, 6852125, 6862, 445210, 450268, 2148, 697959, 6047, 3052, 7045767, 3604, 5280980, 123831, 3845, 441350, 1616, 21120522, 12551, 3870203, 30487, 445211, 445994, 6420149, 45098889, 13294447, 25145077, 3518, 6839, 33037, 91514, 15664, 9815560, 61810, 16615, 445858, 4284, 6950274, 3102, 90274764, 5494443, 254, 7127816, 3744, 1082702, 38945, 2519, 2737071, 12192007, 198695, 5288636, 5447130, 446429, 441314, 121026, 1615, 3698, 1676, 4474777, 60748, 445825, 44585557, 16143, 104904, 5494411, 11071, 3893, 40632, 8367, 46936338, 10130337, 12097317, 10205, 17754076, 447364, 25058136, 61361, 5394, 14052, and 178052.

Amino acids. The amino-acid set contains the following 18 standard amino acids: alanine, arginine, asparagine, aspartate, glutamine, glutamate, glycine, histidine, isoleucine, leucine, lysine, phenylalanine, proline, serine, threonine, tryptophan, tyrosine, and valine.

C. Additional Results

On-top pair-density Fig. 8 shows the on-top pair-density error, $|\Pi_{\text{method}}(\mathbf{r}, \mathbf{r}) - \Pi_{\text{CCSD}}(\mathbf{r}, \mathbf{r})|$, for MP2 and MōLe-Λ in methanol. The MP2 error is broad and pronounced, with extended high-error regions around both heavy atoms and along the C–O bond, indicating a substantial misrepresentation of short-range electron correlation across the molecular framework.

By contrast, for MōLe-Λ the largest errors remain localized near the atomic centres, while the diffuse bond-centred and off-atom error visible for MP2 is strongly suppressed. This indicates that MōLe-Λ reproduces the CCSD on-top pair density much more faithfully, reducing the overall magnitude of the error and recovering its correct spatial localization.

Category	Configuration
Transformer	
Number of transformer layers	4
Hidden irreps	`256x0e + 256x1o + 256x2e`
Edge irreps	`1x0e + 1x1o + 1x2e`
Normalization	Separable equivariant layer norm
Radial / GNN block	
Maximum radius	6.0
Radial basis	Bessel
Number of Bessel functions	16
Polynomial cutoff order	5
MACE correlation order	3
Attention block	
Attention type	Sign-equivariant dot-product attention
Latent irreps	`64x0e + 64x1o + 64x2e`
Number of attention heads	8
Norm	Pre-norm with skip connections
T_1 and Λ_1 readouts	
Readout module	MLP single-to-pair readout
Readout irreps	`64x0e + 64x1o + 64x2e`
MLP hidden width	64
T_2 and Λ_2 readouts	
Single→pair module	Tensor-product pair feature module
Single→pair irreps	`64x0e + 64x0o + 64x1e + 64x1o + 64x2e + 64x2o`
Pre-tensor-product irreps	`64x0e + 64x1o + 32x2e`
Pair→quadruple module	Dot-product readout
Training	
Epochs	100
Batch size	1
Optimizer	Muon
Base learning rate	10^{-2}
Muon learning rate	10^{-2}
Weight decay	0
Loss function	summed MSE over amplitudes
Λ_1 loss factor	1.0
Λ_2 loss factor	1.0
Precision	bfloat16 mixed precision
Learning-rate schedule	
Scheduler	Cosine annealing
Minimum learning rate	10^{-6}
Warmup fraction	0.1
Warmup start factor	0.01

Table 2. Model architecture and training hyperparameters for the joint T- and Λ-amplitude model.