SOLVER-ORIENTED ELECTRON DENSITY LEARNING WITH EQUIVARIANT GAUSSIAN SHELLS

Anonymous authors
Paper under double-blind review

ABSTRACT

Learning electron density from molecular geometry supports electronic-property recovery and self-consistent field (SCF) initialization. We introduce GEM-FIELD, an equivariant Gaussian-shell representation with learnable radial parameters, nonlinear angular distributions, and exact electron-number conservation. Its Gaussian expansion provides analytical Coulomb integrals for target atomic-orbital bases, connecting predicted densities to property evaluation and SCF solvers. On QM9-PBE, GEM-FIELD achieves 0.226% dense-grid density NMAE; its compact variant achieves 0.237% while reducing evaluation time by 96.3% relative to ChargE3Net. On larger OE62 molecules, density-derived frontier-orbital errors are 86.7%–92.8% lower than task-specific predictions. On SCFBench OOD molecules with up to 60 atoms, GEM-FIELD achieves 100% convergence, reduces SCF iterations by 40.5%, and lowers end-to-end runtime by 8.98% relative to MINAO. Further evaluations demonstrate transferable initialization across target AO bases and exchange–correlation functionals.

1 INTRODUCTION

Kohn–Sham density functional theory (KS-DFT) provides accurate electronic structures but requires iterative solution of density-dependent equations (Hohenberg & Kohn, 1964; Kohn & Sham, 1965). Each self-consistent field (SCF) iteration reconstructs the Kohn–Sham matrix and solves a generalized eigenvalue problem, making calculations costly for large systems or poor initial guesses. Learning converged electronic structure from molecular geometry can amortize this cost across molecules (Schütt et al., 2019; Hazra et al., 2024). Accordingly, we learn to predict converged electron densities from molecular geometries. The learned density serves two downstream tasks: recovering approximate electronic properties through a single Kohn–Sham diagonalization and providing an initial density matrix for subsequent SCF iterations. Our contribution centers on the density representation and its interface with Gaussian-AO solvers.

Electron density is a natural intermediate because it is continuous, determines multiple observables, and can be represented independently of the target atomic-orbital (AO) basis. Grid and probe models predict densities at spatial query points, but their outputs lack an analytical structure compatible with Gaussian-type orbitals (GTOs), hindering efficient Coulomb-matrix construction in the atomic-orbital representation (Jørgensen & Bhowmik, 2022; Koker et al., 2024). AO-matrix models interface directly with these solvers, while their outputs depend on the chosen orbital basis (Hazra et al., 2024). Auxiliary-basis density models support analytical integration and reuse across target AO bases; their achievable density accuracy is governed by the expressive capacity of the finite auxiliary expansion (Grisafi et al., 2019; Liu et al., 2026). These considerations motivate a density representation with learnable radial structure and expressive angular distributions that retains an efficient Gaussian integral interface.

Three requirements guide our design: flexible density distributions under molecular symmetry constraints, efficient computation, and compatibility with electronic-structure solvers. To meet these requirements, we start from an expansion of isotropic Gaussian functions with learnable widths, weights, and floating centers, combining spatial flexibility with efficient evaluation and analytical integration against target GTO bases. Predicting each center displacement independently, however, introduces restrictions imposed by molecular symmetry. In benzene, permutation and reflection symmetries constrain these displacements at each carbon atom to the corresponding C–H axis,

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.

limiting center placement around the delocalized π-electron cloud above and below the molecular plane. This motivates a shell-based organization of Gaussian centers.

We introduce GEM-FIELD, the Gaussian Equivariant Multi-Shell Field. Each shell distributes the centers of isotropic Gaussian functions over a sphere of learned radius d around an atom and integrates their weighted contributions over all directions. The angular weights transform equivariantly with the molecular geometry, ensuring rotational equivariance of the continuous shell density. The Gaussians within a shell share a learned width σ, while their signed angular weights are obtained by transforming the atom's higher-order equivariant features through a shared nonlinear spherical decoder. Multiple shells with different radii and widths capture spatial structure at different scales. Lebedev quadrature approximates this spherical integral as a finite weighted sum of displaced Gaussians. This Gaussian sum supports analytical Coulomb integration against target AO bases. Combined with numerical exchange–correlation integration and one generalized eigensolution, it provides electronic properties and SCF initial guesses.

Our contributions are threefold. First, GEM-FIELD combines expressive density modeling, efficient evaluation, and analytical GTO compatibility through a learnable real-space representation whose radial and angular structure adapts beyond a predefined auxiliary density basis. Second, it connects density prediction to electronic-property recovery and SCF initialization, with an output representation reusable across target AO bases and exchange–correlation functionals. Third, experiments demonstrate a favorable accuracy–cost balance, strong property extrapolation to larger molecules, and the lowest OOD RIC among the compared SCFBench models with 100% convergence and reduced end-to-end runtime.

2 RELATED WORK

Grid-native methods learn electron densities directly at spatial locations. Early work predicted densities on fixed grids (Brockherde et al., 2017; Gong et al., 2019), whereas DeepDFT and subsequent equivariant models treat arbitrary query locations as graph nodes to support systems of different sizes (Jørgensen & Bhowmik, 2020; 2022; Sunshine et al., 2023; Koker et al., 2024). Other approaches combine coefficient learning with coordinate-specific residuals, operate in plane-wave and Gaussian-orbital function spaces, construct compact grid-centered expansions, perform density super-resolution, or learn self-consistent densities from grid-projected fingerprints (Cheng & Peng, 2023; Kim & Ahn, 2024; Focassio et al., 2023; Li et al., 2025; Lee & Kim, 2024).

These methods suit real-space analysis and periodic plane-wave workflows, but inference cost typically grows with the number of query points. PAW or pseudopotential labels may also omit all-electron near-core structure. Moreover, grid outputs do not natively provide Gaussian-AO density matrices or analytical integrals, so LCAO solvers still require projection, numerical integration, or grid-to-AO conversion.

Atom-centered methods represent continuous densities using auxiliary basis functions or analytical density functions that can be evaluated across target AO bases. SALTED and related equivariant models predict coefficients of atom-centered radial functions and spherical harmonics (Grisafi et al., 2019; Lewis et al., 2021; Grisafi et al., 2023; Rackers et al., 2023). Analytical message-passing models instead learn compact atom-centered density functions (Cuevas-Zuviría & Pacios, 2021), while SCDP uses an eSCN backbone and adds bond-centered virtual nodes, even-tempered Gaussian bases, and learned exponent scaling (Passaro & Zitnick, 2023; Fu et al., 2024). ELECTRA and BOA instead use movable Gaussians and low-rank GTO products, respectively (Elsborg et al., 2025; Klockow et al., 2026). NeuralSCF further learns the Kohn–Sham density map (Song & Feng, 2026). SCFBench evaluates auxiliary-basis density coefficients for SCF initialization and demonstrates their transfer advantage over matrix targets under a unified protocol (Liu et al., 2026).

The achievable density accuracy is governed by the expressive capacity of the chosen density expansion. For a fixed finite auxiliary basis, its radial and angular span sets a representation error floor, even with optimally chosen coefficients. Additional centers, learned basis parameters, and richer angular structure can expand this representational capacity.

Matrix-learning methods predict electronic structures directly in an AO representation. SchNOrb, PhiSNet, and QHNet primarily learn Hamiltonians or related orbital quantities, whereas Hazra et al. and Graph2Mat directly predict density matrices (Schütt et al., 2019; Unke et al., 2021;

Yu et al., 2023; Hazra et al., 2024; Febrer et al., 2025). These outputs can be used directly for orbital solution, electronic-property evaluation, or SCF initialization.

This family provides the most direct interface to LCAO solvers, but matrix dimensions, block structures, and numerical scales depend on the selected AO basis and its ordering. Atom- and pair-based architectures can assemble matrices for molecules of different sizes, while changes in the AO basis require compatible output blocks or basis conversion. Predicted matrices must also satisfy physical constraints such as Hermiticity, electron count, and representability.

3 DATASETS

QM9-PBE. QM9 contains 133,885 molecules composed of H, C, N, O, and F, with at most nine heavy atoms (Ramakrishnan et al., 2014). At the provided geometries, we generate density-fitted PBE/def2-SVP references with PySCF, including converged density matrices, energy components, orbital energies, dipoles, and forces (Perdew et al., 1996; Weigend & Ahlrichs, 2005; Sun et al., 2018). Using seed 42, we assign 120,396, 100, and 13,389 molecules to training, validation, and testing; all density and property baselines use this split. Density models learn the residual between the converged and superposition-of-atomic-densities (SAD) densities.

OE62-OOD. From OE62 (Stuke et al., 2020), we select neutral, closed-shell H/C/N/O/F molecules with at least 30 atoms and sample 1,000 structures stratified by atom count (30–106 atoms). They are excluded from training and model selection and receive the same PBE/def2-SVP property, energy, and force labels as QM9-PBE, providing a size-extrapolation benchmark.

SCFBench. This ChEMBL-derived dataset contains molecules composed of H, C, N, O, F, P, and S (Liu et al., 2026). We use its official 35,089/4,386/4,386 train/validation/ID-test split (molecules with at most 20 atoms) and its separate 1,050-molecule OOD test set (up to 60 atoms). Training targets are SAD residuals constructed from the reference def2-SVP density matrices; the OOD set is used only for evaluation.

4 METHODS

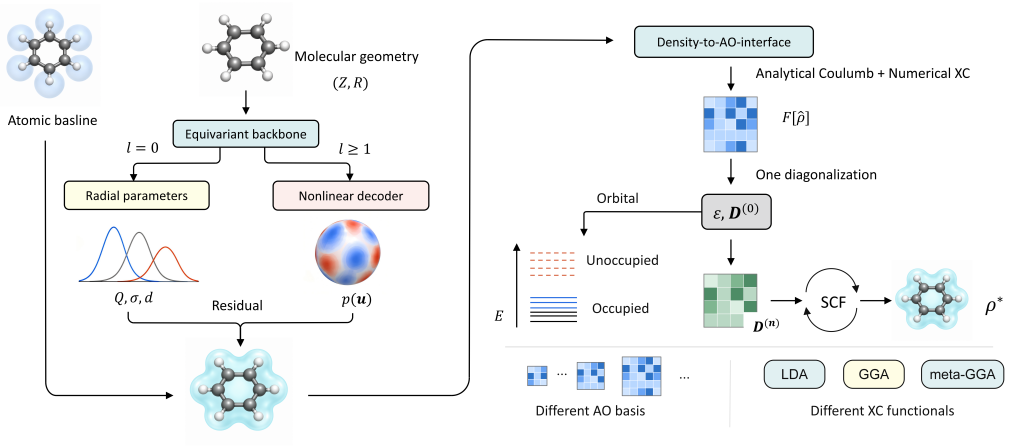


Figure 1: GEM-FIELD maps molecular geometry to an electron-number-conserving Gaussian-shell density for electronic-property evaluation and SCF initialization.

4.1 MODEL ARCHITECTURE

As shown in Figure 1, GEM-FIELD predicts the SAD residual $\Delta\rho = \rho - \rho_{\mathrm{SAD}}$ from atomic numbers Z_i and positions $\mathbf{R}_i$. An EquiformerV2 backbone supplies equivariant features to an element-specific Gaussian-shell readout (Liao et al., 2024). Each shell carries a continuous angular distribution, which Lebedev quadrature converts into a finite Gaussian expansion.

Molecular graph and equivariant features. We represent atoms as nodes and add bidirectional, self-loop-free edges between atoms within 5 Å. A learned atomic-number embedding initializes the $\ell = 0$ scalar channel. Each edge length is expanded in 64 Gaussian radial basis functions; these values and the source and target element information serve as rotation-invariant edge attributes.

At atom i, the backbone maintains natural-parity spherical-tensor features (Thomas et al., 2018)

$$\mathcal{H}_i = \bigoplus_{\ell=0}^{L} C \times \big(\ell, (-1)^\ell\big), \qquad \mathbf{h}_{ic}^{(\ell)} \overset{\mathcal{R}}{\mapsto} D^{(\ell)}(\mathcal{R})\mathbf{h}_{ic}^{(\ell)}. \tag{1}$$

Here, $C = 32$ and $L = m_{\max} = 6$. An edge-degree embedding initializes higher-order features from radial attributes and spherical harmonics. Four EquiformerV2 blocks update these features through edge-aligned $\mathrm{SO}(2)$ attention and equivariant feed-forward layers, retaining degrees $\ell = 0, \ldots, 6$. Appendix A provides implementation details.

Element-specific shell readout. For element Z, each atom receives K_Z active shells. The QM9 model uses $K_{\mathrm{H}} = 4$ and $K_Z = 16$ for C, N, O, and F; the SCFBench model additionally uses $K_Z = 24$ for P and S. Element-specific radial heads accommodate differences in nuclear charge, valence range, and density complexity. Three element-specific two-layer multilayer perceptrons act on the scalar feature $\mathbf{h}_i^{(0)}$ to predict shell charge, width, and displacement radius, where sp denotes Softplus:

$$\widetilde{Q}_{ik} = f_Q^{(Z_i)}\Big(\mathbf{h}_i^{(0)}\Big)_k, \quad \sigma_{ik} = \mathrm{sp}\Big(f_\sigma^{(Z_i)}\Big(\mathbf{h}_i^{(0)}\Big)_k\Big) + 10^{-3}, \quad d_{ik} = \mathrm{sp}\Big(f_d^{(Z_i)}\Big(\mathbf{h}_i^{(0)}\Big)_k\Big). \tag{2}$$

The hidden layers use shifted Softplus activations. We do not constrain $\widetilde{Q}_{ik}$ to be positive because a residual density and its local integral may be negative, whereas the positive maps enforce $\sigma_{ik} > 0$ and $d_{ik} \geq 0$. Element-specific equivariant linear maps convert higher-order features into H_{ang} sets of spherical-harmonic coefficients for each shell. Elements do not share radial or angular readout parameters, but the subsequent nonlinear angular decoder is shared across elements and shells.

Continuous Gaussian shells. Predicting a finite set of Gaussian displacement vectors from degree-one features is restricted by local symmetry. If atom i has stabilizer subgroup G_i, an equivariant displacement $\mathbf{c}_{ik}$ must satisfy $\mathcal{R}\mathbf{c}_{ik} = \mathbf{c}_{ik}$ for every $\mathcal{R} \in G_i$. The admissible vector subspace may therefore be low-dimensional or contain only the zero vector, limiting the representation of multidirectional bonds and lone-pair structure with finitely many independent centers.

GEM-FIELD instead defines a continuous distribution of Gaussian centers on an atom-centered sphere. Let $g_\sigma(\mathbf{r})$ denote a normalized isotropic Gaussian of width σ. Shell k on atom i is

$$\widehat{\Delta\rho}_{ik}(\mathbf{r}) = Q_{ik} \int_{S^2} p_{ik}(\mathbf{u}) g_{\sigma_{ik}}(\mathbf{r} - \mathbf{R}_i + d_{ik}\mathbf{u})\, d\mathbf{u}. \tag{3}$$

Direction $\mathbf{u}$ places a Gaussian at the virtual center $\mathbf{R}_i - d_{ik}\mathbf{u}$. The parameters Q_{ik}, σ_{ik}, and d_{ik} control the integrated shell charge, radial width, and displacement radius, while $p_{ik}(\mathbf{u})$ controls the directional contributions.

Nonlinear angular decoding. An element-specific equivariant linear layer produces H_{ang} coefficient sets for each shell. Coefficient set h defines a spherical scalar field

$$f_{ikh}(\mathbf{u}) = \sum_{\ell=1}^{L} \sum_{m=-\ell}^{\ell} a_{ik\ell m}^{(h)} Y_{\ell m}(\mathbf{u}), \qquad h = 1, \ldots, H_{\mathrm{ang}}. \tag{4}$$

We use $H_{\mathrm{ang}} = 8$. A shared pointwise decoder $\Phi : \mathbb{R}^8 \to \mathbb{R}$ acts on every direction, element, and shell:

$$z_{ik}(\mathbf{u}) = \Phi\big(f_{ik1}(\mathbf{u}), \ldots, f_{ikH_{\mathrm{ang}}}(\mathbf{u})\big). \tag{5}$$

The decoder contains a width-16 hidden layer, a SiLU activation, and a scalar output. Because each input field satisfies $f'_{ikh}(\mathbf{u}) = f_{ikh}(\mathcal{R}^{-1}\mathbf{u})$, the shared pointwise map satisfies

$$z'_{ik}(\mathbf{u}) = z_{ik}(\mathcal{R}^{-1}\mathbf{u}). \tag{6}$$

The continuous angular field is therefore rotation equivariant. The pointwise nonlinearity generates higher-order angular couplings unavailable to a linear truncated harmonic expansion, while the output remains a spherical function and is not projected back onto truncated harmonic coefficients.

Angular normalization and charge conservation. The nonlinear decoder should redistribute a shell angularly without changing its integrated charge. We subtract the spherical mean of z_{ik} and add a uniform background:

$$\bar{z}_{ik} = \frac{1}{4\pi}\int_{S^2} z_{ik}(\mathbf{u})\, d\mathbf{u}, \qquad p_{ik}(\mathbf{u}) = \frac{1}{4\pi} + z_{ik}(\mathbf{u}) - \bar{z}_{ik}. \tag{7}$$

Consequently, $\int_{S^2} p_{ik}(\mathbf{u})d\mathbf{u} = 1$. The field p_{ik} is a signed angular weight rather than a probability density. Since p_{ik} and $g_{\sigma_{ik}}$ are normalized, the spatial integral of each shell equals Q_{ik}.

We also apply a uniform charge correction to all active shells in a molecule. With $N_{\mathrm{sh}} = \sum_i K_{Z_i}$, the corrected charge is

$$Q_{ik} = \widetilde{Q}_{ik} + \frac{\Delta N_e - \sum_j \sum_{s=1}^{K_{Z_j}} \widetilde{Q}_{js}}{N_{\mathrm{sh}}}, \qquad \Delta N_e = \int_{\mathbb{R}^3} \Delta\rho(\mathbf{r})\, d\mathbf{r}. \tag{8}$$

Thus, $\int \widehat{\Delta\rho}(\mathbf{r})d\mathbf{r} = \Delta N_e$, and the total density has the correct electron count. When SAD is built from neutral atomic densities, ΔN_e is the negative molecular charge and is zero for the neutral closed-shell systems considered here.

Discrete Gaussian sum. After summing over all atoms and shells, the complete residual density is

$$\begin{aligned}\widehat{\Delta\rho}(\mathbf{r}) &= \sum_i \sum_{k=1}^{K_{Z_i}} Q_{ik} \int_{S^2} p_{ik}(\mathbf{u}) g_{\sigma_{ik}}(\mathbf{r} - \mathbf{R}_i + d_{ik}\mathbf{u})\, d\mathbf{u} \\ &\approx \sum_i \sum_{k=1}^{K_{Z_i}} \sum_{j=1}^{J_{\mathrm{Leb}}} \frac{\omega_j Q_{ik} p_{ik}(\mathbf{u}_j)}{\pi^{3/2}\sigma_{ik}^3} \exp\left[-\frac{\|\mathbf{r} - \mathbf{R}_i + d_{ik}\mathbf{u}_j\|^2}{\sigma_{ik}^2}\right].\end{aligned} \tag{9}$$

We use an order-15 Lebedev rule with $J_{\mathrm{Leb}} = 86$ directions to convert each continuous spherical integral into a finite weighted sum of Gaussian functions (Lebedev, 1975). For term (i, k, j), the coefficient, exponent, and virtual center are $\omega_j Q_{ik} p_{ik}(\mathbf{u}_j)/(\pi^{3/2}\sigma_{ik}^3)$, σ_{ik}^{-2}, and $\mathbf{R}_i - d_{ik}\mathbf{u}_j$, respectively. The same quadrature rule is used for angular normalization, so each discretized shell retains integrated charge Q_{ik}. In its continuous definition, GEM-FIELD is equivalent to an infinite-order radial spherical-harmonic expansion, which underpins its expressive power; Appendix D provides a detailed proof.

4.2 TRAINING AND ELECTRONIC-STRUCTURE INTERFACE

Training and evaluation. PySCF constructs SAD in the def2-SVP AO basis, and the supervision target is the difference between the converged and SAD densities (Sun et al., 2018). The model fits this signed residual using the molecule-averaged spatial L_1 loss

$$\mathcal{L}_{\mathrm{train}} = \frac{1}{B}\sum_{b=1}^{B} \int_{\mathbb{R}^3} \left|\widehat{\Delta\rho}_b(\mathbf{r}) - \Delta\rho_b(\mathbf{r})\right| d\mathbf{r}. \tag{10}$$

To reduce storage, density values are generated on the fly from density matrices during training and evaluation. Training and validation use PySCF level-0 grids. At each training visit, we sample 1,000 grid points uniformly without replacement and estimate the integral using absolute quadrature weights and a full-to-sampled grid-size correction. After model selection, frozen checkpoints are tested on the complete level-0 grid and the complete unpruned level-3 grid. The latter evaluates density accuracy on a dense grid and sensitivity to quadrature resolution. Both evaluations compute the molecule-level normalized mean absolute error $\mathrm{NMAE}_b = \frac{\int_{\mathbb{R}^3} |\hat{\rho}_b(\mathbf{r}) - \rho_b(\mathbf{r})|\, d\mathbf{r}}{N_{e,b}}$, and report the mean over molecules. Detailed training settings and hyperparameter configurations are provided in Appendix A.

Fock construction from the predicted density. The total prediction $\hat{\rho} = \rho_{\mathrm{SAD}} + \widehat{\Delta\rho}$ defines a one-shot PBE Fock matrix in the target AO basis (Perdew et al., 1996). Writing $q_{ikj} = \omega_j Q_{ik} p_{ik}(\mathbf{u}_j)$ for the integrated charge of a discrete Gaussian,

$$F_{\mu\nu}[\hat{\rho}] = h_{\mu\nu} + J_{\mu\nu}^{\mathrm{SAD}} + J_{\mu\nu}^{\mathrm{res}} + V_{\mu\nu}^{\mathrm{xc}}[\hat{\rho}], \qquad J_{\mu\nu}^{\mathrm{res}} = \sum_{i,k,j} q_{ikj}(\mu\nu \mid g_{ik,\mathbf{u}_j}). \tag{11}$$

Table 1: Electron-density prediction on the all-electron QM9-PBE test set. The two error columns report NMAE on the level-0 and complete unpruned level-3 grids. Inference time is normalized to GEM-FIELD with $L_{\max} = 4$.

Model	# Param.	Level-0 NMAE (%)	Unpruned level-3 NMAE (%)	Inference time ↓
GEM-FIELD ($L_{\max} = 6$)	3.54M	**0.037**	0.226	1.35×
GEM-FIELD ($L_{\max} = 4$)	**1.53M**	0.055	0.237	1.00×
SCDP ($L_{\max} = 4$)	1.98M	0.107	0.359	1.44×
ELECTRA	1.68M	0.269	0.360	**0.15**×
ChargE3Net ($L_{\max} = 4$)	1.91M	0.145	**0.198**	26.90×
BOA (def2-TZVP)	1.58M	0.072	0.272	1.60×

Here, h is the one-electron contribution, and PySCF's direct-J engine evaluates J^{SAD} from the SAD density matrix. Every component of the discrete Gaussian residual is a displaced s-type Gaussian, so its three-center two-electron Coulomb integral $(\mu\nu \mid g_{ik,\mathbf{u}_j})$ with an AO pair is analytical. Because the exchange–correlation functional is nonlinear, $V^{\mathrm{xc}}[\hat{\rho}]$ is not separated into SAD and residual terms; PySCF instead evaluates it numerically from the total predicted density and its gradient on a level-3 grid. At the few grid points where the predicted density is negative, the exchange–correlation potential and energy density are set to zero; Coulomb construction instead retains the signed density contributions, avoiding the net-charge bias that clipping would introduce.

After assembling the Fock matrix, we solve only one generalized eigenvalue problem:

$$\mathbf{FC} = \mathbf{SC}\boldsymbol{\varepsilon}, \qquad \mathbf{D}^{(0)} = 2\mathbf{C}_{\mathrm{occ}}\mathbf{C}_{\mathrm{occ}}^{\mathsf{T}}. \tag{12}$$

The interface supports two downstream uses. For property recovery, orbital energies are read from $\boldsymbol{\varepsilon}$, energy components and total energies are evaluated from $\mathbf{D}^{(0)}$, and approximate forces are computed using the first-diagonalization quantities. Dipole moments follow directly from $\hat{\rho}$. For SCF initialization, $\mathbf{D}^{(0)}$ is passed to the conventional restricted Kohn–Sham solver, which iterates to the specified convergence thresholds.

5 EXPERIMENTS

5.1 ELECTRON-DENSITY BENCHMARK

We retrain ChargE3Net, SCDP, ELECTRA, and BOA on the same QM9-PBE split and supervision target (Koker et al., 2024; Fu et al., 2024; Elsborg et al., 2025; Klockow et al., 2026). Every model learns the signed residual from SAD. Residual learning reduces the dynamic range near nuclei, but the total reference remains all-electron and is not pseudopotential-smoothed. Fixed-voxel convolutional models require interpolation and a changed input definition on our nonuniform grids, so they are excluded from the quantitative comparison. The evaluated set covers the query-based ChargE3Net and the analytical or basis-function representations of SCDP, ELECTRA, and BOA.

All four baselines use approximately two million parameters or fewer and are trained on a single GPU with at most 32 GB of memory. The number of grid points sampled per molecule is set separately for each model to fit within the GPU memory limit. All methods share the data split, supervision target, and evaluation molecules. Complete training configurations for the reported checkpoints are provided in Appendix B.1.

GEM-FIELD achieves a favorable balance between density accuracy and evaluation efficiency. At comparable parameter counts, its $L_{\max} = 4$ variant achieves lower errors than SCDP and BOA on both grids while evaluating faster. On the complete unpruned level-3 grid, it achieves 0.237% NMAE, close to ChargE3Net's best result of 0.198%, while reducing evaluation time by 96.3%. Increasing $L_{\max}$ to six further improves density accuracy, yielding the lowest level-0 NMAE of 0.037% and a dense-grid NMAE of 0.226%. Together, these results establish GEM-FIELD as an accurate and computationally efficient density representation for the following property and solver evaluations.

Table 2: Electronic-property MAEs on 13,389 QM9-PBE (ID) and 1,000 OE62 (OOD) molecules. Bold values indicate the lower error within each split.

	Energy-component MAE (meV atom^{-1})				Orbital-energy MAE (meV)			Dipole MAE (D)
Method	E_{kin}	E_{ne}	E_J	E_{xc}	HOMO	LUMO	Gap	$\lvert\boldsymbol{\mu}\rvert$
In distribution								
GEM-FIELD	94.83	**201.28**	**116.59**	10.12	**8.64**	**8.18**	**4.23**	**0.0083**
EquiformerV2	**2.07**	718.99	260.92	**0.81**	25.53	25.02	43.51	0.0147
Out of distribution								
GEM-FIELD	**129.80**	**265.02**	**156.96**	**14.22**	**27.49**	**30.61**	**33.94**	**0.1830**
EquiformerV2	161.94	1,205,020.91	604,164.28	54.09	209.55	230.02	472.79	0.1848

5.2 ELECTRONIC-PROPERTY BENCHMARK

We evaluate total energies, forces, energy components, frontier orbitals, and dipoles derived from predicted densities on 13,389 QM9-PBE test molecules. GEM-FIELD receives no property-specific retraining. For comparison, separate EquiformerV2 models are trained for each energy component, frontier-orbital target, and the dipole vector. Each of three machine-learned classical interatomic potentials is jointly trained on energies and forces. The tables report mean absolute errors (MAEs), with energy components normalized per atom. Appendices B.2 and B.3 give the corresponding training configurations.

The main benefit of GEM-FIELD emerges under distribution shift, where the density intermediate preserves consistency across electronic properties. On ID molecules, its electron–nuclear and Coulomb MAEs are 201.28 and 116.59 meV/atom, 72.0% and 55.3% below direct prediction. EquiformerV2 remains more accurate for kinetic and exchange–correlation energies, reflecting the local advantage of optimizing each scalar target directly. This advantage does not transfer to OE62. There, GEM-FIELD obtains the lowest errors for all four components, which increase to only 1.32–1.40 times their ID values. Its electron–nuclear and Coulomb errors remain 265.02 and 156.96 meV/atom. The direct counterparts instead reach 1.21×10^6 and 6.04×10^5 meV/atom. Direct regression can therefore optimize individual ID scalars, whereas density-mediated inference better resists system-size shift.

The same pattern extends to Fock-derived frontier orbitals and density-derived dipoles. On ID molecules, GEM-FIELD lowers the HOMO, LUMO, and gap errors to 8.64, 8.18, and 4.23 meV, respectively. These values improve on direct prediction by 66.1%–90.3%, while the dipole MAE decreases by 43.9% to 0.0083 D. On OE62, all three orbital errors remain below 34 meV and are 86.7%–92.8% below EquiformerV2. The dipole errors are comparable at 0.1830 D for GEM-FIELD and 0.1848 D for the direct model. A single predicted density therefore retains a pronounced orbital advantage while matching the target-specific dipole model on larger molecules. The larger OOD dipole error also identifies accumulated long-range charge error as a remaining limitation.

We further compare total energies and atomic forces on the same split using EquiformerV2, NequIP, and MACE as direct interatomic-potential baselines (Liao et al., 2024; Batzner et al., 2022; Batatia et al., 2022). Total-energy error is measured per molecule. Force MAE covers all Cartesian atomic components.

For total energies and atomic forces, the principal advantage of GEM-FIELD is again its transfer across molecular size. On ID molecules, its energy MAE is 0.61 meV/molecule, 93.5% below EquiformerV2. This accuracy reflects cancellation at the molecule level: the signed sum of the four electronic-component errors has an MAE of 0.61 meV/molecule, whereas the mean sum of their absolute errors is 7.55 eV/molecule. EquiformerV2 attains the lowest ID force MAE, 2.34 meV/Å, compared with 4.81 meV/Å for GEM-FIELD. However, GEM-FIELD has the lowest force RMSE, 12.77 meV/Å, indicating better control of large deviations. On OE62, its energy MAE, force MAE, and force RMSE remain 13.32 meV/molecule, 24.65, and 61.94 meV/Å, respectively. These values improve on the strongest interatomic-potential baselines by 93.9%, 76.4%, and 57.3%. The OOD advantage shows that ID error cancellation is not the sole source of accuracy. Task-specific potentials retain an advantage for typical ID force components, but the density intermediate yields

Table 3: Total-energy and force errors for density-derived and direct interatomic-potential predictions.

Model	Energy MAE (meV molecule^{-1})		Force MAE (meV Å^{-1})		Force RMSE (meV Å^{-1})	
	ID	OOD	ID	OOD	ID	OOD
GEM-FIELD	**0.61**	**13.32**	4.81	**24.65**	**12.77**	**61.94**
EquiformerV2	9.46	3,706.67	**2.34**	108.34	13.90	155.50
NequIP ($L_{\max} = 1$)	95.64	981.60	8.39	114.21	19.12	158.01
NequIP ($L_{\max} = 2$)	69.78	486.84	6.66	110.02	15.81	153.82
MACE ($L_{\max} = 1$)	51.03	257.17	12.15	109.09	21.50	150.31
MACE ($L_{\max} = 2$)	44.51	218.00	11.32	104.59	21.02	145.06

Table 4: SCFBench convergence and RIC. Baseline results follow the benchmark's unified protocol, including QHFlow from its Table 10 (Liu et al., 2026). A dash denotes an unreported parameter count; bold values mark the best metrics and smallest reported parameter count.

Prediction target	Method	# Param.	ID		OOD	
			Conv. (%) ↑	RIC (%) ↓	Conv. (%) ↑	RIC (%) ↓
Hamiltonian	QHNet	20.5M	**100%**	63.20%	97.43%	179.47%
	QHFlow	-	**100%**	**56.80%**	98.29%	146.97%
Density matrix	QHNet	20.5M	**100%**	70.45%	99.71%	91.69%
Density coefficients (def2-universal-jfit)	NequIP-L	50.0M	**100%**	63.78%	**100%**	66.68%
Electron density	**GEM-FIELD (ours)**	**3.6M**	**100%**	58.51%	**100%**	**59.49%**

more transferable joint energy–force predictions. Because the predicted density matrix is not self-consistent, the reported forces are not rigorously conservative. Appendix G shows by finite differences on water, ethanol, acetamide, and benzene that the analytic forces closely approximate the gradient of the one-shot energy at these geometries.

5.3 SCF ACCELERATION

We test whether predicted densities provide high-quality SCF initial guesses on SCFBench (Liu et al., 2026). The benchmark contains an ID test set and an OOD set of larger molecules. We train GEM-FIELD on the provided density matrices using the official split. Following the benchmark protocol, model-seeded SCF calculations are compared with default MINAO initialization. The primary metric is relative iteration count (RIC), where lower values indicate fewer iterations.

GEM-FIELD achieves the lowest OOD RIC among the compared models, with 100% convergence on both splits and the smallest reported parameter count. All 4,386 ID calculations converge, with an RIC of 58.51%, compared with 56.80% for QHFlow and 63.78% for the auxiliary-density baseline. All 1,050 OOD calculations also converge, and their RIC of 59.49% improves on the previous best by 7.19 points. Hamiltonian- and density-matrix-based models instead lose convergence or require substantially more iterations on larger molecules. Using unrounded values, the GEM-FIELD RIC increases by only 0.97 points from ID to OOD. Its predicted density therefore provides reliable initialization while maintaining stable acceleration under molecular-size shift. With analytical Coulomb construction, end-to-end wall-clock time decreases by 18.69% on ID and 8.98% on OOD, corresponding to $1.23\times$ and $1.10\times$ speedups over MINAO. Numerical Coulomb construction further improves these speedups to $1.32\times$ and $1.17\times$, respectively, with nearly identical SCF iteration counts and 100% convergence. Appendix C provides the implementation details and complete timing analysis for both approaches.

Relative iteration savings also transfer to functionals and AO bases different from the PBE/def2-SVP training setting. Cross-functional tests retain def2-SVP and replace PBE with BLYP, SCAN, B3LYP, or PBE0. For the SCAN meta-GGA test, the kinetic-energy density is approximated using the

Table 5: Cross-functional and cross-basis SCF transfer. Entries are RIC values in percent; lower is better.

Model	def2-SVP				PBE		
	BLYP	SCAN	B3LYP	PBE0	def2-TZVP	def2-TZVPPD	def2-QZVP
In distribution							
NequIP-L	**71.38%**	88.15%	**84.63%**	**85.99%**	76.68%	77.07%	77.81%
GEM-FIELD	72.08%	**85.40%**	84.77%	86.47%	**71.65%**	**71.44%**	**71.93%**
Out of distribution							
NequIP-L	71.22%	86.45%	**83.72%**	**85.51%**	75.24%	75.81%	75.98%
GEM-FIELD	**70.43%**	**83.53%**	84.07%	85.89%	**68.99%**	**68.72%**	**69.08%**

von Weizsäcker form, whereas for the B3LYP and PBE0 hybrid-functional tests, the Hartree–Fock exchange matrix K is constructed from the MINAO initial-guess density matrix. Cross-basis tests retain PBE and use def2-TZVP, def2-TZVPPD, or def2-QZVP. Across functionals, GEM-FIELD averages 82.18% RIC on ID and 80.98% on OOD, improving on NequIP-L by 0.36 and 0.75 points. The largest functional advantage occurs for SCAN, at 2.75 ID and 2.92 OOD points. Basis transfer produces a clearer separation: GEM-FIELD lowers all three RIC values, by 5.51 points on average for ID and 6.75 points for OOD. Across all seven targets, no OOD RIC exceeds its corresponding ID result; instead, the OOD values are 0.58–2.85 points lower. The learned density is therefore reusable beyond its training calculation level and is particularly robust to basis-set expansion and molecular-size growth.

5.4 ABLATION STUDIES

Table 6: Ablations on QM9-PBE evaluated on the level-0 and complete unpruned level-3 grids.

Model	Setting	Level-0 NMAE ↓ (%)	Unpruned level-3 NMAE ↓ (%)
EquiformerV2	$L_{\max} = 6$	**0.037**	**0.226**
EquiformerV2	$L_{\max} = 5$	0.041	0.241
EquiformerV2	$L_{\max} = 4$	0.055	0.237
EquiformerV2	$L_{\max} = 3$	0.066	0.257
eSCN	$L_{\max} = 3$	0.079	0.263
NequIP	$L_{\max} = 3$	0.117	0.266
NequIP	$L_{\max} = 3$, linear decoder	0.346	0.660

The ablations identify the nonlinear angular decoder as the dominant source of accuracy improvement at both grid resolutions. At fixed $L_{\max} = 3$, backbone choice has a smaller effect, whereas increasing the angular-momentum cutoff generally improves accuracy and gives the best results at $L_{\max} = 6$. The water-molecule spectrum further shows that nonlinear decoding activates more spherical-harmonic modes and extends angular expressivity beyond the backbone cutoff (Appendix E).

6 CONCLUSION

This work establishes a solver-oriented paradigm for electron-density learning: electron density is not merely a prediction target, but a general intermediate representation connecting molecular geometry, electronic properties, and quantum-chemistry solvers. Through continuous Gaussian shells, nonlinear angular expressivity, and exact charge conservation, GEM-FIELD converts geometric neural representations into an analytical electronic object that can be integrated, differentiated, and reevaluated in different target bases. The same predicted density consequently supports dense-grid evaluation, electronic-property recovery, one-shot Fock construction, energy and force calculation, and SCF initialization, while remaining transferable across molecular sizes and computational settings. More importantly, this representation expands the role of machine learning from predicting density values to providing an electronic-structure interface that can be embedded directly into existing Gaussian-AO software and solver workflows. The present study validates this paradigm for neutral closed-shell molecules, the PBE/def2-SVP training setting, and the element range considered here. Future work will refine the meta-GGA and hybrid-functional interfaces by learning kinetic-energy density and nonlocal exchange contributions, and extend the framework to charged and spin-resolved open-shell systems with appropriate charge- and spin-conservation constraints. Overall, GEM-FIELD demonstrates the potential of physically constrained analytical electron density as a next-generation interface for machine-learning-accelerated quantum chemistry.

AI USE STATEMENT

We used generative AI tools to improve the clarity of the manuscript, identify potentially relevant literature, and assist with implementing and debugging research code. The authors checked suggested references against the original publications and reviewed and tested AI-assisted code. The authors verified the scientific claims and reported results and take full responsibility for the manuscript and accompanying artifacts.

REPRODUCIBILITY STATEMENT

Dataset sources and split protocols are described in Section 3; model, training, and evaluation details are provided in Section 4 and the Appendix. The anonymized supplementary code package contains implementations, dependency specifications, configurations, and split manifests.Reproducing the reported numerical results therefore requires obtaining the cited source datasets and obtaining or regenerating the required reference density matrices before running the supplied code.

REFERENCES

Ilyes Batatia, Dávid Péter Kovács, Gregor N. C. Simm, Christoph Ortner, and Gábor Csányi. MACE: Higher order equivariant message passing neural networks for fast and accurate force fields. In *Advances in Neural Information Processing Systems*, volume 35, pp. 11423–11436, 2022. doi: 10.52202/068431-0830. URL https://papers.nips.cc/paper_files/paper/2022/hash/4a36c3c51af11ed9f34615b81edb5bbc-Abstract-Conference.html.

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

Felix Brockherde, Leslie Vogt, Li Li, Mark E. Tuckerman, Kieron Burke, and Klaus-Robert Müller. Bypassing the Kohn–Sham equations with machine learning. *Nature Communications*, 8:872, 2017. doi: 10.1038/s41467-017-00839-3.

Chaoran Cheng and Jian Peng. Equivariant neural operator learning with graphon convolution. In *Advances in Neural Information Processing Systems*, volume 36, pp. 61960–61984, 2023. doi: 10.52202/075280-2709. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/c362fbc0d182c6b4b8dadb90177239e4-Abstract-Conference.html.

Bruno Cuevas-Zuviría and Luis F. Pacios. Machine learning of analytical electron density in large molecules through message-passing. *Journal of Chemical Information and Modeling*, 61(6): 2658–2666, 2021. doi: 10.1021/acs.jcim.1c00227.

Jonas Elsborg, Luca Thiede, Alán Aspuru-Guzik, Tejs Vegge, and Arghya Bhowmik. ELECTRA: A cartesian network for 3d charge density prediction with floating orbitals. In *Advances in Neural Information Processing Systems*, volume 38, pp. 31897–31926, 2025. doi: 10.52202/085713-0947. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/288b63aa98084366c4536ba0574a0f22-Abstract-Conference.html.

Pol Febrer, Peter Bjørn Jørgensen, Miguel Pruneda, Alberto García, Pablo Ordejón, and Arghya Bhowmik. Graph2Mat: Universal graph to matrix conversion for electron density prediction. *Machine Learning: Science and Technology*, 6(2):025013, 2025. doi: 10.1088/2632-2153/adc871.

Bruno Focassio, Michelangelo Domina, Urvesh Patil, Adalberto Fazzio, and Stefano Sanvito. Linear Jacobi–Legendre expansion of the charge density for machine learning-accelerated electronic structure calculations. *npj Computational Materials*, 9:87, 2023. doi: 10.1038/s41524-023-01053-0.

Xiang Fu, Andrew Rosen, Kyle Bystrom, Rui Wang, Albert Musaelian, Boris Kozinsky, Tess Smidt, and Tommi Jaakkola. A recipe for charge density prediction. In *Advances in Neural Information Processing Systems*, volume 37, pp. 9727–9752, 2024. doi: 10.52202/079017-0310. URL https://papers.nips.cc/paper_files/paper/2024/hash/12c118ef87fde56a10bd858842781b34-Abstract-Conference.html.

Sheng Gong, Tian Xie, Taishan Zhu, Shuo Wang, Eric R. Fadel, Yawei Li, and Jeffrey C. Grossman. Predicting charge density distribution of materials using a local-environment-based graph convolutional network. *Physical Review B*, 100(18):184103, 2019. doi: 10.1103/PhysRevB.100.184103.

Andrea Grisafi, Alberto Fabrizio, Benjamin Meyer, David M. Wilkins, Clemence Corminboeuf, and Michele Ceriotti. Transferable machine-learning model of the electron density. *ACS Central Science*, 5(1):57–64, 2019. doi: 10.1021/acscentsci.8b00551.

Andrea Grisafi, Alan M. Lewis, Mariana Rossi, and Michele Ceriotti. Electronic-structure properties from atom-centered predictions of the electron density. *Journal of Chemical Theory and Computation*, 19(14):4451–4460, 2023. doi: 10.1021/acs.jctc.2c00850.

Suman Hazra, Urvesh Patil, and Stefano Sanvito. Predicting the one-particle density matrix with machine learning. *Journal of Chemical Theory and Computation*, 20(11):4569–4578, 2024. doi: 10.1021/acs.jctc.4c00042.

Pierre Hohenberg and Walter Kohn. Inhomogeneous electron gas. *Physical Review*, 136(3B): B864–B871, 1964. doi: 10.1103/PhysRev.136.B864.

Peter Bjørn Jørgensen and Arghya Bhowmik. DeepDFT: Neural message passing network for accurate charge density prediction. *arXiv preprint arXiv:2011.03346*, 2020. URL https://arxiv.org/abs/2011.03346.

Peter Bjørn Jørgensen and Arghya Bhowmik. Equivariant graph neural networks for fast electron density estimation of molecules, liquids, and solids. *npj Computational Materials*, 8:183, 2022. doi: 10.1038/s41524-022-00863-y.

Seongsu Kim and Sungsoo Ahn. Gaussian plane-wave neural operator for electron density estimation. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 23805–23824, 2024. URL https://proceedings.mlr.press/v235/kim24b.html.

Manuel Viktor Klockow, Marc Ickler, Peter Lippmann, and Fred A. Hamprecht. A function-centric graph neural network approach for predicting electron densities. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/718573eff1cb169316783d3e08514b5b-Abstract-Conference.html.

Walter Kohn and Lu Jeu Sham. Self-consistent equations including exchange and correlation effects. *Physical Review*, 140(4A):A1133–A1138, 1965. doi: 10.1103/PhysRev.140.A1133.

Teddy Koker, Keegan Quigley, Eric Taw, Kevin Tibbetts, and Lin Li. Higher-order equivariant neural networks for charge density prediction in materials. *npj Computational Materials*, 10:161, 2024. doi: 10.1038/s41524-024-01343-1.

V. I. Lebedev. Values of the nodes and weights of ninth to seventeenth order Gauss–Markov quadrature formulae invariant under the octahedron group with inversion. *USSR Computational Mathematics and Mathematical Physics*, 15(1):44–51, 1975. doi: 10.1016/0041-5553(75)90133-0.

Ryong-Gyu Lee and Yong-Hoon Kim. Convolutional network learning of self-consistent electron density via grid-projected atomic fingerprints. *npj Computational Materials*, 10:248, 2024. doi: 10.1038/s41524-024-01433-0.

Alan M. Lewis, Andrea Grisafi, Michele Ceriotti, and Mariana Rossi. Learning electron densities in the condensed phase. *Journal of Chemical Theory and Computation*, 17(11):7203–7214, 2021. doi: 10.1021/acs.jctc.1c00576.

Chenghan Li, Or Sharir, Shunyue Yuan, and Garnet Kin-Lic Chan. Image super-resolution inspired electron density prediction. *Nature Communications*, 16:4811, 2025. doi: 10.1038/s41467-025-60095-8.

Yi-Lun Liao, Brandon M. Wood, Abhishek Das, and Tess Smidt. EquiformerV2: Improved equivariant transformer for scaling to higher-degree representations. In *International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=mCOBKZmrzD.

Zhe Liu, Yuyan Ni, Zhichen Pu, Qiming Sun, Siyuan Liu, and Wen Yan. Towards a transferable acceleration method for density functional theory. In *International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=JNuk3yGDKE.

Saro Passaro and C. Lawrence Zitnick. Reducing SO(3) convolutions to SO(2) for efficient equivariant GNNs. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 27420–27438, 2023. URL https://proceedings.mlr.press/v202/passaro23a.html.

John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Physical Review Letters*, 77(18):3865–3868, 1996. doi: 10.1103/PhysRevLett.77.3865.

Joshua A. Rackers, Lucas Tecot, Mario Geiger, and Tess E. Smidt. A recipe for cracking the quantum scaling limit with machine learned electron densities. *Machine Learning: Science and Technology*, 4(1):015027, 2023. doi: 10.1088/2632-2153/acb314.

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

Kristof T. Schütt, Michael Gastegger, Alexandre Tkatchenko, Klaus-Robert Müller, and Reinhard J. Maurer. Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions. *Nature Communications*, 10:5024, 2019. doi: 10.1038/s41467-019-12875-2.

Feitong Song and Ji Feng. Neural network self-consistent fields for density functional theory. *npj Computational Materials*, 12:289, 2026. doi: 10.1038/s41524-026-02110-0.

Annika Stuke, Christian Kunkel, Dorothea Golze, Milica Todorović, Johannes T. Margraf, Karsten Reuter, Patrick Rinke, and Harald Oberhofer. Atomic structures and orbital energies of 61,489 crystal-forming organic molecules. *Scientific Data*, 7:58, 2020. doi: 10.1038/s41597-020-0385-y.

Qiming Sun, Timothy C. Berkelbach, Nick S. Blunt, George H. Booth, Sheng Guo, Zhendong Li, Junzi Liu, James D. McClain, Elvira R. Sayfutyarova, Sandeep Sharma, Sebastian Wouters, and Garnet Kin-Lic Chan. PySCF: The python-based simulations of chemistry framework. *WIREs Computational Molecular Science*, 8(1):e1340, 2018. doi: 10.1002/wcms.1340.

Ethan M. Sunshine, Muhammed Shuaibi, Zachary W. Ulissi, and John R. Kitchin. Chemical properties from graph neural network-predicted electron densities. *The Journal of Physical Chemistry C*, 127 (48):23459–23466, 2023. doi: 10.1021/acs.jpcc.3c06157.

Nathaniel Thomas, Tess Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation- and translation-equivariant neural networks for 3d point clouds. *arXiv preprint arXiv:1802.08219*, 2018. URL https://arxiv.org/abs/1802.08219.

Oliver T. Unke, Mihail Bogojeski, Michael Gastegger, Mario Geiger, Tess E. Smidt, and Klaus-Robert Müller. SE(3)-equivariant prediction of molecular wavefunctions and electronic densities. In *Advances in Neural Information Processing Systems*, volume 34, pp. 14434–14447, 2021. URL https://papers.nips.cc/paper/2021/hash/78f1893678afbeaa90b1fa01b9cfb860-Abstract.html.

Florian Weigend and Reinhart Ahlrichs. 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. doi: 10.1039/B508541A.

Xiaojie Wu, Qiming Sun, Zhichen Pu, Tianze Zheng, Wenzhi Ma, Wen Yan, Yu Xia, Zhengxiao Wu, Mian Huo, Xiang Li, Weiluo Ren, Sheng Gong, Yumin Zhang, and Weihao Gao. Enhancing GPU-acceleration in the python-based simulations of chemistry frameworks. *WIREs Computational Molecular Science*, 15(2):e70008, 2025. doi: 10.1002/wcms.70008.

Haiyang Yu, Zhao Xu, Xiaofeng Qian, Xiaoning Qian, and Shuiwang Ji. Efficient and equivariant graph networks for predicting quantum Hamiltonian. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 40412–40424, 2023. URL https://proceedings.mlr.press/v202/yu23i.html.

A Model Training and Implementation Details

A.1 Implementation of the GEM-Field Backbone and Density Readout

GEM-Field is implemented in PyTorch and is trained in single precision by default. Molecular geometries are represented internally in bohr, and the graph cutoff is $5.0\,\text{Å}$. Atom pairs inside the cutoff are recorded as directed edges without self-loops. Atomic numbers are mapped by a learnable embedding to the $\ell = 0$ channels, and edge lengths are expanded in 64 equally spaced Gaussian radial basis functions. Node-degree embeddings and message aggregation are both normalized using an average neighbor count of 8.

The main model contains four EquiformerV2 blocks, with 32 spherical channels at every angular degree. Each block uses four $\mathrm{SO}(2)$ attention heads, an attention hidden width of 32, attention-weight and value widths of 8, and an equivariant feed-forward width of 64. We use separable S^2 activations, scaled SiLU activations, and spherical-harmonic RMS normalization. Attention, projection, and stochastic-depth dropout rates are all zero. Table A1 summarizes the principal architecture settings for the QM9-PBE and SCFBench models.

Table A1: Principal architecture settings for GEM-Field. "Active shells" denotes model density shells, not AO shells.

Setting	QM9-PBE	SCFBench
EquiformerV2 blocks	4	4
$L_{\max} = m_{\max}$	6	6
Spherical channels per degree	32	32
$\mathrm{SO}(2)$ attention heads	4	4
Radial basis functions	64	64
Cutoff radius	$5.0\,\text{Å}$	$5.0\,\text{Å}$
Angular latent heads H_{ang}	8	8
Pointwise decoder hidden width	16	16
Lebedev degree/directions	15/86	15/86
Active H shells	4	4
Active C/N/O/F shells	16/16/16/16	16/16/16/16
Active P/S shells	n/a	24/24
# Param.	~ 3.54M	~ 3.60M

The density readout uses element-specific shell-parameter networks. Charge, width, and displacement radius are predicted by two-layer MLPs whose hidden width equals the 32 scalar channels. The charge branch uses a shifted Softplus activation and does not constrain its output to be positive. The width and displacement branches output $\mathrm{softplus}(x) + 10^{-3}$ and $\mathrm{softplus}(x)$, respectively. Higher-order features are first mapped by element-specific equivariant linear layers to eight spherical fields, after which a shared $8 \to 16 \to 1$ pointwise decoder produces nonlinear angular weights. The hidden layer uses SiLU. Spherical-mean subtraction and a molecule-wide shell-charge correction enforce single-shell normalization and conservation of the total electron number, respectively.

A.2 Training Objective, Sampling, and Optimization

All GEM-Field density models fit the signed residual between the reference and SAD densities. Training grids are generated from PySCF level-0 atom-centered integration grids. Each time a molecule is visited, 1,000 points are sampled uniformly without replacement from its complete grid. The point loss is $|w_g||\hat{\rho}_g - \rho_g|$, where w_g is the numerical-integration weight. The sampled loss of each molecule is multiplied by $N_{\text{grid}}/N_{\text{sample}}$ to estimate the unnormalized L_1 integral on the complete grid, and the batch loss is the molecular mean of this estimator. Validation and model selection use the complete level-0 grid of every molecule without sampling. After selecting a checkpoint, the same test set is also evaluated on the complete unpruned level-3 grid, which uses nonnegative quadrature weights for the QM9 elements. Integrated errors are accumulated in float64.

We optimize with AdamW at an initial learning rate of 10^{-3}. All other settings use the PyTorch defaults: $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-8}$, a weight decay of 10^{-2}, and no AMSGrad. The learning

Table A2: Training settings for the two principal GEM-FIELD models.

Training setting	QM9-PBE	SCFBench
Train/validation/test molecules	120,396/100/13,389	35,089/4,386/4,386
Effective batch size	64	64
Grid points per training molecule	1,000	1,000
Initial learning rate	10^{-3}	10^{-3}
Maximum updates	10^{6}	5×10^{5}
Validation interval	1,000	1,000
Early-stopping patience	100,000	100,000
Gradient clipping	1,000	1,000
Random seed	42	42

rate is warmed up linearly over the first 1,000 updates and then decayed as

$$\eta_t = \eta_0 \times \begin{cases} (t+1)/1000, & t < 1000, \\ 0.96^{(t-1000)/10000}, & t \geq 1000. \end{cases} \tag{A1}$$

The global gradient norm is clipped at 1,000. Training runs for at most 1,000,000 optimization steps, and validation is performed every 1,000 steps. Models are selected by the molecule-averaged validation L_1 error, with early stopping after 100,000 updates without improvement. The random seed is 42. We use neither mixed precision nor gradient accumulation for GEM-FIELD.

The QM9-PBE model uses a fixed random split. The SCFBench model reads the official main-data split directly and uses no OOD molecule for training or model selection.

A.3 TRAINING CONSISTENCY ACROSS ABLATIONS

The backbone and angular-momentum ablations share the data splits, SAD-residual target, L_1 loss, grid sampling, and optimization settings described above. EquiformerV2 models with $L_{\max} = 3, 4, 5, 6$ change only the maximum angular momentum of the backbone and density readout. To suppress high-frequency angular aliasing introduced by nonlinear decoding, models with $L_{\max} \leq 4$, $L_{\max} = 5$, and $L_{\max} = 6$ use Lebedev quadrature degrees 11, 13, and 15, respectively. The NequIP and eSCN comparisons fix $L_{\max} = 3$ and retain the same nonlinear density readout. The linear-decoder ablation uses the same $L_{\max} = 3$ NequIP backbone and removes only the $8 \to 16 \to 1$ pointwise nonlinear map. This design separately attributes the effects of the backbone, angular-momentum cutoff, and decoder.

B DETAILED BASELINE CONFIGURATIONS

B.1 DENSITY-PREDICTION BASELINES

ChargE3Net, SCDP, ELECTRA, and BOA share the QM9-PBE split, signed SAD-residual target, molecule-averaged integrated L_1 loss, and test molecules. They also share AdamW, an initial learning rate of 10^{-3}, the schedule in Appendix A, a gradient-clipping threshold of 1,000, validation every 1,000 steps, and a random seed of 42. Under a common single-GPU memory limit of 32 GB, the four models sample 50, 100, 1,000, and 1,000 level-0 grid points per molecule, respectively. Gradient accumulation is used where needed to maintain an effective batch size of 64. All frozen checkpoints are reevaluated on the same complete level-0 and unpruned level-3 test grids.

SCDP constructs a virtual node at each bond center and assigns it the oxygen node type. The orbital cutoff is 5.0 Å; exponents are not trained, and both target variance and output scaling are set to one. ELECTRA uses $4K \to 3K \to 2K$ decoders for weights and scalar multipliers, and $4K \to 3K \to 3K$ branches for displacement and covariance, with $K = 5$. BOA uses radial correction and a quadratic density readout; def2-TZVP is its internal representation basis, while the reference densities remain PBE/def2-SVP.

Table B1: Density-prediction baseline configurations.

Model	Principal architecture	Graph and geometry	Output representation	# Param.
ChargE3Net	3 interaction layers, multiplicity 500, $L_{\max} = 4$, 20 Gaussian radial functions	4.0 Å cutoff; at most 20 neighbors	Direct density readout at query points after atom encoding	1.91M
SCDP	8-layer eSCN, $L_{\max} = 4$, $m_{\max} = 2$, sphere/hidden/edge widths 32/64/32, 64 spherical samples	6.0 Å cutoff; at most 24 neighbors; bond-centered virtual nodes	Coefficients in an uncontracted def2-QZVPPD, even-tempered ($\beta = 2.0$) auxiliary basis	1.98M
ELECTRA	6-layer MiaoMiaoNet, width 36, maximum body order 3, maximum tensor way 3	20.0 Å cutoff; Aufbau atomic descriptors and principal inertia axes	Five movable anisotropic Gaussians per valence electron, with signed coefficients	1.68M
BOA	3 BOA blocks, hidden width 12, 4 orbitals	Edge/message/orbital cutoffs of 3.0/6.0/3.0 Å	Low-rank quadratic products of atom-centered GTOs in uncontracted def2-TZVP	1.58M

B.2 Direct-Property EquiformerV2 Models

We train an independent EquiformerV2 for each target without parameter sharing. Every model uses a 5.0 Å cutoff, at most 500 neighbors, $L_{\max} = m_{\max} = 4$, 96 spherical channels, 128 distance bases, 64 edge channels, and a spherical-grid resolution of 18. The attention hidden width is 48, with four attention heads, 64 alpha channels, 24 value channels, and a feed-forward width of 96.

The energy-component readout applies a $96 \to 96 \to 1$ SiLU MLP to invariant atomic features and sums the outputs over each molecule. The orbital and gap readouts first mean-pool the invariant atomic features and then apply the same MLP. For molecule b, the dipole model predicts an atomic partial charge $q_i = f_q(\mathbf{h}_i^{(0)})$ from each invariant feature and a local equivariant dipole $\boldsymbol{\mu}_i^{\mathrm{loc}} = f_\mu(\mathbf{h}_i^{(1)})$ from each $\ell = 1$ feature. Because QM9 molecules are neutral, the charges are centered within each molecule before aggregation:

$$\widetilde{q}_i = q_i - \frac{1}{N_b} \sum_{j \in b} q_j, \qquad \sum_{i \in b} \widetilde{q}_i = 0, \tag{B1}$$

$$\widehat{\boldsymbol{\mu}}_b = \sum_{i \in b} \left(\widetilde{q}_i \mathbf{R}_i + \boldsymbol{\mu}_i^{\mathrm{loc}} \right). \tag{B2}$$

Here N_b is the number of atoms in molecule b, and $\mathbf{R}_i$ is measured in Å. Charge centering makes the charge-position contribution invariant to a global translation of the molecule.

Energy-component models use six blocks; orbital, gap, and dipole models use five. Energy-component, orbital, and gap targets use MAE on normalized values. Dipoles are trained against the full vector without target normalization. For a batch of B molecules, the training loss and reported magnitude MAE are

$$\mathcal{L}_\mu = \frac{1}{B} \sum_{b=1}^{B} \left\| \widehat{\boldsymbol{\mu}}_b - \boldsymbol{\mu}_b \right\|_2, \qquad \mathrm{MAE}_{|\mu|} = \frac{1}{B} \sum_{b=1}^{B} \left| \left\| \widehat{\boldsymbol{\mu}}_b \right\|_2 - \left\| \boldsymbol{\mu}_b \right\|_2 \right|, \tag{B3}$$

both in D. Validation and checkpoint selection use per-atom energy MAE, orbital or gap MAE, or $\mathrm{MAE}_{|\mu|}$, as appropriate. All models use AdamW with the target-specific constant learning rates in Table B2, without warmup or learning-rate decay. Training uses batch size 32 without gradient accumulation and runs for at most 100 epochs. Early stopping occurs after 10 epochs without validation-MAE improvement, with a minimum improvement of zero. The split is 120,396/100/13,389 and the random seed is 42.

Energy components are expressed in eV and subtract elementwise atomic references fitted by least squares on the training set. Energy residuals, orbital energies, and gaps are standardized using their training-set statistics. Dipole vectors are expressed in D and are not normalized.

Table B2: Training configurations for direct-property EquiformerV2 models.

	Training configuration				
Target	Blocks	Attn. dropout	Constant LR	Weight decay	Atomic reference
E_{kin}	6	0	1.5×10^{-4}	0	yes
E_{ne}	6	0	1.5×10^{-4}	0	yes
E_J	6	0	1.5×10^{-4}	0	yes
E_{xc}	6	0	1.5×10^{-4}	0	yes
HOMO	5	0.2	5.0×10^{-4}	0.005	no
LUMO	5	0.2	5.0×10^{-4}	0.005	no
HOMO–LUMO gap	5	0.2	5.0×10^{-4}	0.005	no
$\boldsymbol{\mu}$	5	0.2	5.0×10^{-4}	0.005	no

Table B3: Architectures used for machine-learned interatomic-potential comparisons.

Model	Equivariant representation and backbone	Radial and message settings	Energy and force readout
EquiformerV2	6 blocks, $L_{\max} = m_{\max} = 4$, 96 spherical channels, 4 attention heads	128 Gaussian distance bases; edge widths 128/64/64; at most 500 neighbors	Equivariant feed-forward atomic-energy readout; terminal $\mathrm{SO}(2)$-equivariant attention predicts forces directly
NequIP ($L_{\max} = 1$)	4 message layers, 32 features, both natural and unnatural parity	8 fixed Bessel bases; sixth-order polynomial cutoff; 2×64 radial MLP	Fixed elementwise atomic-energy shifts; elementwise energy scales determined from training-force RMS
NequIP ($L_{\max} = 2$)	As above, with irreducible features through $\ell = 2$	Same as $L_{\max} = 1$ NequIP	Same as $L_{\max} = 1$ NequIP
MACE ($L_{\max} = 1$)	Hidden irreps $128 \times 0e + 128 \times 1o$; maximum correlation order 3	Maximum spherical-harmonic degree $\ell = 3$; 5.0 Å cutoff	Standard MACE energy and force readouts
MACE ($L_{\max} = 2$)	Hidden irreps $128{\times}0e{+}128{\times}1o{+}128{\times}2e$; maximum correlation order 3	Maximum spherical-harmonic degree $\ell = 3$; 5.0 Å cutoff	Standard MACE energy and force readouts

B.3 Machine-Learned Interatomic Potentials

All potential models use total-energy and atomic-force labels from the same QM9-PBE split. Coordinates are in angstrom, energies in eV, and forces in eV $\mathrm{Å}^{-1}$. Every model is trained for 30 epochs with a random seed of 42 and a nonperiodic cutoff of 5.0 Å. Before training, elementwise atomic-energy references fitted on the training set are subtracted from the total energy.

The EquiformerV2 energy target is normalized by the standard deviation of the training residuals. Its checkpoint is selected by validation force-component MAE. Test results for every model are generated from the best validation checkpoint.

C Wall-clock Analysis

C.1 End-to-End SCF Timing

End-to-end GPU4PySCF timings include model inference, initial density-matrix construction, and subsequent PBE/def2-SVP SCF (Wu et al., 2025). Table C1 reports per-molecule means for GEM-Field-seeded and default MINAO-initialized calculations.

Inference is negligible in both splits, while initial density-matrix construction is the dominant overhead. Including this cost, GEM-Field reduces total runtime by 18.69% on ID and 8.98% on

Table B4: Optimization settings for machine-learned interatomic potentials.

Model	Batch	Optimizer and initial LR	Schedule	Energy/force weights	EMA	Other settings
EquiformerV2	4	AdamW, 1.5×10^{-4}	5-epoch warmup, cosine decay	2/100	0.999	energy MAE; vector L_2-MAE force; clipping 100; no weight decay
NequIP ($L_{\max} = 1, 2$)	5	Adam, 10^{-2}	ReduceLR-OnPlateau, factor 0.6, patience 5, threshold 0.2, minimum LR 10^{-6}	1/1	0.999	per-atom energy loss; single precision
MACE ($L_{\max} = 1, 2$)	10	Adam with AMSGrad, 10^{-2}	stage two starts at epoch 14	1/100	0.99	RMS-force scaling; single precision

Table C1: End-to-end timing of GEM-FIELD-seeded and MINAO-initialized SCF calculations.

Split	Mean time (s)					Speedup ↑	Accel. (%) ↑
	Inference	DM build	Seeded SCF	GEM-FIELD total	MINAO SCF		
ID test	0.024	0.899	2.691	3.614	4.445	1.23×	99.43%
OOD test	0.025	3.161	6.724	9.910	10.888	1.10×	84.67%

OOD. It accelerates 99.43% and 84.67% of molecules, respectively. Density-matrix construction accounts for 31.90% of OOD GEM-FIELD time and is the main optimization target.

The following analysis replaces analytical three-center residual Coulomb integrals with a numerical grid Hartree potential. This implementation comparison tests whether reducing density-matrix construction cost can improve wall-clock performance without changing SCF convergence.

C.2 Implementation of Numerical Coulomb-Matrix Construction

The analytical path contracts three-center two-electron integrals between every normalized displaced s-type Gaussian and each AO pair. The numerical path does not construct these three-center integrals. For the kth normalized Gaussian, with charge q_k, width σ_k, and center $\mathbf{C}_k$, its Hartree potential is evaluated analytically at grid points as

$$v_{H,k}(\mathbf{r}) = q_k \frac{\mathrm{erf}(\|\mathbf{r} - \mathbf{C}_k\|/\sigma_k)}{\|\mathbf{r} - \mathbf{C}_k\|}. \tag{C1}$$

At $\mathbf{r} = \mathbf{C}_k$, we use the limiting value $2q_k/(\sqrt{\pi}\sigma_k)$. The total residual Hartree potential is the sum over all Gaussian components, and the numerical residual Coulomb matrix is

$$J_{\mu\nu}^{\mathrm{res,num}} \approx \sum_g w_g \phi_\mu(\mathbf{r}_g)\phi_\nu(\mathbf{r}_g)v_H^{\mathrm{res}}(\mathbf{r}_g). \tag{C2}$$

The implementation uses PySCF level-3 atom-centered grids. Residual densities, gradients, and Hartree potentials are computed in GPU chunks. The numerical Hartree and PBE exchange–correlation potentials are contracted in the same AO-grid chunks, avoiding a second traversal of AO values. The SAD J matrix is still evaluated analytically using the direct J engine in PySCF.

C.3 End-to-End Effect on the Complete SCFBench Evaluation

We evaluate the analytical and numerical Coulomb implementations on all 4,386 ID and 1,050 OOD molecules at PBE/def2-SVP. For each path, aggregate speedup is the total MINAO SCF time divided by the total GEM-FIELD time from the same run, reducing bias from system-load differences

Table C2: End-to-end analytical and numerical J results on SCFBench.

Convergence					
Split	J path	Molecules	Convergence (%)	Mean RIC (%)	Model cycles
ID	analytical	4,386	100	58.5119	34,194
ID	numerical	4,386	100	58.5071	34,190
OOD	analytical	1,050	100	59.4857	9,474
OOD	numerical	1,050	100	59.5188	9,479

End-to-end timing				
Split	J path	GEM-FIELD time (s)	Aggregate speedup	Molecules accelerated (%)
ID	analytical	15,850.86	$1.2299\times$	99.43
ID	numerical	14,873.88	$1.3220\times$	99.95
OOD	analytical	10,405.88	$1.0986\times$	84.67
OOD	numerical	9,540.77	$1.1667\times$	93.52

between runs. The GEM-FIELD total includes model inference, initial density-matrix construction, and the model-initialized SCF calculation.

Numerical J does not alter the main SCF-convergence conclusion. The mean RIC changes by -0.0048 percentage points for ID and $+0.0331$ percentage points for OOD, while both paths retain 100% convergence. Relative to the analytical path, the numerical path reduces the total GEM-FIELD runtime by 6.16% for ID and 8.31% for OOD. Aggregate speedup rises from $1.2299\times$ to $1.3220\times$ and from $1.0986\times$ to $1.1667\times$, respectively. For OOD molecules, the accelerated fraction increases from 84.67% to 93.52%, indicating a stronger runtime advantage for larger molecules.

The gain comes from merging computational paths rather than relaxing numerical accuracy or reducing the number of SCF iterations. The analytical method must generate an auxiliary integral environment for every displaced Gaussian center and contract three-center integrals, whereas the numerical method reuses the grid and AO chunks already required for exchange–correlation integration. Its limitation is that numerical-J accuracy depends on the integration grid and may be more sensitive to tighter Gaussian components. The level-3 grid is appropriate for the PBE/def2-SVP deployment studied here; changes to the basis, element range, or integration accuracy should be accompanied by a new assessment of matrix errors and SCF-convergence agreement relative to the analytical path.

D MATHEMATICAL CHARACTERIZATION OF GEM-FIELD EXPRESSIVITY

A linear decoder corresponds to a conventional finite-order radial spherical-harmonic expansion. With the nonlinear decoder, the continuous model admits an infinite-order radial spherical-harmonic expansion without a hard angular-degree cutoff; its numerically effective angular resolution is instead set by the spherical quadrature.

Consider one continuous Gaussian shell centered at atom $\mathbf{R}_i$. Let $\mathbf{x} = \mathbf{r} - \mathbf{R}_i$, $r = \|\mathbf{x}\|$, and $\widehat{\mathbf{x}} = \mathbf{x}/r$, and define

$$g_\sigma(\mathbf{y}) = \frac{1}{\pi^{3/2}\sigma^3} \exp\left(-\frac{\|\mathbf{y}\|^2}{\sigma^2}\right). \tag{D1}$$

A shell with charge Q, displacement radius d, and angular weight $p(\mathbf{u})$ has density

$$\rho(\mathbf{x}) = Q \int_{S^2} p(\mathbf{u}) g_\sigma(\mathbf{x} + d\mathbf{u})\, d\Omega_{\mathbf{u}}. \tag{D2}$$

Substituting $p(\mathbf{u}) = \sum_{\ell m} p_{\ell m} Y_{\ell m}(\mathbf{u})$ and using

$$e^{-\kappa \widehat{\mathbf{x}}\cdot\mathbf{u}} = 4\pi \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} (-1)^\ell i_\ell(\kappa) Y_{\ell m}(\widehat{\mathbf{x}}) Y_{\ell m}(\mathbf{u}), \qquad \kappa = \frac{2rd}{\sigma^2}, \tag{D3}$$

where i_ℓ is the modified spherical Bessel function, gives

$$\rho(\mathbf{x}) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} p_{\ell m} K_\ell(r; Q, d, \sigma) Y_{\ell m}(\widehat{\mathbf{x}}), \tag{D4}$$

with

$$K_\ell(r; Q, d, \sigma) = \frac{4\pi Q}{\pi^{3/2}\sigma^3} \exp\left(-\frac{r^2 + d^2}{\sigma^2}\right) (-1)^\ell i_\ell\left(\frac{2rd}{\sigma^2}\right). \tag{D5}$$

The Gaussian-shell integral is therefore diagonal in the spherical-harmonic basis: an angular coefficient $p_{\ell m}$ is mapped to a radial function in the same (ℓ, m) channel, without changing angular degree. With multiple shells, the radial coefficient in each channel is simply a linear combination of different K_ℓ functions.

Without the nonlinear decoder, band-limited backbone fields can be combined only linearly into

$$p_{\mathrm{lin}}(\mathbf{u}) = \frac{1}{4\pi} + \sum_{\ell=1}^{L_{\max}} \sum_{m=-\ell}^{\ell} a_{\ell m} Y_{\ell m}(\mathbf{u}). \tag{D6}$$

For orthonormal real spherical harmonics, $Y_{00} = 1/\sqrt{4\pi}$. When the shell index is restored, the uniform term therefore corresponds to

$$a_{k00} = \frac{1}{\sqrt{4\pi}}, \qquad a_{k00} Y_{00}(\mathbf{u}) = \frac{1}{4\pi}, \tag{D7}$$

for every shell k. Substitution into Equation D2 and summation over shells yields

$$\rho_{\mathrm{lin}}(\mathbf{x}) = \sum_{\ell=0}^{L_{\max}} \sum_{m=-\ell}^{\ell} F_{\ell m}(r) Y_{\ell m}(\widehat{\mathbf{x}}), \tag{D8}$$

$$F_{\ell m}(r) = \sum_k a_{k\ell m} K_\ell(r; Q_k, d_k, \sigma_k). \tag{D9}$$

This is precisely a conventional finite-order radial spherical-harmonic expansion. Additional shells enrich only the radial bases $F_{\ell m}(r)$; the angular degree remains strictly limited to $\ell \le L_{\max}$.

GEM-FIELD first constructs H band-limited spherical fields from equivariant features with $\ell \le L_{\max}$,

$$f_h(\mathbf{u}) = \sum_{\ell=1}^{L_{\max}} \sum_{m=-\ell}^{\ell} a_{\ell m}^{(h)} Y_{\ell m}(\mathbf{u}), \qquad h = 1, \ldots, H, \tag{D10}$$

and then applies a shared pointwise nonlinear network,

$$z(\mathbf{u}) = \mathbf{w}_2^{\mathsf{T}} \,\mathrm{SiLU}(\mathbf{W}_1 \mathbf{f}(\mathbf{u}) + \mathbf{b}_1) + b_2, \tag{D11}$$

$$p_{\mathrm{NL}}(\mathbf{u}) = \frac{1}{4\pi} + z(\mathbf{u}) - \frac{1}{4\pi} \int_{S^2} z(\mathbf{v})\, d\Omega_{\mathbf{v}}. \tag{D12}$$

SiLU satisfies $\mathrm{SiLU}(t) - \mathrm{SiLU}(-t) = t$, so the decoder contains a linear decoder as a special case. More importantly, SiLU is nonpolynomial. For an admissible input $f_1(\mathbf{u}) \propto Y_{10}(\mathbf{u}) \propto \cos\theta$, the network can produce

$$z(\theta) = A\,\mathrm{SiLU}(B + C\cos\theta) + D, \qquad AC \neq 0. \tag{D13}$$

If z contained only finitely many spherical-harmonic degrees, axial symmetry would require it to be a finite Legendre polynomial in $\cos\theta$. The expression above is nonpolynomial in $\cos\theta$, so its Legendre expansion necessarily extends to arbitrarily high degree. More generally, pointwise nonlinearities map products and higher-order compositions of finite-degree spherical fields into higher angular-momentum channels, giving

$$z(\mathbf{u}) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} b_{\ell m} Y_{\ell m}(\mathbf{u}). \tag{D14}$$

Mean correction fixes only the $\ell = 0$ component and leaves all coefficients with $\ell \geq 1$ unchanged. Substituting this expansion into the Gaussian-shell integral gives the multi-shell density

$$\rho_{\mathrm{NL}}(\mathbf{x}) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \left[\sum_{k} b_{k\ell m} K_{\ell}(r; Q_k, d_k, \sigma_k) \right] Y_{\ell m}(\widehat{\mathbf{x}}). \tag{D15}$$

Thus, the nonlinear decoder transforms finite-degree equivariant features into an infinite-order angular expansion, while Gaussian shells provide a learnable radial response for every populated angular-momentum channel. In its continuous definition, GEM-FIELD is consequently equivalent to a structured infinite-order radial spherical-harmonic expansion and does not inherit a hard angular cutoff from the backbone $L_{\max}$.

The implementation uses $U = 86$ Lebedev directions,

$$\rho^{(U)}(\mathbf{x}) = Q \sum_{j=1}^{U} \Omega_j p(\mathbf{u}_j) g_{\sigma}(\mathbf{x} + d\mathbf{u}_j), \qquad U = 86. \tag{D16}$$

Although the continuous model has no hard spherical-harmonic cutoff, a finite quadrature limits the reliably resolved angular frequencies. The practical angular resolution is therefore set by the spherical integration grid rather than by $L_{\max}$ alone.

E Angular-Momentum Degree and Mode Utilization of the Nonlinear Decoder

To test the expressivity result in Appendix D, we analyze how the nonlinear decoder changes the empirical angular spectrum. We compare two independently trained NequIP models with identical data, splits, backbone size, and $\ell_{\max} = 3$; they differ only in whether nonlinear angular decoding is used. The analysis considers the 16 active oxygen shells of a water molecule. Each continuous angular function is evaluated at the 5,810 directions of a degree-131 Lebedev grid and projected onto the standard real spherical-harmonic basis through $\ell = 20$. All mode statistics exclude the isotropic $\ell = 0$ component. The maximum relative L_2 projection error is 2.73×10^{-6}, indicating that the projection range resolves the dominant angular structure produced by the decoder.

We use three complementary spectral measures. N_{99} is the minimum number of real spherical-harmonic modes needed to explain 99% of the anisotropic power and quantifies how many independent modes participate in the angular representation. ℓ_{99} is the highest angular degree required to reach 99% cumulative power. A specific (ℓ, m) mode is considered significant if it contributes at least 1% of the anisotropic power.

The linear decoder concentrates power strictly at $\ell \leq 3$. Averaged over all shells, the power fraction at $\ell > 3$ is only 1.16×10^{-15}, consistent with numerical rounding (Figure E1a). With nonlinear decoding, 13 shells retain more than 1% of their total anisotropic power at $\ell > 3$. The mean, median, and maximum high-degree power fractions are 7.29%, 3.98%, and 22.42%, respectively (Figure E1b). Twelve shells contain at least one $\ell > 3$ mode contributing more than 1%. The mean ℓ_{99} rises from 2.88 to 4.19, and its maximum rises from 3 to 6 (Figure E1d). At the 10^{-6} display threshold, the nonlinear spectrum extends to $\ell = 12$, whereas the linear spectrum terminates at $\ell = 3$.

The nonlinear decoder also increases angular-mode utilization. The linear model requires an average of 4.00 modes to explain 99% of the anisotropic power, compared with 6.81 modes for the nonlinear model, an increase of 70.3%. The maximum N_{99} correspondingly rises from 5 to 12 (Figure E1c). At the per-mode 1% threshold, the mean number of significant modes increases from 4.00 to 5.50, or 37.5%. The added expressivity therefore does not arise from a single high-degree mode; it is distributed across more (ℓ, m) channels.

The complete shell projections provide a spatial view of this spectral expansion (Figure E2). Angular functions from the linear decoder retain the smooth forms permitted by $\ell_{\max} \leq 3$. The nonlinear decoder produces significant modes with $\ell_{\max} > 3$ in 12 shells. Shell 8 reaches $\ell_{\max} = 6$, and shells 3 and 9 reach $\ell_{\max} = 5$. These higher-order components create more localized positive and negative regions and richer nodal structure, showing that the additional modes are actively used to construct the shell angular distributions.

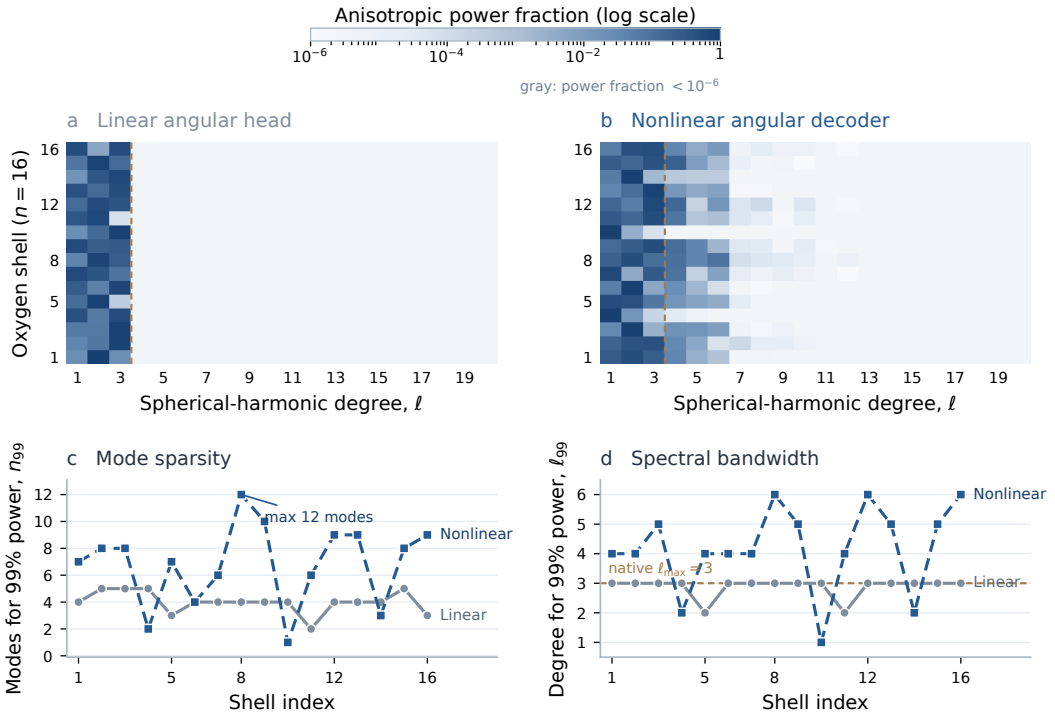


Figure E1: **The nonlinear decoder increases angular-momentum degree and mode utilization.** (a,b) Fraction of anisotropic power at each spherical-harmonic degree for the 16 oxygen shells produced by the linear angular head and nonlinear angular decoder. The heat maps use a shared low-saturation blue logarithmic scale from 10^{-6} to 1; gray denotes power below 10^{-6}. (c) Minimum number of real spherical-harmonic modes N_{99} required by each shell to reach 99% anisotropic power. (d) Angular degree ℓ_{99} required to reach 99% power. Solid gray-blue lines with circles and dashed dark-blue lines with squares denote the linear and nonlinear decoders, respectively. The golden-brown dashed line marks the common native backbone limit $\ell_{\max} = 3$.

These results directly support the increase in implicit angular degree produced by the nonlinear decoder. Although the explicit backbone features remain limited to $\ell_{\max} = 3$, the decoded angular functions use dominant components at $\ell = 4$–6 and retain resolvable spectral tails at still higher degree. The nonlinear decoder therefore raises both effective angular degree and mode utilization, allowing a fixed low-degree backbone to form a richer angular density representation.

F Numerical Equivariance and Derived-Property Stability

F.1 Numerical Equivariance Test

We freeze a trained EquiformerV2 density model and vary only the Lebedev degree at inference. The model uses an $\ell_{\max} = 6$ backbone and a nonlinear angular decoder. It was trained with degree-15 Lebedev quadrature. The test molecules are H_2O, CH_4, CH_2O, CH_3OH, and CH_3F. The tests span 16 Lebedev rules from degree 3 to 131. Degrees 15, 17, and 131 contain 86, 110, and 5,810 directions, respectively.

Each molecule is transformed by 20 Haar-random rotations and three fixed right-angle rotations. This gives 115 tests for each evaluation mode and quadrature degree. The primary metric is the weighted global relative density L_1 error at corresponding physical positions before and after rotation. We average this metric over all molecules and rotations. Mean errors in the scalar, first moment, and second moment serve as auxiliary checks. A degree passes persistently when it and all higher tested degrees remain below a specified threshold.

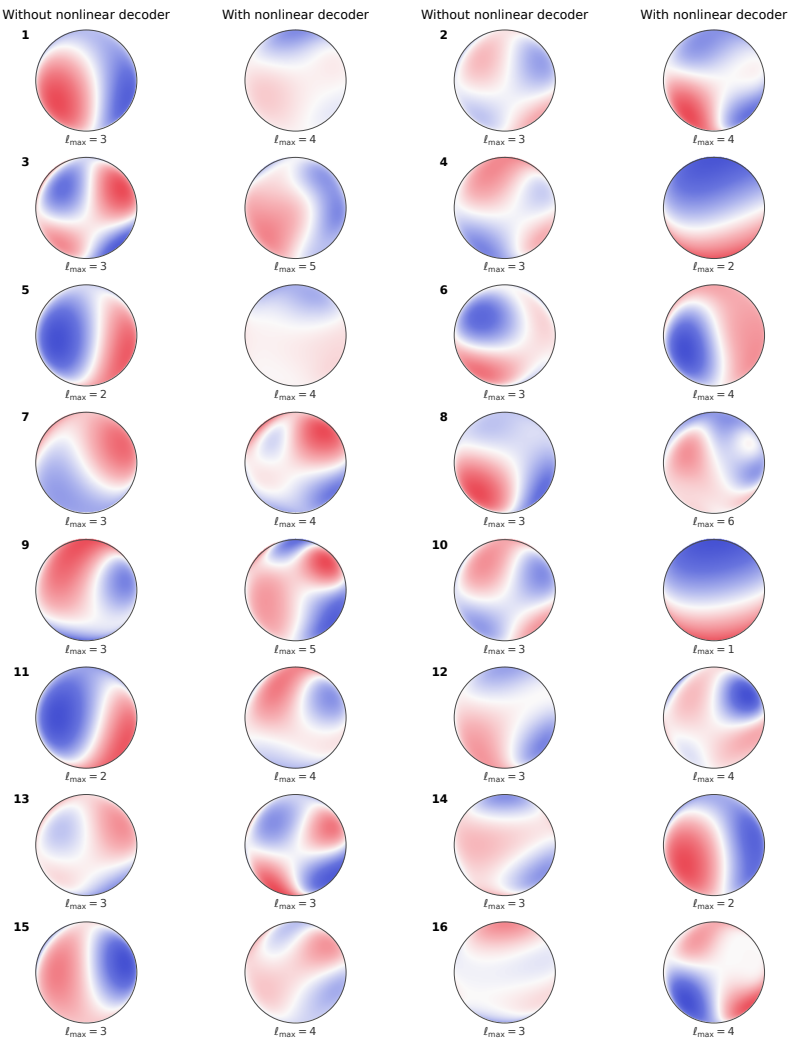


Figure E2: **The nonlinear decoder increases the spatial complexity of oxygen-shell angular distributions.** The 16 oxygen shells are shown in sequence. For each shell, the linear decoder is on the left and the nonlinear decoder on the right. Disks show the orthographic projection of $p(\mathbf{u}) - 1/(4\pi)$; red and blue denote positive and negative angular deviations. Each shell pair shares a zero-centered color scale. The $\ell_{\max}$ below each disk is the highest spherical-harmonic degree containing at least one mode that contributes no less than 1% of the anisotropic power.

Figure F1 reports decoder-only and end-to-end equivariance errors. At degree 13, their mean relative L_1 errors are 3.4685×10^{-3} and 3.4817×10^{-3}, respectively. Neither meets the 10^{-3} threshold. At degree 15, the errors fall to 4.5367×10^{-4} and 4.7606×10^{-4}. Every higher tested degree remains below the threshold. Degree 15 is therefore the minimum persistent passing degree for both modes.

The minimum passing degrees are 13, 15, and 19 at thresholds of 10^{-2}, 10^{-3}, and 10^{-4}, respectively. The high-degree end-to-end plateau reflects backbone reevaluation and floating-point error rather than Lebedev quadrature.

F.2 Derived-Property Test

For the same five molecules, we compute total energy, HOMO–LUMO gap, and dipole moment using one-shot GPU4PySCF PBE/def2-SVP calculations. Table F1 and Figure F2 report mean absolute

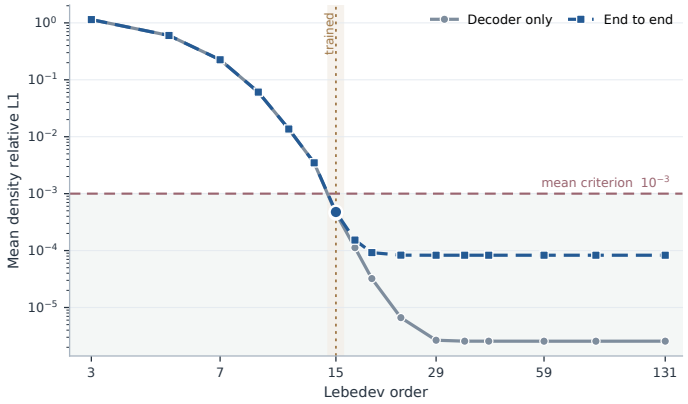


Figure F1: **Degree-15 Lebedev quadrature satisfies the target numerical $\mathrm{SO}(3)$ equivariance threshold.** Solid gray-blue lines with circles show decoder-only reevaluation. Dashed dark-blue lines with squares show complete-model reevaluation after molecular rotation. The vertical axis is the mean relative density L_1 error over all molecules and rotations.

differences relative to the degree-59 reference. All three properties stabilize rapidly. At degree 15, the mean energy, gap, and dipole differences are 8.63×10^{-8} Ha, 9.44×10^{-5} eV, and 2.11×10^{-4} D, respectively. At degree 19, they decrease to 1.83×10^{-8} Ha, 1.89×10^{-5} eV, and 3.55×10^{-5} D.

Table F1: Stability of derived electronic properties across Lebedev degrees. Values are mean absolute differences over five test molecules relative to degree 59.

Degree	Mean energy difference (Ha)	Mean gap difference (eV)	Mean dipole difference (D)
3	1.4858×10^{-1}	1.9581	1.5310
7	5.5176×10^{-4}	6.0595×10^{-2}	4.7128×10^{-2}
11	4.2743×10^{-7}	1.0144×10^{-3}	2.9645×10^{-3}
15	8.6317×10^{-8}	9.4429×10^{-5}	2.1053×10^{-4}
19	1.8281×10^{-8}	1.8934×10^{-5}	3.5506×10^{-5}
29	3.4361×10^{-10}	3.3444×10^{-7}	8.8360×10^{-7}
41	9.5834×10^{-11}	2.5564×10^{-7}	1.5887×10^{-7}
59	0	0	0

From degree 11 through 59, the mean energy error relative to PBE remains approximately 1.57×10^{-5} Ha. Mean gap errors remain between 2.54×10^{-2} and 2.62×10^{-2} eV. Mean dipole errors remain between 3.26×10^{-2} and 3.37×10^{-2} D. The residual property errors beyond degree 15 therefore do not arise from Lebedev discretization. Together, both tests support the degree-15, 86-direction rule used for training and inference.

G Analytic Evaluation of Density-Derived Properties

G.1 Energy and Orbital Properties from the One-Shot Density Matrix

The one-shot generalized eigensolution provides the occupied-orbital density matrix and its energy-weighted counterpart,

$$D^{(0)}_{\mu\nu} = 2 \sum_{i \in \mathrm{occ}} C_{\mu i} C_{\nu i}, \qquad W^{(0)}_{\mu\nu} = 2 \sum_{i \in \mathrm{occ}} \varepsilon_i C_{\mu i} C_{\nu i}. \tag{G1}$$

For the closed-shell PBE calculations, the total energy evaluated from $\mathbf{D}^{(0)}$ is

$$E[\mathbf{D}^{(0)}] = E_{\mathrm{nn}} + \mathrm{Tr}[\mathbf{D}^{(0)}\mathbf{h}] + E_J^{\mathrm{DF}}[\mathbf{D}^{(0)}] + E_{\mathrm{xc}}[\rho_{\mathbf{D}^{(0)}}], \qquad \rho_{\mathbf{D}^{(0)}}(\mathbf{r}) = \sum_{\mu\nu} D^{(0)}_{\mu\nu} \chi_\mu(\mathbf{r}) \chi_\nu(\mathbf{r}), \tag{G2}$$

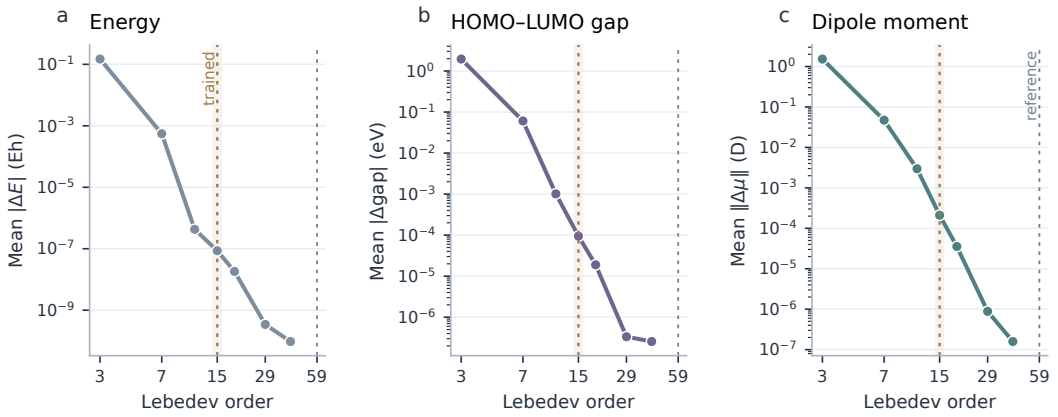


Figure F2: **Degree-15 Lebedev quadrature gives stable derived electronic properties.** (a–c) Mean absolute differences in total energy, HOMO–LUMO gap, and dipole moment relative to degree 59. All vertical axes are logarithmic. Golden-brown dashed lines and light bands mark the degree-15 training rule. Gray dashed lines mark the degree-59 reference. Reference zeros are omitted from the logarithmic axes.

where $\mathbf{h} = \mathbf{T} + \mathbf{V}_{\mathrm{ne}}$ and E_{nn} is the nuclear-repulsion energy. This expression yields the reported kinetic, electron–nuclear, Coulomb, exchange–correlation, and total energies without another orbital diagonalization. The HOMO and LUMO energies, and hence their gap, are selected directly from $\boldsymbol{\varepsilon}$.

We evaluate the Coulomb energy by density fitting in an auxiliary basis $\{\beta_P\}$. Define the three-center integrals, Coulomb metric, density projection, and fitted coefficients as

$$B_{P,\mu\nu} = (P \mid \mu\nu), \qquad V_{PQ} = (P \mid Q), \qquad b_P = \sum_{\mu\nu} D^{(0)}_{\mu\nu} B_{P,\mu\nu}, \qquad \mathbf{c} = \mathbf{V}^{-1}\mathbf{b}. \tag{G3}$$

The density-fitted Coulomb contribution is then

$$E_J^{\mathrm{DF}} = \frac{1}{2}\mathbf{b}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{b} = \frac{1}{2}\mathbf{b}^{\mathsf{T}}\mathbf{c}. \tag{G4}$$

G.2 First-Diagonalization Force Approximation with Density Fitting

The reported force is obtained by evaluating the stationary restricted-PBE nuclear-gradient expression once with the first-diagonalization orbitals, eigenvalues, occupations, and density matrix, without subsequent SCF iterations. Because the Fock matrix is constructed from the predicted density before diagonalization, the resulting $\mathbf{D}^{(0)}$ is generally not a self-consistent stationary point of $E[\mathbf{D}]$. Equation G5 defines the force approximation evaluated at $\mathbf{D}^{(0)}$. For nuclear coordinate $R_{A\alpha}$, we denote an explicit integral derivative by a superscript $(A\alpha)$. The force component is

$$F_{A\alpha} = -\left[E_{\mathrm{nn}}^{(A\alpha)} + \sum_{\mu\nu} D^{(0)}_{\mu\nu} h^{(A\alpha)}_{\mu\nu} - \sum_{\mu\nu} W^{(0)}_{\mu\nu} S^{(A\alpha)}_{\mu\nu} + \left(E_J^{\mathrm{DF}}\right)^{(A\alpha)} + E_{\mathrm{xc}}^{(A\alpha)}\right]. \tag{G5}$$

The overlap term is the Pulay contribution from the geometry-dependent AO basis. The one-electron derivative includes both AO-basis response and the derivative of the electron–nuclear operator.

Density fitting is differentiated explicitly rather than treated as geometry independent. At fixed $\mathbf{D}^{(0)}$, the Coulomb derivative is

$$\begin{aligned} \left(E_J^{\mathrm{DF}}\right)^{(A\alpha)} &= \mathbf{c}^{\mathsf{T}}\mathbf{b}^{(A\alpha)} - \frac{1}{2}\mathbf{c}^{\mathsf{T}}\mathbf{V}^{(A\alpha)}\mathbf{c}, \\ b_P^{(A\alpha)} &= \sum_{\mu\nu} D^{(0)}_{\mu\nu} B^{(A\alpha)}_{P,\mu\nu}. \end{aligned} \tag{G6}$$

Thus, the force includes derivatives of both the three-center integrals and the auxiliary-basis Coulomb metric. These terms contain the responses of the AO and auxiliary basis functions centered on the moving nuclei.

Table G1: Agreement between the first-diagonalization force and the central finite-difference derivative of the one-shot energy. MAE is over all Cartesian components in Ha/Bohr; "cos" is the cosine similarity between the complete force vectors.

	$h = 10^{-3}$ Bohr		$h = 10^{-4}$ Bohr		$h = 10^{-5}$ Bohr	
Molecule	MAE	cos	MAE	cos	MAE	cos
H_2O	1.293×10^{-4}	0.999611	1.421×10^{-4}	0.999530	3.662×10^{-4}	0.996485
C_2H_5OH	2.381×10^{-4}	0.999254	2.776×10^{-4}	0.998917	1.229×10^{-3}	0.980209
CH_3CONH_2	1.591×10^{-4}	0.998956	2.397×10^{-4}	0.998127	2.044×10^{-3}	0.875431
C_6H_6	5.770×10^{-5}	0.999933	2.620×10^{-4}	0.996676	2.543×10^{-3}	0.831573

For PBE, write the exchange–correlation energy density as $f_{\mathrm{xc}}(\rho, \gamma)$, where $\gamma = |\nabla\rho|^2$. On the atom-centered numerical grid, its explicit derivative is

$$E_{\mathrm{xc}}^{(A\alpha)} = \sum_g \left[w_g^{(A\alpha)} f_{\mathrm{xc},g} + w_g \left(v_{\rho,g} \rho_g^{(A\alpha)} + v_{\gamma,g} \gamma_g^{(A\alpha)} \right) \right], \tag{G7}$$

where $v_\rho = \partial f_{\mathrm{xc}}/\partial\rho$ and $v_\gamma = \partial f_{\mathrm{xc}}/\partial\gamma$. We enable the full grid response, including derivatives of quadrature weights, AO values, and density gradients on the moving atom-centered grid.

G.3 Finite-Difference Check of the Force Approximation

For the geometry-dependent one-shot energy $\widetilde{E}(\mathbf{R}) = E[\mathbf{D}^{(0)}(\mathbf{R}); \mathbf{R}]$, the total derivative follows the chain

$$\mathbf{R} \longmapsto \hat{\rho}(\mathbf{R}) \longmapsto \mathbf{F}[\hat{\rho}(\mathbf{R})] \longmapsto \big(\boldsymbol{\varepsilon}(\mathbf{R}), \mathbf{C}(\mathbf{R})\big) \longmapsto \mathbf{D}^{(0)}(\mathbf{R}) \longmapsto \widetilde{E}(\mathbf{R}). \tag{G8}$$

Equation G5 contains the integral, AO-basis, density-fitting, and exchange–correlation grid derivatives in the stationary gradient kernel. The finite difference of $\widetilde{E}$ also includes the nonstationary electronic response propagated through Equation G8. We measure this contribution using the central finite-difference force

$$F_{A\alpha}^{\mathrm{FD}}(h) = -\frac{\widetilde{E}(\mathbf{R} + h\mathbf{e}_{A\alpha}) - \widetilde{E}(\mathbf{R} - h\mathbf{e}_{A\alpha})}{2h}. \tag{G9}$$

At every displaced geometry, we rerun density inference, rebuild $\mathbf{D}^{(0)}$, and evaluate the same density-fitted PBE/def2-SVP energy with GPU4PySCF. The test geometries of water, ethanol, acetamide, and benzene are generated using RDKit ETKDGv3, followed by addition of explicit hydrogens and MMFF optimization for at most 500 iterations.

The 10^{-3}-Bohr step gives the stable comparison: the component MAE is 5.77×10^{-5}–2.38×10^{-4} Ha/Bohr (2.97–12.25 meV/Å), the relative MAE is 1.45–3.48%, and every force-vector cosine exceeds 0.9989. The 10^{-4}-Bohr results remain similar, whereas 10^{-5} Bohr is degraded by subtractive cancellation and numerical noise from single-precision model inference and rebuilt atom-centered grids. At a self-consistent stationary solution, the contraction of the energy gradient with the orbital response vanishes, and Equation G5 becomes the negative total energy derivative. The observed local agreement shows that the nonstationary response contribution is small for the four geometries tested here.

G.4 Dipole Moment from the Predicted Continuous Density

The dipole is evaluated from the predicted density $\hat{\rho} = \rho_{\mathrm{SAD}} + \widehat{\Delta\rho}$ rather than reconstructed from $\mathbf{D}^{(0)}$. With the coordinate origin used for the molecular calculation,

$$\boldsymbol{\mu}[\hat{\rho}] = \sum_A Z_A \mathbf{R}_A - \int_{\mathbb{R}^3} \mathbf{r}\, \hat{\rho}(\mathbf{r})\, d\mathbf{r}. \tag{G10}$$

Define $\mathbf{r}_{\mu\nu} = \int \chi_\mu(\mathbf{r}) \mathbf{r} \chi_\nu(\mathbf{r}) d\mathbf{r}$ and $q_{ikj} = \omega_j Q_{ik} p_{ik}(\mathbf{u}_j)$. Each normalized residual Gaussian has center $\mathbf{C}_{ikj} = \mathbf{R}_i - d_{ik}\mathbf{u}_j$ and first moment $q_{ikj}\mathbf{C}_{ikj}$. The analytic dipole used in evaluation is therefore

$$\boldsymbol{\mu}[\hat{\rho}] = \sum_A Z_A \mathbf{R}_A - \sum_{\mu\nu} D_{\mu\nu}^{\mathrm{SAD}} \mathbf{r}_{\mu\nu} - \sum_{i,k,j} q_{ikj} \mathbf{C}_{ikj}. \tag{G11}$$

The Gaussian widths do not enter this first moment because every component is normalized and centered at $\mathbf{C}_{ikj}$.