ORBITAL SPECIFIC EXCHANGE: ADVANCING HYBRID DENSITY FUNCTIONAL THEORY WITH GEOMETRIC DEEP LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

One of the main challenges of hybrid Density Functional Theory (DFT) is selection of the exact-exchange fraction (α). While state-of-the-art methods adapt α per molecule, they remain limited to a global scalar, ignoring that distinct electronic environments require distinct physical treatments, i.e., the optimal α would be different for various parts of the same molecule. We propose a generalized formalism that treats exchange mixing not as a scalar hyperparameter, but as a learnable orbital-specific interaction map. We illustrate it with OSE26—an equivariant neural network predicting a dense mixing matrix that modulates the Hartree–Fock exchange operator. This exchange-functional implementation provides self-consistency and roto-translational energy invariance. Trained and validated on MSR-ACC-TAE and RDB7, OSE26 achieves near-chemical accuracy for reaction-barrier prediction, attaining a $\sim$1 kcal/mol MAE on RDB7 and reducing MAE by 11–56% relative to the considered baselines. This move from system-wide scalars to learnable orbital-level matrices promises a more broadly applicable hybrid-functional paradigm.

1 INTRODUCTION

Density Functional Theory (DFT) is a widely used computational approach in chemistry and physics for predicting electronic and atomic structure in many-body systems, largely because it offers a favorable accuracy–cost trade-off. More accurate methods, such as coupled cluster Čížek (1966) and Quantum Monte Carlo McMillan (1965); Metropolis et al. (1953), can achieve higher precision but at substantially greater computational cost and with poor scaling to large systems. The key source of approximation in DFT is the exchange–correlation (XC) energy, which accounts for many-electron interaction effects; since its exact form is unknown, constructing accurate XC approximations is crucial for robust DFT predictions. Many different levels of such density functional approximations (DFAs) exist. Traditional XC functionals are built from analytical forms depending on the electron density (LDA Vosko et al. (1980); Perdew & Zunger (1981); Perdew & Wang (1992)), the density and its gradient magnitude $|\nabla\rho(\mathbf{r})|$ (GGA Perdew & Yue (1986); Perdew et al. (1992; 1996)), or these plus additional local ingredients (meta-GGA Perdew et al. (1999); Tao et al. (2003)), typically with multiple parameters fitted to experiments or higher-level calculations. Hybrid DFAs Ernzerhof & Perdew (1998) further improve accuracy by mixing a fraction of exact (Hartree–Fock) exchange with semilocal exchange–correlation.

Hybrid DFAs are common workhorses of molecular quantum chemistry: they routinely deliver robust geometries, reaction barriers, and electronic structure descriptors at tractable cost, and are widely deployed in high-throughput screening pipelines and benchmark datasets Adamo & Barone (1999); Mardirossian & Head-Gordon (2017). However, global hybrid DFAs inherit a structural bottleneck: they require choosing a single global mixing fraction α of exact (Hartree–Fock) exchange, which implicitly enforces a *single* system-wide balance between the exact-exchange contribution and the semilocal exchange–correlation description for all electronic environments in a system. This limitation controls the applicability domain of hybrid across thermochemistry, kinetics, and noncovalent interactions.

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

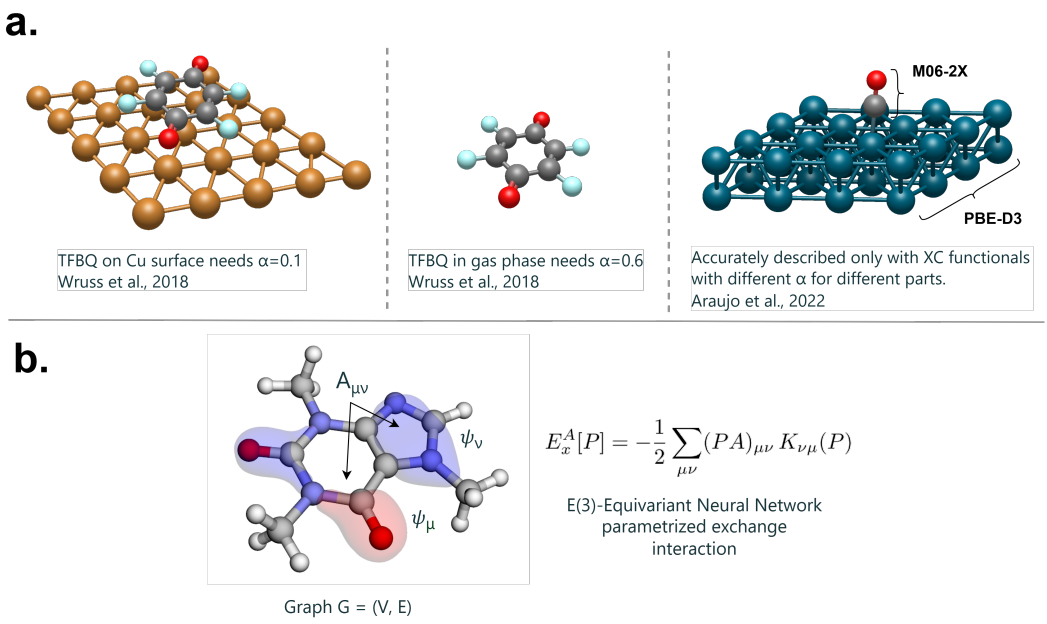


Figure 1: (a). Examples of systems requiring different exact-exchange fractions for the same molecule. Left: TFBQ adsorbed on $Cu_2O(111)$ surface requires $\alpha \approx 0.10$. Center: Gas-phase TFBQ requires $\alpha \approx 0.62$ for correct electron affinity. Right: CO adsorption on Pd slab, where the metal center and adsorbate have conflicting α requirements. (b) Proposed method: neural network parametrized exact exchange mixing adapted to local chemical environments.

This limitation becomes acute in real chemical systems (Fig. 1 a). The tetrafluorobenzoquinone (TFBQ) electron acceptor illustrates this fundamental limitation: gas-phase calculations require $\alpha \approx 0.62$ for correct electron affinity, surface adsorption on $Cu_2O(111)$ requires $\alpha \approx 0.10$ Wruss et al. (2018), a sixfold variation in optimal α for the same molecule depending on the system. On catalytic oxide surfaces such as CO oxidation and TiO_2, the metal center requires higher α (0.30–0.40) for accurate d-electron self-interaction correction, while adsorbate species over-localize at the same α Zhao & Kulik (2019). Practically, this incompatibility has led researchers to use different functionals for different parts of the same system—for example, employing PBE-D3 for periodic metal surfaces while applying hybrid functionals like M06-2X for adsorbates Araujo et al. (2022) — achieving mean absolute errors of 2.4 kcal/mol for adsorption energies. These examples demonstrate that no single global α can satisfy the conflicting requirements of different electronic structures.

Adaptive hybrid strategies aim to predict an optimal α per system (e.g., per molecule), improving accuracy relative to fixed-α baselines Khan et al. (2025). Yet these methods remain fundamentally limited: a molecule is not electronically homogeneous. Different regions—covalent bonds, charge-transfer interfaces, conjugated π-systems, transition-metal centers—exhibit distinct self-interaction and delocalization characteristics that are not well captured by a single scalar. Modern range-separated and local hybrids already introduce exchange flexibility beyond a fixed global mixing fraction: range-separated hybrids vary the exchange treatment with interelectronic distance, whereas local hybrids use a position-dependent mixing factor $a(\mathbf{r})$ Jaramillo et al. (2003). OSE addresses a complementary axis by assigning geometry-conditioned weights to AO-pair exchange channels. These approaches are distinct and potentially complementary. In this work step forward is made with adaptation on the *orbital-pair level*, replacing usual scalar α of hybrid density functionals with a learnable, geometry dependent, AO$\times$AO mixing map.

OSE remains within the hybrid-functional family Perdew & Schmidt (2001); Zhang & Ren (2023) but introduce orbital-pair dependence into the exchange treatment. Since the Ladder is fundamentally defined by the ingredients used to construct the energy density, explicit inclusion of geometric semantics and orbital Hilbert space (orbital-orbital mixing is geometry dependent) constitutes a dis-

tinct parameterization within the hybrid-functional paradigm. While double-hybrids (fifth rung) add nonlocal MP2-like correlation to capture virtual-orbital physics Su et al. (2018), we modify the structure of exchange functionals themselves to capture each orbital-orbital exchange interactions, depending on their specific chemical environment. This formulation offers a promising avenue for rectifying the intrinsic pathologies of hybrid functionals, achieving accuracy without ascending from the occupied to the virtual orbital space (thereby avoiding the steep $\mathcal{O}(N^5)$ scaling penalty).

Our main **contributions** are as follows:

1. We present new approach, named OSE26 (Orbital Specific Exchange), generalizing global hybrid functionals by replacing the scalar exact-exchange fraction α with a learnable AO$\times$AO mixing matrix $A_{\mu\nu}$, that rescales each orbital-pair contribution to Hartree–Fock exchange.
2. We implement this formalism not only post-SCF (typical for higher rung XC functionals) but also in efficient self-consistent functional with corresponding analytical nuclear forces.
3. We train it on diverse chemical reaction energies, and obtain near-chemical-accuracy on both reaction barriers and atomization energies at hybrid-like cost, improving over established hybrid and double-hybrid benchmarks on RDB7 and MSR-ACC-TAE25.

2 Background

2.1 Kohn–Sham DFT and hybrid functionals

Kohn–Sham DFT Hohenberg & Kohn (1964); Kohn & Sham (1965) replaces the interacting electronic problem by non-interacting electrons moving in an effective potential, determined self-consistently. In practice, the XC functional must be approximated. Semilocal DFAs (LDA/GGA/meta-GGA) provide computational efficiency and broad transferability Perdew et al. (1996); Sun et al. (2015), while hybrids blend semilocal exchange with a fraction of exact exchange improving reaction energies, frontier gaps and barriers.

A typical global hybrid has exchange energy

$$E_x^{\text{hyb}}[P] = \alpha\, E_x^{\text{HF}}[P] + (1-\alpha)\, E_x^{\text{DFA}}[P], \tag{1}$$

with fixed α (e.g., $\alpha = 0.25$ in PBE0 Adamo & Barone (1999)). Despite their widespread adoption, global hybrids are known to be optimal only for a fixed range of chemical properties. This is evidenced by the proliferation of specialized functionals—a "zoo" of hybrids with different α values optimized for narrow domains (thermochemistry, kinetics, noncovalent interactions) Goerigk et al. (2017); Schreiner et al. (2006); Chéron et al. (2012). In practice, researchers often manually calibrate α for specific problem classes Demchenko & Reshchikov (2020), but this task-specific tuning is typically omitted in large-scale high-throughput quantum chemistry databases Jain et al. (2013), leading to systematic biases that propagate into downstream applications.

2.2 ML for density functionals

Machine learning has been used to learn XC energies/potentials from density-based descriptors Zheng et al. (2004); Ryabov et al. (2020; 2022); Nagai et al. (2020), and to enforce constraints such as fractional-charge behavior Kirkpatrick et al. (2021). An NN-based XC functional typically maps electron-density descriptors to the local XC energy density. The molecular XC energy then is computed by integrating this energy density over the grid and combining it with the Coulomb and kinetic contributions to form the total energy. The network parameters are trained by minimizing the discrepancy between predicted and reference total energies. In many applications, one uses energy differences between related systems as training targets, since they carry direct physical and chemical relevance. Recent work has explored scalar-adaptive hybrids that predict system-level α Khan et al. (2025), achieving improved accuracy over fixed-α baselines. Our approach is orthogonal: we keep the semilocal XC integration unchanged and instead learn a structured modulation of the exact exchange operator at the orbital-pair level.

2.3 ENERGY EXPRESSION: FROM SCALAR α TO MATRIX A

We extend the hybrid Kohn–Sham construction by introducing a symmetric modulation matrix $A \in \mathbb{R}^{n\times n}$ acting in AO space. For spin-unrestricted calculations, let P^σ denote the AO density matrix of spin channel $\sigma \in \{\uparrow, \downarrow\}$, and let $K(P^\sigma)$ be the corresponding Hartree–Fock exchange matrix. The same geometry-dependent mixing matrix A is applied to both spin channels. We define the learned exchange energy as

$$E_x^A[\{P^\sigma\}] = -\frac{1}{2}\sum_\sigma \mathrm{Tr}[P^\sigma A K(P^\sigma)] = -\frac{1}{2}\sum_\sigma \sum_{\mu\nu} (P^\sigma A)_{\mu\nu} K_{\nu\mu}(P^\sigma), \tag{2}$$

where

$$(P^\sigma A)_{\mu\nu} = \sum_\kappa P^\sigma_{\mu\kappa} A_{\kappa\nu}. \tag{3}$$

For each spin channel, this matrix-product form can be viewed as replacing the usual contraction $P^\sigma K(P^\sigma)$ by a contraction of the exact-exchange matrix $K(P^\sigma)$ with a spin density filtered through the AO-space mixing matrix A.

For $A = \alpha I$, it reduces to

$$E_x^A = -\frac{\alpha}{2}\sum_\sigma \mathrm{Tr}[P^\sigma K(P^\sigma)]\,, \tag{4}$$

recovering the conventional spin-resolved Hartree–Fock exchange contribution multiplied by the global mixing fraction α. The total energy is written as

$$E_{\mathrm{total}} = E_{\mathrm{nuc}} + T_s + J + E_{xc}^{\mathrm{DFA}} + E_x^A + E_{\mathrm{disp}}, \tag{5}$$

where E_{nuc} is the nuclear repulsion energy, T_s is the non-interacting kinetic energy, J is the classical Coulomb energy, E_{xc}^{DFA} is the semilocal exchange–correlation part inherited from the baseline density functional, E_x^A is the learned matrix-product exchange term, and E_{disp} is the D3(BJ) dispersion correction.Kohn & Sham (1965); Grimme et al. (2010; 2011)

3 MODEL

We construct a molecular graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ from atomic numbers $\mathbf{Z} = \{Z_i\}_{i=1}^N$ and coordinates $\mathbf{R} = \{\mathbf{r}_i\}_{i=1}^N$, where nodes represent atoms and edges $(i, j) \in \mathcal{E}$ connect pairs with overlapping basis functions within cutoff. MACE Batatia et al. (2022) produces equivariant per-atom embeddings $\mathbf{n}_i$ that encode local chemical environment.

We predict the mixing matrix A using the Graph2Mat approach Febrer et al. (2025), which ensures rotational equivariance. Following Graph2Mat, orbital basis functions are grouped by angular momentum l, and the block-structured matrix A is constructed using equivariant sequences of tensor product layers:

$$\widehat{A}_{\nu\nu} = f_n^{Z_\nu}(\mathbf{n}_\nu), \quad \widehat{A}_{\kappa\nu} = f_e^{Z_\kappa, Z_\nu}(\mathbf{n}_\kappa, \mathbf{n}_\nu), \tag{6}$$

where superscripts denote predictions of element-type-specific network, and f_n and f_e are node and edge specific networks. Each network handles all angular momentum combinations for its element type(s). See detailed model description in Appendix. E.

$$A = \alpha(I + \widehat{A}), \tag{7}$$

Predicted $A_{\mu\nu}$ modulates exchange energy in Eq. 2. We initialize the network such that $\widehat{A} = 0$, yielding $A = \alpha I$ with $\alpha = 0.2$ and recovering the B3LYP baseline at the start of training (see Appendix. E).

Table 1: Dataset composition.

DATASET	COUNT	DESCRIPTION
MSR-ACC	16,502	Atomization energies, CCSD(T)/CBS
RDB7	3,902	Reaction barriers, CCSD(T)-F12/cc-pVDZ
S66x8	451	Noncovalent dimers, CCSD(T)/CBS
Total	**20,855**	
MSR-ACC (test)	1,200	Atomization energies, CCSD(T)/CBS
RDB7 (test)	6,883	Reaction barriers, CCSD(T)-F12/cc-pVDZ
Total	**8,083**	

4 EXPERIMENTS

4.1 DATASET

To enable efficient training on large reaction datasets, we adopt a precomputation strategy that separates baseline DFT calculations from neural network training. For each molecule, we perform a one-time B3LYP/def2-QZVP SCF calculation with PySCF Sun et al. (2020) and store all non-trainable components of total energy Eq. 5: the spin-resolved density matrices $P^{\uparrow}$ and $P^{\downarrow}$, the total density matrix $P = P^{\uparrow} + P^{\downarrow}$, the corresponding exchange matrices $K(P^{\uparrow})$ and $K(P^{\downarrow})$, nuclear repulsion energy E_{nuc}, kinetic energy $T_s[P]$, Coulomb energy $J[P]$, semilocal B3LYP XC energy $E_{\text{xc}}^{\text{B3LYP}}$ (excluding HF exchange), D3(BJ) dispersion correction E_{disp} (B3LYP), and molecular geometry. All listed baseline quantities are precomputed once using unrestricted B3LYP with the def2-QZVP basis set before training.

We train on reaction energies rather than absolute molecular energies, because energy differences have well-defined reference values and, in stoichiometric combinations, largely cancel systematic biases present in absolute total energies. We train on two complementary large-scale datasets of high-level computed reaction energies: the MSR-ACC atomization-energy benchmark Ehlert et al. (2026), RDB7 reaction-barrier database Liu et al. (2025) and subset of S66x8 Brauer et al. (2016) noncovalent interaction curves. This combined dataset (Table 1) provides diverse coverage across thermochemistry, kinetics, and weak interactions, enabling robust generalization across distinct chemical reactions.

4.2 TRAINING

For reaction products j and reagents i with stoichiometric coefficients c_j and c_i, the predicted reaction energy is

$$\Delta E_{\text{rxn}}^{\text{pred}} = \sum_{\text{products}} c_j E(\mathbf{Z}_j, \mathbf{R}_j) - \sum_{\text{reactants}} c_i E(\mathbf{Z}_i, \mathbf{R}_i). \tag{8}$$

Training on diverse reaction datasets requires balancing contributions across energy scales. Large-magnitude reactions such as atomization energies (hundreds of kcal/mol) dominate standard loss functions like MSE, while weak interactions like dispersion and hydrogen bonding (1–10 kcal/mol) contribute negligibly to the gradient. We use a scaled mean squared error loss:

$$\mathcal{L} = \frac{1}{N} \sum_{k=1}^{N} \frac{(\Delta E_k^{\text{pred}} - \Delta E_k^{\text{ref}})^2}{\epsilon + |\Delta E_k^{\text{ref}}|}, \tag{9}$$

where the denominator normalizes each reaction by its magnitude, N is the number of reactions in the dataset, and ϵ is a small constant chosen such that $\epsilon = 0.01 \ll |\Delta E_k^{\text{ref}}|$ for all reactions.

Table 2: Mean absolute reaction/atomization energy errors (MAE) and standard deviations (std) in kcal/mol for the MSR and RDB7 test sets.

DATASET	METHOD	MAE $\pm$ STD [kcal/mol]
MSR Test (Atomization Energy)	OSE26	$\mathbf{0.85 \pm 0.85}$
	ωB97X-V	4.27 ± 3.15
	B3LYP D_3BJ	5.26 ± 4.03
	M06-2X	3.38 ± 2.40
RDB7 Test (Kinetics)	OSE26	$\mathbf{1.23 \pm 1.33}$
	ωB97M(2)	1.68 ± 1.87
	ωB97X-D3	2.59 ± 2.30
	MN15	2.08 ± 1.77

We used Adafactor Shazeer & Stern (2018) optimizer with learning rate 5×10^{-4} and gradient clipping threshold 0.5. Gradients are accumulated over 10 reactions before each optimizer step. Training runs for up to 34 epochs. With precomputed quantities, one epoch requires $\sim$120 minutes on NVIDIA RTX5090 GPU and Intel Core i7-12700K.

4.3 Results

The validation datasets probe distinct physical and chemical properties: RDB7 evaluates reaction barrier heights, while MSR-ACC-TAE assesses atomization energies, which traditionally demand different functional designs. In these tests, the predicted atomization energies are computed using the (Eq. equation 5) and compared to the corresponding reference values.

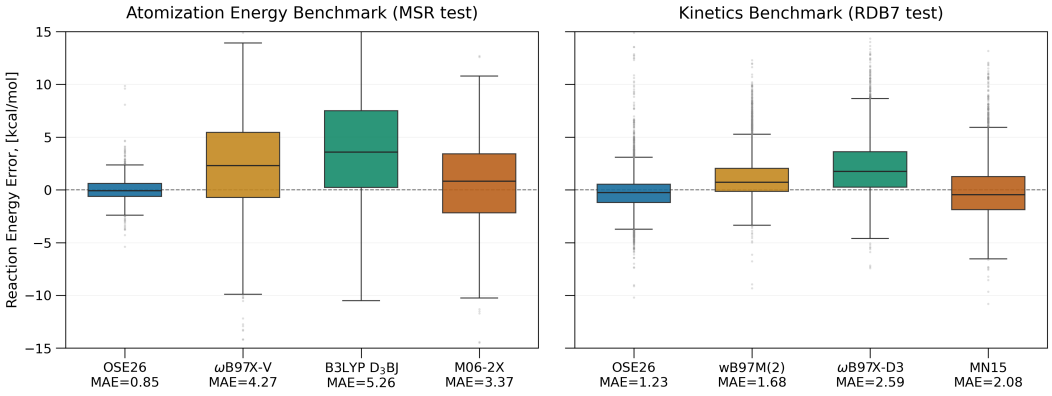


Figure 2: Boxplot comparison of atomization energy errors on the MSR-ACC-TAE25 validation benchmark. Mean absolute errors (MAE) are shown in parentheses below each method name (left). Boxplot comparison of proposed NN-based method in reaction energy errors at RDB7 validation dataset for ωB97M(2), ωB97X-D3, and MN15 methods on the RDB7 validation dataset. (right)

Left plot on the figure 2 reports the distribution of atomization energy errors across the test set and summarizes the corresponding MAE values, OSE26 gives MAE of 0.85 kcal/mol. Right plot on the Figure 2 presents a boxplot comparison of reaction energy errors for proposed NN-based method and the most accurate DFT functionals including double-hybrid (ωB97M(2), range-separated hybrid ωB97X-D3 and global hybrid MN15) evaluated on the RDB7 validation dataset.

Remarkably, our neural network model achieves superior accuracy compared to the state-of-the-art double-hybrid functional ωB97M(2) on the RDB7 dataset, while requiring only the computational resources of standard hybrid functionals. Whereas ωB97M(2) necessitates two orders of magnitude

more computational time due to its perturbative second-order correlation treatment, our approach delivers MAE of 1.23 kcal/mol (compared to 1.68 kcal/mol for ωB97M(2)).

5 Discussion

5.1 Theoretical Context and Functional Design

Historically, the introduction of double hybrid functionals (the fifth rung of Jacob's Ladder) was driven by the need to resolve the stubborn dichotomy between reaction kinetics and thermodynamics. Standard global hybrids typically face a compromise: high fractions of exact exchange improve barrier heights by reduction of self-interaction error, but often degrade atomization energies due to uncompensated static correlation. Double hybrids address this by incorporating nonlocal MP2 correlation from virtual orbitals to balance these competing requirements Grimme (2006). Remarkably, our findings suggest that ascending to the virtual orbital space is not strictly necessary to decouple these physical descriptions. By exploiting the geometric semantics within the occupied orbitals Hilbert subspace, OSE26 achieves a similar simultaneous accuracy for both RDB7 (reaction barriers) and MSR (atomization energies) benchmarks (see 5.3 Section) without changing the formal asymptotic scaling of exact-exchange evaluation. In addition, OSE26 admits an efficient self-consistent implementation with an explicit exchange potential and analytic nuclear forces (see Appendices A and B), capabilities that are still rarely available for modern local hybrids and double hybrids.

Orbital-specific exchange formalism occupies a unique position in the landscape of flexible functional design, distinct from both global adaptive schemes and standard local hybrid strategies. Adaptive global hybrids, such as those employing machine-learned schemes to predict optimal mixing parameters Khan et al. (2025), rely on geometric semantics to vary the exchange fraction α, but they apply this parameter globally, as a constant scalar across the entire integration space. Conversely, standard local hybrid functionals modulate $\alpha(\mathbf{r})$ at each point in real space. Local hybrids and OSE parameterize exchange flexibility along different axes: the former via a real-space factor, the latter via AO-pair weights. Consequently, at a point $\mathbf{r}$ where several orbitals contribute, a local hybrid still applies a single $a(\mathbf{r})$, whereas OSE can assign different weights to different AO-pair exchange channels. We therefore present AO-pair weighting as a complementary parameterization of exchange flexibility, rather than as a replacement for real-space local mixing or range separation. Our approach circumvents this bottleneck by operating directly in the orbital Hilbert space, allowing each AO pair to carry a learned exchange weight. This AO-pair parameterization is an inductive bias for the exchange energy; in this work we demonstrate improvements in reaction-energy accuracy. The resulting exchange modulation can be visualized through local exchange-energy densities; see Appendix D. While the local exchange energy density used in standard local hybrids is not unique (defined only up to a Laplacian or gauge transformation), presented orbital-pair formalism operates on well-defined integral quantities. This ensures that the modulation of exchange is robust and independent of the arbitrary choice of the energy density gauge, avoiding this problem by construction Jaramillo et al. (2003).

5.2 Computational efficiency

The dominant cost is evaluating $K(P)$ during hybrid SCF (identical to all hybrid functionals). For the self-consistent implementation, the computational cost is similar to that of range-separated hybrids. Alternatively, OSE26 can be efficiently evaluated as a post-SCF correction to a hybrid functional (see. A). With density fitting (RI-K) Weigend (2008), this scales as $\mathcal{O}(N^2 N_{\text{aux}})$ where N_{aux} is the size of auxiliary basis set, feasible for 100+ atom systems. Neural network inference adds negligible overhead (< 1 s per molecule on GPU).

5.3 Limitations

In the present construction the mixing matrix is predicted from nuclear geometry alone, $A = A(\mathbf{R})$, and the same A is shared by both spin channels. Consequently, charge states or spin multiplicities that share (nearly) the same nuclear framework receive identical orbital-pair weights, even though their exact-exchange requirements can differ substantially. A natural next step is to make the map

density-dependent, $A = A(\mathbf{R}, P^\sigma)$, so that distinct electronic states can induce distinct AO-pair mixing while retaining the OSE energy form. More broadly, the predictor is trained in def2-QZVP on main-group elements (H–Ar+Br); transferring to other bases or to d/f-block chemistry will require retraining or basis-independent descriptors. Our evaluation targets ground-state reaction energetics and equilibrium geometries and does not systematically probe densities, response properties, or regimes with extreme density-functional pathologies (e.g. strongly one-electron or long-range charge-transfer cases). These choices reflect the scope of a first study of the OSE formalism rather than an exhaustive claim of universality.

5.4 Broader Impact Statement

Hybrid DFT functionals are among the most widely used tools in computational chemistry: catalyst design Butera (2024), materials discovery Jain et al. (2013), and computer-aided drug design LaPointe & Weaver (2007), where systematic errors of a few kcal/mol in activation energies propagate via Arrhenius-type relations into order-of-magnitude errors in predicted rates and selectivities Liu et al. (2025). Improvements in functional accuracy can therefore directly translate into more reliable candidate rankings and better experimental decisions across these domains. This work is a methodological contribution at the level of electronic-structure theory and does not target specific deployment scenarios. We do not identify societal risks specific to this work beyond those shared with the broader field of computational chemistry.

6 Conclusion

We introduced OSE26, an orbital-pair generalization of hybrid density functional theory in which the scalar exact-exchange fraction is replaced by a geometry-dependent AO×AO interaction map predicted by an E(3)-equivariant neural network. Unlike global adaptive hybrids, which assign a single mixing parameter to an entire molecule, OSE26 modulates the exchange operator at the level of orbital pairs, providing a structured inductive bias between conventional hybrid DFT and fully data-driven energy correction models.

Across atomization energies and reaction barriers, OSE26 improves to near chemical accuracy and outperforms both scalar adaptive-hybrid and delta-learning baselines under the same neural backbone. Matched comparisons support a benefit from the OSE exchange-operator parameterization over adaptive global α and unconstrained Δ-learning under the same backbone; a retrained diagonal-only A OSE is weaker than full OSE, showing that off-diagonal AO-pair terms contribute materially beyond diagonal-only modulation. Additional self-consistent calculations indicate that the learned functional can be evaluated with stable SCF convergence and hybrid-like computational cost. This exchange functional implementation provides self-consistency, roto-translational energy invariance and analytical nuclear forces.

AI use statement

We have not used generative AI tools to help develop theoretical models or conceptual frameworks, formulate mathematical claims, provide critical ingredients for proving mathematical claims, propose or refine hypotheses, design or provide feedback on research methodology or experiments, implement methods, support qualitative and thematic data analysis, interpret results and generation of synthetic data sets, assist with translation, cleaning and reformation of dataset. Writing proofs is not applicable to this work. Additionally, we used generative AI tools to propose keywords for a research paper. We have reviewed all AI-assisted work: we checked LLM-generated language corrections. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

References

Carlo Adamo and Vincenzo Barone. Toward reliable density functional methods without adjustable parameters: The pbe0 model. *The Journal of Chemical Physics*, 110(13):6158–6170, April 1999. ISSN 1089-7690. doi: 10.1063/1.478522. URL http://dx.doi.org/10.1063/1.478522.

Rafael B. Araujo, Gabriel L. S. Rodrigues, Egon Campos dos Santos, and Lars G. M. Pettersson. Adsorption energies on transition metal surfaces: towards an accurate and balanced description. *Nature Communications*, 13(1), November 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-34507-y. URL http://dx.doi.org/10.1038/s41467-022-34507-y.

Ilyes Batatia, David P Kovacs, Gregor Simm, Christoph Ortner, and Gabor Csanyi. Mace: Higher order equivariant message passing neural networks for fast and accurate force fields. In *Advances in Neural Information Processing Systems 35*, NeurIPS 2022, pp. 11423–11436. Neural Information Processing Systems Foundation, Inc. (NeurIPS), 2022. doi: 10.52202/068431-0830. URL http://dx.doi.org/10.52202/068431-0830.

Brina Brauer, Manoj K. Kesharwani, Sebastian Kozuch, and Jan M. L. Martin. The s66x8 benchmark for noncovalent interactions revisited: explicitly correlated ab initio methods and density functional theory. *Physical Chemistry Chemical Physics*, 18(31):20905–20925, 2016. ISSN 1463-9084. doi: 10.1039/c6cp00688d. URL http://dx.doi.org/10.1039/C6CP00688D.

Vyacheslav S. Bryantsev, Mamadou S. Diallo, Adri C. T. van Duin, and William A. Goddard. Evaluation of b3lyp, x3lyp, and m06-class density functionals for predicting the binding energies of neutral, protonated, and deprotonated water clusters. *Journal of Chemical Theory and Computation*, 5(4):1016–1026, March 2009. ISSN 1549-9626. doi: 10.1021/ct800549f. URL http://dx.doi.org/10.1021/ct800549f.

Kyle R Bryenton, Adebayo A Adeleke, Stephen G Dale, and Erin R Johnson. Delocalization error: The greatest outstanding challenge in density-functional theory. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 13(2), March 2023.

Valeria Butera. Density functional theory methods applied to homogeneous and heterogeneous catalysis: a short review and a practical user guide. *Physical Chemistry Chemical Physics*, 26 (10):7950–7970, 2024. ISSN 1463-9084. doi: 10.1039/d4cp00266k. URL http://dx.doi.org/10.1039/D4CP00266K.

Eike Caldeweyher, Sebastian Ehlert, Andreas Hansen, Hagen Neugebauer, Sebastian Spicher, Christoph Bannwarth, and Stefan Grimme. A generally applicable atomic-charge dependent london dispersion correction. *The Journal of Chemical Physics*, 150(15), April 2019. ISSN 1089-7690. doi: 10.1063/1.5090222. URL http://dx.doi.org/10.1063/1.5090222.

Nicolas Chéron, Denis Jacquemin, and Paul Fleurat-Lessard. A qualitative failure of B3LYP for textbook organic reactions. *Phys. Chem. Chem. Phys.*, 14(19):7170–7175, May 2012.

D. O. Demchenko and M. A. Reshchikov. Koopmans' tuning of hse hybrid density functional for calculations of defects in semiconductors: A case study of carbon acceptor in gan. *Journal of Applied Physics*, 127(15), April 2020. ISSN 1089-7550. doi: 10.1063/1.5140661. URL http://dx.doi.org/10.1063/1.5140661.

Sebastian Ehlert, Jan Hermann, Thijs Vogels, Victor Garcia Satorras, Stephanie Lanius, Marwin Segler, Klaas J. H. Giesbertz, Derk P. Kooi, Kenji Takeda, Chin-Wei Huang, Giulia Luise, Rianne van den Berg, Paola Gori-Giorgi, and Amir Karton. Accurate chemistry collection: Coupled cluster atomization energies for broad chemical space. *Scientific Data*, 13(1), April 2026. ISSN 2052-4463. doi: 10.1038/s41597-026-07200-8. URL http://dx.doi.org/10.1038/s41597-026-07200-8.

Matthias Ernzerhof and John P. Perdew. Generalized gradient approximation to the angle- and system-averaged exchange hole. *The Journal of Chemical Physics*, 109(9):3313–3320, 1998. ISSN 1089-7690. doi: 10.1063/1.476928. URL http://dx.doi.org/10.1063/1.476928.

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, April 2025. ISSN 2632-2153. doi: 10.1088/2632-2153/adc871. URL http://dx.doi.org/10.1088/2632-2153/adc871.

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

Lars Goerigk, Andreas Hansen, Christoph Bauer, Stephan Ehrlich, Asim Najibi, and Stefan Grimme. A look at the density functional theory zoo with the advanced gmtkn55 database for general main group thermochemistry, kinetics and noncovalent interactions. *Physical Chemistry Chemical Physics*, 19(48):32184–32215, 2017. ISSN 1463-9084. doi: 10.1039/c7cp04913g. URL http://dx.doi.org/10.1039/C7CP04913G.

Stefan Grimme. Semiempirical hybrid density functional with perturbative second-order correlation. *The Journal of Chemical Physics*, 124(3), January 2006. ISSN 1089-7690. doi: 10.1063/1.2148954. URL http://dx.doi.org/10.1063/1.2148954.

Stefan Grimme, Jens Antony, Stephan Ehrlich, and Helge Krieg. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.*, 132(15):154104, April 2010.

Stefan Grimme, Stephan Ehrlich, and Lars Goerigk. Effect of the damping function in dispersion corrected density functional theory. *Journal of Computational Chemistry*, 32(7):1456–1465, March 2011. ISSN 1096-987X. doi: 10.1002/jcc.21759. URL http://dx.doi.org/10.1002/jcc.21759.

P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Physical Review*, 136(3B):B864–B871, November 1964. ISSN 0031-899X. doi: 10.1103/physrev.136.b864. URL http://dx.doi.org/10.1103/PhysRev.136.B864.

Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, and Kristin A. Persson. Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL Materials*, 1(1), 2013. ISSN 2166-532X. doi: 10.1063/1.4812323. URL http://dx.doi.org/10.1063/1.4812323.

Juanita Jaramillo, Gustavo E. Scuseria, and Matthias Ernzerhof. Local hybrid functionals. *The Journal of Chemical Physics*, 118(3):1068–1073, January 2003. ISSN 1089-7690. doi: 10.1063/1.1528936. URL http://dx.doi.org/10.1063/1.1528936.

Danish Khan, Alastair J. A. Price, Bing Huang, Maximilian L. Ach, and O. Anatole von Lilienfeld. Adapting hybrid density functionals with machine learning. *Science Advances*, 11(5), January 2025. ISSN 2375-2548. doi: 10.1126/sciadv.adt7769. URL http://dx.doi.org/10.1126/sciadv.adt7769.

James Kirkpatrick, Brendan McMorrow, David H. P. Turban, Alexander L. Gaunt, James S. Spencer, Alexander G. D. G. Matthews, Annette Obika, Louis Thiry, Meire Fortunato, David Pfau, Lara Román Castellanos, Stig Petersen, Alexander W. R. Nelson, Pushmeet Kohli, Paula Mori-Sánchez, Demis Hassabis, and Aron J. Cohen. Pushing the frontiers of density functionals by solving the fractional electron problem. *Science*, 374(6573):1385–1389, December 2021. ISSN 1095-9203. doi: 10.1126/science.abj6511. URL http://dx.doi.org/10.1126/science.abj6511.

Sascha Klawohn, Hilke Bahmann, and Martin Kaupp. Implementation of molecular gradients for local hybrid density functionals using seminumerical integration techniques. *Journal of Chemical Theory and Computation*, 12(9):4254–4262, August 2016. ISSN 1549-9626. doi: 10.1021/acs.jctc.6b00486. URL http://dx.doi.org/10.1021/acs.jctc.6b00486.

W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Physical Review*, 140(4A):A1133–A1138, November 1965. ISSN 0031-899X. doi: 10.1103/physrev.140.a1133. URL http://dx.doi.org/10.1103/PhysRev.140.A1133.

Kirill Kulaev, Alexander Ryabov, Michael Medvedev, Evgeny Burnaev, and Vladimir Vanovskiy. On the practical applicability of modern dft functionals for chemical computations. case study of dm21 applicability for geometry optimization, 2025. URL https://arxiv.org/abs/2501.12149.

Shenna LaPointe and Donald Weaver. A review of density functional theory quantum mechanics as applied to pharmaceutically relevant systems. *Current Computer Aided-Drug Design*, 3(4):290–296, December 2007. ISSN 1573-4099. doi: 10.2174/157340907782799390. URL http://dx.doi.org/10.2174/157340907782799390.

Xiao Liu, Kevin A. Spiekermann, Angiras Menon, William H. Green, and Martin Head-Gordon. Revisiting a large and diverse data set for barrier heights and reaction energies: best practices in density functional theory calculations for chemical kinetics. *Physical Chemistry Chemical Physics*, 27(25):13326–13339, 2025. ISSN 1463-9084. doi: 10.1039/d5cp01181g. URL http://dx.doi.org/10.1039/D5CP01181G.

Narbe Mardirossian and Martin Head-Gordon. 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. ISSN 1362-3028. doi: 10.1080/00268976.2017.1333644. URL http://dx.doi.org/10.1080/00268976.2017.1333644.

W L McMillan. Ground state of Liquid He4. *Phys. Rev.*, 138(2A):A442–A451, April 1965.

Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953. ISSN 1089-7690. doi: 10.1063/1.1699114. URL http://dx.doi.org/10.1063/1.1699114.

Ryo Nagai, Ryosuke Akashi, and Osamu Sugino. Completing density functional theory by machine learning hidden messages from molecules. *npj Computational Materials*, 6(1), May 2020. ISSN 2057-3960. doi: 10.1038/s41524-020-0310-0. URL http://dx.doi.org/10.1038/s41524-020-0310-0.

J. P. Perdew and Alex Zunger. Self-interaction correction to density-functional approximations for many-electron systems. *Physical Review B*, 23(10):5048–5079, May 1981. ISSN 0163-1829. doi: 10.1103/physrevb.23.5048. URL http://dx.doi.org/10.1103/PhysRevB.23.5048.

John P. Perdew and Karla Schmidt. Jacob's ladder of density functional approximations for the exchange-correlation energy. In *AIP Conference Proceedings*, volume 577, pp. 1–20. AIP, 2001. doi: 10.1063/1.1390175. URL https://doi.org/10.1063/1.1390175.

John P. Perdew and Yue Wang. Accurate and simple analytic representation of the electron-gas correlation energy. *Physical Review B*, 45(23):13244–13249, 1992. ISSN 1095-3795. doi: 10.1103/physrevb.45.13244. URL http://dx.doi.org/10.1103/PhysRevB.45.13244.

John P. Perdew and Wang Yue. Accurate and simple density functional for the electronic exchange energy: Generalized gradient approximation. *Physical Review B*, 33(12):8800–8802, 1986. ISSN 0163-1829. doi: 10.1103/physrevb.33.8800. URL http://dx.doi.org/10.1103/PhysRevB.33.8800.

John P. Perdew, J. A. Chevary, S. H. Vosko, Koblar A. Jackson, Mark R. Pederson, D. J. Singh, and Carlos Fiolhais. Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. *Physical Review B*, 46(11):6671–6687, 1992. ISSN 1095-3795. doi: 10.1103/physrevb.46.6671. URL http://dx.doi.org/10.1103/PhysRevB.46.6671.

John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Physical Review Letters*, 77(18):3865–3868, October 1996. ISSN 1079-7114. doi: 10.1103/physrevlett.77.3865. URL http://dx.doi.org/10.1103/PhysRevLett.77.3865.

John P. Perdew, Stefan Kurth, Aleš Zupan, and Peter Blaha. Accurate density functional with correct formal properties: A step beyond the generalized gradient approximation. *Physical Review Letters*, 82(12):2544–2547, March 1999. ISSN 1079-7114. doi: 10.1103/physrevlett.82.2544. URL http://dx.doi.org/10.1103/PhysRevLett.82.2544.

Alexander Ryabov, Iskander Akhatov, and Petr Zhilyaev. Neural network interpolation of exchange-correlation functional. *Scientific Reports*, 10(1), May 2020. ISSN 2045-2322. doi: 10.1038/s41598-020-64619-8. URL http://dx.doi.org/10.1038/s41598-020-64619-8.

Alexander Ryabov, Iskander Akhatov, and Petr Zhilyaev. Application of two-component neural network for exchange-correlation functional interpolation. *Scientific Reports*, 12(1), August 2022. ISSN 2045-2322. doi: 10.1038/s41598-022-18083-1. URL http://dx.doi.org/10.1038/s41598-022-18083-1.

Peter R. Schreiner, Andrey A. Fokin, Robert A. Pascal, and Armin de Meijere. Many density functional theory approaches fail to give reliable large hydrocarbon isomer energy differences. *Organic Letters*, 8(17):3635–3638, 2006. ISSN 1523-7052. doi: 10.1021/ol0610486. URL http://dx.doi.org/10.1021/ol0610486.

Noam Shazeer and Mitchell Stern. Adafactor: Adaptive learning rates with sublinear memory cost, 2018. URL https://arxiv.org/abs/1804.04235.

Neil Qiang Su, Zhenyu Zhu, and Xin Xu. Doubly hybrid density functionals that correctly describe both density and energy for atoms. *Proceedings of the National Academy of Sciences*, 115(10): 2287–2292, February 2018. ISSN 1091-6490. doi: 10.1073/pnas.1713047115. URL http://dx.doi.org/10.1073/pnas.1713047115.

Jianwei Sun, Adrienn Ruzsinszky, and John P. Perdew. Strongly constrained and appropriately normed semilocal density functional. *Physical Review Letters*, 115(3), 2015. ISSN 1079-7114. doi: 10.1103/physrevlett.115.036402. URL http://dx.doi.org/10.1103/PhysRevLett.115.036402.

Qiming Sun, Xing Zhang, Samragni Banerjee, Peng Bao, Marc Barbry, Nick S Blunt, Nikolay A Bogdanov, George H Booth, Jia Chen, Zhi-Hao Cui, Janus J Eriksen, Yang Gao, Sheng Guo, Jan Hermann, Matthew R Hermes, Kevin Koh, Peter Koval, Susi Lehtola, Zhendong Li, Junzi Liu, Narbe Mardirossian, James D McClain, Mario Motta, Bastien Mussard, Hung Q Pham, Artem Pulkin, Wirawan Purwanto, Paul J Robinson, Enrico Ronca, Elvira R Sayfutyarova, Maximilian Scheurer, Henry F Schurkus, James E T Smith, Chong Sun, Shi-Ning Sun, Shiv Upadhyay, Lucas K Wagner, Xiao Wang, Alec White, James Daniel Whitfield, Mark J Williamson, Sebastian Wouters, Jun Yang, Jason M Yu, Tianyu Zhu, Timothy C Berkelbach, Sandeep Sharma, Alexander Yu Sokolov, and Garnet Kin-Lic Chan. Recent developments in the PySCF program package. *J. Chem. Phys.*, 153(2):024109, July 2020.

Jianmin Tao, John P. Perdew, Viktor N. Staroverov, and Gustavo E. Scuseria. Climbing the density functional ladder: Nonempirical meta–generalized gradient approximation designed for molecules and solids. *Physical Review Letters*, 91(14), 2003. ISSN 1079-7114. doi: 10.1103/physrevlett.91.146401. URL http://dx.doi.org/10.1103/PhysRevLett.91.146401.

S. H. Vosko, L. Wilk, and M. Nusair. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Canadian Journal of Physics*, 58(8):1200–1211, August 1980. ISSN 1208-6045. doi: 10.1139/p80-159. URL http://dx.doi.org/10.1139/p80-159.

Florian Weigend. Hartree-Fock exchange fitting basis sets for H to Rn. *J. Comput. Chem.*, 29(2): 167–175, January 2008.

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. ISSN 1463-9084. doi: 10.1039/b508541a. URL http://dx.doi.org/10.1039/B508541A.

Elisabeth Wruss, Egbert Zojer, and Oliver T. Hofmann. Distinguishing between charge-transfer mechanisms at organic/inorganic interfaces employing hybrid functionals. *The Journal of Physical Chemistry C*, 122(26):14640–14653, May 2018. ISSN 1932-7455. doi: 10.1021/acs.jpcc.8b03699. URL http://dx.doi.org/10.1021/acs.jpcc.8b03699.

Jin Xiao, Yingfeng Zhang, Bowen Li, Shuwen Zhang, Ya Gao, Wei Chen, Han Wang, John Z. H. Zhang, and Tong Zhu. A deep learning-augmented density functional framework for reaction modeling with chemical accuracy. *JACS Au*, 5(8):3892–3903, 2025. ISSN 2691-3704. doi: 10.1021/jacsau.5c00541. URL http://dx.doi.org/10.1021/jacsau.5c00541.

Igor Ying Zhang and Xinguo Ren. Introduction to the fifth-rung density functional approximations: Concept, formulation, and applications, 2023. URL https://arxiv.org/abs/2301.12119.

Qing Zhao and Heather J. Kulik. Stable surfaces that bind too tightly: Can range-separated hybrids or dft+u improve paradoxical descriptions of surface chemistry? *The Journal of Physical Chemistry Letters*, 10(17):5090–5098, August 2019. ISSN 1948-7185. doi: 10.1021/acs.jpclett.9b01650. URL http://dx.doi.org/10.1021/acs.jpclett.9b01650.

Xiao Zheng, LiHong Hu, XiuJun Wang, and GuanHua Chen. A generalized exchange-correlation functional: the neural-networks approach. *Chemical Physics Letters*, 390(1-3):186–192, May 2004. ISSN 0009-2614. doi: 10.1016/j.cplett.2004.04.020. URL http://dx.doi.org/10.1016/j.cplett.2004.04.020.

Jiří Čížek. On the correlation problem in atomic and molecular systems. calculation of wavefunction components in ursell-type expansion using quantum-field theoretical methods. *The Journal of Chemical Physics*, 45(11):4256–4266, December 1966. ISSN 1089-7690. doi: 10.1063/1.1727484. URL http://dx.doi.org/10.1063/1.1727484.

A SELF-CONSISTENT POTENTIAL

In this section we derive the effective exchange potential used in the Kohn-Sham SCF for the OSE26 exchange functional. The matrix-product exchange energy is defined as

$$E_x^A[\{P^\sigma\}] = -\frac{1}{2}\sum_\sigma \sum_{\mu\nu\kappa} P^\sigma_{\mu\kappa} A_{\kappa\nu} K_{\nu\mu}(P^\sigma). \tag{10}$$

The exchange contribution entering the SCF equations is defined as the effective Fock matrix associated with $E_x^A[\{P^\sigma\}]$. For each spin channel, it is obtained by differentiating the exchange energy with respect to the corresponding spin density matrix P^σ. The derivative contains two contributions: one from the explicit factor P^σ and one from the implicit dependence of $K(P^\sigma)$ on P^σ. For each spin channel σ, the corresponding exchange contribution to the Fock matrix is

$$v^{A,x,\sigma} = -\frac{1}{2}\,\mathrm{sym}[AK(P^\sigma) + K(P^\sigma A)]\,, \qquad \mathrm{sym}(X) = \frac{1}{2}\left(X + X^\mathsf{T}\right). \tag{11}$$

In the implementation, $K(P^\sigma)$ and $K(P^\sigma A)$ are evaluated separately for each spin channel using the same PySCF exchange routines as in the reference hybrid functional. At each SCF iteration, the spin-resolved density matrices $P^\uparrow$ and $P^\downarrow$ are updated from the occupied orbitals, and the corresponding exchange contributions $v_x^{A,\uparrow}$ and $v_x^{A,\downarrow}$ are recomputed. For $A = \alpha I$, it reduces independently in each spin channel to

$$v_x^{A,\sigma} = -\alpha K(P^\sigma),$$

recovering the conventional global-hybrid exchange operator.

For the trained OSE26 functional, SCF convergence dynamics are very stable in practice. OSE26 typically converges in essentially the same number of iterations as the underlying B3LYP reference, or at most requires two to three additional SCF steps. Figure 3 illustrates SCF histories for B3LYP and OSE26.

To confirm that OSE26 remains accurate in a fully self-consistent setting, we performed an additional check on 30 reactions randomly sampled from the RDB7. We evaluated OSE26 both as a post-SCF correction on converged B3LYP orbitals and after convergence of SCF with them. For this subset, we report that all SCF calculations were converged (from minao initial guess), post-SCF and self-consistent OSE26 reaction energies agree within ~ 0.1 kcal/mol. Similarly, we report SCF results for the water-cluster dissociation-energy benchmark presented later App. D.

The OSE26 exchange operator involves two exact-exchange builds per spin channel, corresponding to $K(P^\sigma)$ and $K(P^\sigma A)$ in equation 11. In terms of both formal scaling the resulting SCF cost is comparable to that of range-separated hybrid functionals, which likewise require multiple exchange builds per iteration. Overall cost is significantly lower and more predictable than for modern local hybrid density functionals such as DM21 Kirkpatrick et al. (2021), whose SCF can exceed the cost of CCSD(T) even for small molecules due to the expensive grid-based XC potential and its convergence issuesKulaev et al. (2025), it has been shown that maintaining SCF robustness requires incorporating explicit SCF-related regularization terms into the training loss Kirkpatrick et al. (2021).

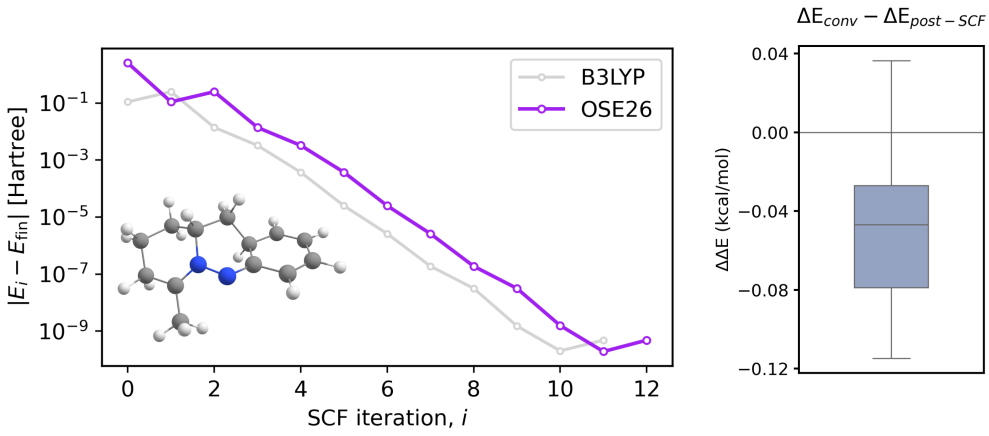


Figure 3: SCF convergence histories of KS-DFT with B3LYP and OSE26 on the +1 charged molecule visualized at the picture. (left). Comparison between reaction energies ΔE_{conv} from fully converged SCF with OSE26 and energies $\Delta E_{\text{post-SCF}}$ calculated directly from B3LYP orbitals (as used in training and test procedures) on the set of molecules randomly sampled from RDB7 test subset (right).

B Analytical nuclear forces

For a nuclear coordinate R_A, the OSE26 contribution to the force is

$$F_x^A = -\frac{\partial E_x^A}{\partial R_A}.$$

Differentiating equation 10 with respect to R_A gives

$$\frac{\partial E_x^A}{\partial R_A} = -\frac{1}{2}\sum_\sigma \mathrm{Tr}\left[P^\sigma \frac{\partial A}{\partial R_A} K(P^\sigma)\right] - \frac{1}{2}\sum_\sigma \mathrm{Tr}\left[P^\sigma A \frac{\partial K(P^\sigma)}{\partial R_A}\right]. \tag{12}$$

where the first term describes the geometry response of the learned mixing matrix A, while the second term is the usual exchange-integral response, with both contributions summed over spin channels. Forces therefore evaluated together with standard forces terms: Coulomb nuclear-electron, nuclear-nuclear, DFA XC gradient terms with the standard Pulay contributions associated with the AO basis and overlap derivatives.

Figure 4 summarizes the performance of OSE26 for equilibrium geometries on the LMGB35 benchmark Caldeweyher et al. (2019) set of small main-group molecules with experimentally defined bond lengths in gas phase. The MAE in bond lengths amounts to 0.53 pm for OSE26, which is slightly better than the 0.63 pm error reported for the neural-network functional DM21Kulaev et al. (2025) and better than 0.77 pm error obtained with the range-separated hybrid ωB97X-V. This demonstrates that the improved formal structure of OSE26 does not come at the expense of geometric accuracy and, in fact, yields bond-length errors on par with other modern methods.

This behaviour differs markedly from that of modern local hybrid density functionals. For such functionals, analytic nuclear gradients have not yet been worked out in a broadly available form; to the best of our knowledge, only a single seminumerical gradient implementation has been reported for a specific family of local hybrids.Klawohn et al. (2016) In practice, geometry optimization with local hybrid density functionals often relies on numerical finite-difference gradients, which can be substantially more expensive than high-level wavefunction methods such as CCSD(T).Kulaev et al. (2025). A similar situation holds for many double-hybrid functionals: analytic gradients exist for selected implementations, but a general, widely adopted framework for self-consistent double-hybrid forces is still lacking, which limits their routine use in large-scale geometry optimization, molecular dynamics and Gibbs free energy estimation. In contrast, OSE26 admits a straightforward analytic force formulation, making geometry optimization no more demanding than for conventional global or range-separated hybrids.

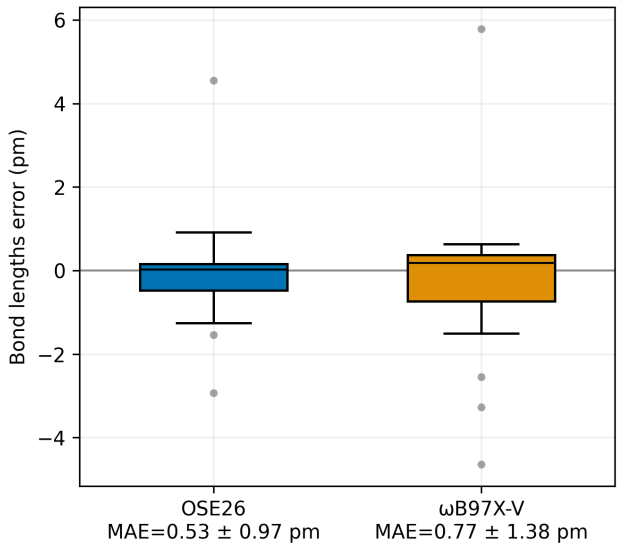


Figure 4: Box plots of errors and MAE (in pm) on the LMGB35 benchmark set for OSE26 and range-separated hybrid ωB97X-V.

C Ablation study

We compare OSE26 to simpler ML baselines under a shared backbone and training setup. All models use the same MACE architecture (with an MLP head of similar size to Graph2Mat), the same reaction-energy dataset, batch size (10 reactions), and Adafactor optimizer.

We consider four alternative strategies in addition to full OSE26: (i) an adaptive hybrid that still uses a conventional global hybrid XC functional but predicts a system-dependent scalar mixing parameter $\alpha(Z, R)$ (similar to Khan et al. (2025)), (ii) a Δ-learning baseline that leaves the underlying B3LYP-D3(BJ) XC functional untouched and learns a single scalar correction ΔE to the total energy, and (iii) OSE26-diag, a controlled architectural ablation trained with enforced zero off-diagonal elements of A. Full OSE26, in contrast, directly modifies the exchange operator via a learned AO–AO mixing matrix. (iv) DeePHF Xiao et al. (2025): which implements the energy correction $\Delta E = \sum_I F_{\mathrm{NN}}(\mathrm{desc}_I)$ from eigenvalues of local projected density matrices, added on top of B3LYP-D3(BJ) with original model hyperparameters. Table 3 reports mean absolute errors (MAEs) on RDB7 and MSR.

Table 3: Test RDB7 and MSR MAEs (kcal/mol) for OSE26 and baselines.

Method	RDB7 test MAE	MSR test MAE
Adaptive hybrid α	1.65	1.57
MACE Δ-learning	1.38	1.23
OSE26-diag (diagonal A)	1.55	1.03
OSE26 (ours)	**1.23**	**0.85**
DeePHF	2.70	4.48
B3LYP-D3(BJ)	2.79	5.26

We interpret these results as indicating that the way the correction is parameterized (basis and number of degrees of freedom) matters. In the adaptive hybrid, the α lives in a two-dimensional basis (E_x^{HF}, E_x^{DFA}) giving only one effective degree of freedom with a strong bias toward global rescaling. In the delta-learning baseline, the correction is represented in the high-dimensional latent space of the MACE network, with many weakly constrained degrees of freedom. In OSE26, the learnable parameters act in the high-dimensional AO$\times$AO exchange-matrix basis, this stronger inductive bias likely explains why OSE26 achieves lower errors than all baselines (see Fig. 5).

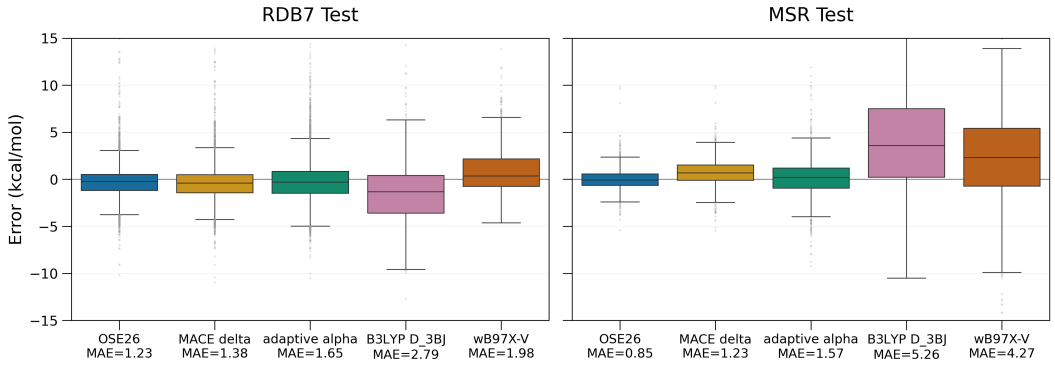


Figure 5: Ablation study results. Boxplot comparison of different method accuracy on reaction energy test data

D OSE VISUALIZATION

We use additional hold-out dataset of neutral and charged water clusters with energies estimated at the CCSD(T)/CBS Bryantsev et al. (2009) to visualize behaviour of OSE26 in details (OSE had no water clusters in training except the S66x8 neutral H_2O-H_2O dimer dissociation curve and in fact no charged systems in whole training dataset) and stress test system size extensity of OSE architecture. Fig. 6 shows that error does not grow (even lowering) with the system's size.

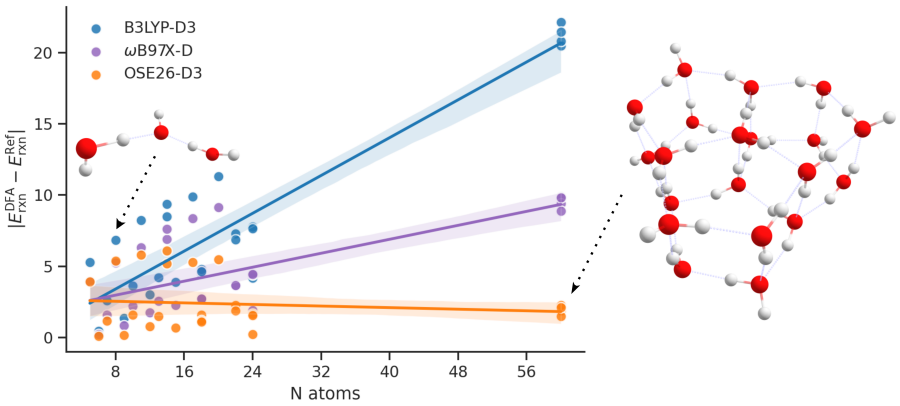


Figure 6: Dependence of the absolute error in water cluster binding energy on the number of atoms.

With employed dispersion corrections the energy errors in these systems are closely related to charge delocalization Bryenton et al. (2023); therefore, since OSE predicts better energetics in these systems, it is useful to visualize where in real space this modulation occurs in order to better understand the neural-network result. This can be done by comparing local exchange-energy densities $\varepsilon_x(\mathbf{r})$ for Hartree–Fock and OSE exchange:

$$\varepsilon_x^{\mathrm{HF}}(\mathbf{r}) = -\frac{1}{2}\sum_{\sigma}\sum_{\mu\nu\lambda\kappa} P_{\mu\nu}^{\sigma}\, P_{\lambda\kappa}^{\sigma}\, \phi_\mu(\mathbf{r})\, \phi_\lambda(\mathbf{r}) \int \frac{\phi_\nu(\mathbf{r}')\, \phi_\kappa(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\, d\mathbf{r}', \tag{13}$$

$$\varepsilon_x^{\mathrm{OSE}}(\mathbf{r}) = -\frac{1}{2}\sum_{\sigma}\sum_{\mu\nu\lambda\kappa\alpha} P_{\mu\alpha}^{\sigma}\, A_{\alpha\nu}\, P_{\lambda\kappa}^{\sigma}\, \phi_\mu(\mathbf{r})\, \phi_\lambda(\mathbf{r}) \int \frac{\phi_\nu(\mathbf{r}')\, \phi_\kappa(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\, d\mathbf{r}', \tag{14}$$

To illustrate the real-space effect of the learned exchange modulation, we examine the $OH^-(H_2O)_3$ cluster real space grid. Figure 7 compares the difference of local exchange-energy densities relative to B3LYP for self-consistent OSE26 and for the range-separated hybrid CAM-B3LYP. In the

hydrogen-bond region associated with the OH^- lone pair, OSE26 reduces the local exchange contribution relative to B3LYP, consistent with weaker binding and reduced charge delocalization; regions of enhanced OSE exchange coincide with lower electron density comparable with overall effect of range separation of HF exchange in CAM B3LYP.

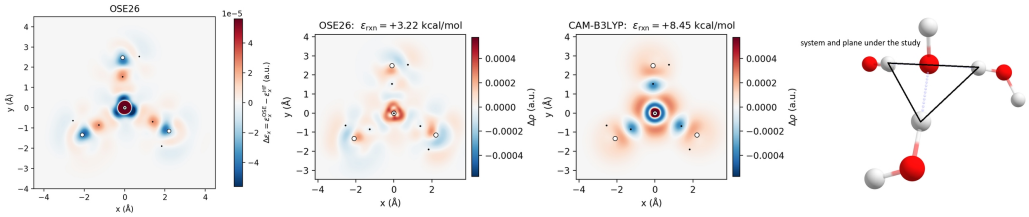


Figure 7: Real-space exchange modulation on the $OH^-(H_2O)_3$ cluster. Difference of local exchange-energy densities relative to B3LYP, $\Delta\varepsilon_x(\mathbf{r}) = \varepsilon_x(\mathbf{r}) - \varepsilon_x^{\mathrm{B3LYP}}(\mathbf{r})$, evaluated on the real space grid for self-consistent OSE26 (left) and CAM-B3LYP (right). OSE26 and the range-separated hybrid produce spatially distinct exchange redistributions; in the OH^- hydrogen-bond region, OSE26 weakens local exchange relative to B3LYP, consistent with reduced delocalization and improved binding energetics.

E Model Details

Each atom i carries an AO basis of size d_i, and the total number of AOs is

$$n_{\mathrm{AO}} = \sum_{i=1}^{N} d_i. \tag{15}$$

The AO mixing matrix $A \in \mathbb{R}^{n_{\mathrm{AO}} \times n_{\mathrm{AO}}}$ is organized into atom–atom blocks $A_{ij} \in \mathbb{R}^{d_i \times d_j}$, where A_{ij} collects all couplings between basis functions on atoms i and j. Diagonal blocks A_{ii} couple basis functions on the same atom, while off-diagonal blocks A_{ij} couple basis functions centered on different atoms.

Equivariant features Let $\mathbf{n}_i$ be the per-atom embedding produced by an $E(3)$-equivariant encoder (e.g. MACE Batatia et al. (2022)). It can be written as a direct sum of $O(3)$ irreps:

$$\mathbf{n}_i \in \bigoplus_{\ell=0}^{\ell_{\max}} \left(\mathbb{R}^{c_\ell} \otimes \mathcal{H}^{(\ell)} \right), \tag{16}$$

where $\mathcal{H}^{(\ell)}$ is the $(2\ell+1)$-dimensional irrep of degree ℓ and c_ℓ is the channel multiplicity. Edge geometry enters through relative vectors $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$ via rotationally invariant radial features $\phi(\|\mathbf{r}_{ij}\|)$ and equivariant spherical harmonics $Y_{\ell m}(\widehat{\mathbf{r}}_{ij})$.

Graph2Mat readout Graph2Mat constructs an equivariant matrix by predicting it block-by-block. Let $\widehat{A} \in \mathbb{R}^{n_{\mathrm{AO}} \times n_{\mathrm{AO}}}$ be partitioned into atom–atom blocks $\widehat{A}_{ij} \in \mathbb{R}^{d_i \times d_j}$, where d_i is the number of AOs on atom i. A molecular graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is formed so that $(i,j) \in \mathcal{E}$ whenever the basis functions on atoms i and j overlap (within a cutoff).

Let $\mathbf{n}_i$ denote the $E(3)$-equivariant embedding of atom i produced by an encoder. Graph2Mat defines two equivariant block-generating functions: a node function for diagonal blocks and an edge function for off-diagonal blocks, i.e.

$$\widehat{A}_{ii} = f_n^{Z_i}(\mathbf{n}_i), \tag{17}$$

$$\widehat{A}_{ij} = f_e^{Z_i, Z_j}(\mathbf{n}_i, \mathbf{n}_j), \qquad (i,j) \in \mathcal{E}. \tag{18}$$

Different element types (and element pairs) are handled by different f_n^Z and $f_e^{Z,Z'}$.

In the Graph2Mat, these functions are implemented using an equivariant tensor product: f_n is a tensor square and f_e is a tensor product. In e3nn Geiger & Smidt (2022) notation, a fully-connected tensor product produces outputs of the form

$$\mathrm{TP}(x, y)_w \;=\; \sum_{u,v} w_{uvw} \left(x_u \otimes y_v\right), \tag{19}$$

where w_{uvw} are trainable weights and $\otimes$ denotes the equivariant (Clebsch–Gordan) coupling.

Symmetrization. Because the exchange kernel is Hermitian, we constrain $\widehat{A}$ to be symmetric. Graph2Mat enforces this at the block level rather than by a post-hoc average $(\widehat{A} + \widehat{A}^\top)/2$. For each undirected neighbour pair $\{i, j\}$ the model predicts a single off-diagonal block $\widehat{A}_{ij}$ and sets

$$\widehat{A}_{ji} \;=\; \widehat{A}_{ij}^\top, \tag{20}$$

so the opposite directed edge is never predicted independently. Diagonal blocks are generated with a reduced tensor product whose irrep channels obey the block symmetry ij $=$ ji Febrer et al. (2025), which yields $\widehat{A}_{ii} = \widehat{A}_{ii}^\top$ by construction.

Baseline limit The block construction defines a matrix $\widehat{A}$. Then,

$$A = \alpha(I + \widehat{A}), \tag{21}$$

when NN parameters are initialized as zeros $\widehat{A} = 0 \Rightarrow A = \alpha I$, recovering the baseline global hybrid functional.

Neural network hyperparameters The model uses a MACE Batatia et al. (2022) encoder followed by a Graph2Mat Febrer et al. (2025) readout that predicts diagonal (node) and off-diagonal (edge) blocks. The AO basis is def2-QZVP Weigend & Ahlrichs (2005). Model supports all main-group chemistry elements before Ar and additional Br: H, He, Li, Be, B, C, N, O, F, Ne, Na, Mg, Al, Si, P, S, Cl, Ar, Br.

MACE encoder: Cutoff radius $r_{\max} = 6.0$ Å, 6 Bessel radial basis functions with polynomial cutoff of order 6, angular resolution $\ell_{\max} = 2$, two interaction layers, hidden irreps $8 \times 0\mathrm{e} + 8 \times 1\mathrm{o}$ and MLP irreps $8 \times 0\mathrm{e}$, 19 element types, average number of neighbours 8.0, correlation order 1.

Graph2Mat readout: Edge hidden irreps $4 \times 0\mathrm{e} + 4 \times 1\mathrm{o}$.