LEARNING EQUIVARIANT NON-LOCAL ELECTRON DENSITY FUNCTIONALS

Anonymous authors
Paper under double-blind review

ABSTRACT

The accuracy of density functional theory hinges on the approximation of *non-local* contributions to the exchange-correlation (XC) functional. To date, machine-learned and human-designed approximations suffer from insufficient accuracy, limited scalability, or dependence on costly reference data. To address these issues, we introduce Global Graph Exchange Correlation (GG-XC), a novel non-local XC functional based on equivariant graph neural networks. GG-XC combines semi-local functionals with a non-local feature density parametrized by an equivariant nuclei-centered point cloud representation of the electron density to capture long-range interactions. By differentiating through a self-consistent field solver, we train GG-XC requiring only energy targets. In our empirical evaluation, we find GG-XC to accurately reconstruct 'gold-standard' CCSD(T) energies on MD17. On out-of-distribution conformations of 3BPA, GG-XC reduces the relative MAE by $35\,\%$ to $50\,\%$. Remarkably, GG-XC excels in data efficiency and molecular size extrapolation on QM9, matching force fields trained on 5 times more and larger molecules. On identical training sets, GG-XC yields on average $51\,\%$ lower MAEs.

1 INTRODUCTION

Kohn-Sham Density Functional Theory (KS-DFT) is the backbone of computational material and drug discovery (Jones, 2015). It is a quantum mechanical method to approximate the ground state energy of an N_{el}-electron system by finding the electron density $\rho \in \mathcal{D}_{N_{\text{el}}} = \left\{\rho : \mathbb{R}^3 \to \mathbb{R}_+ | \int_{\mathbb{R}^3} \rho(r) dr = N_{\text{el}}\right\}$ that minimizes the energy functional $E : \mathcal{D}_{N_{\text{el}}} \to \mathbb{R}$. This functional maps electron densities to energies and is composed of

$$E[\rho] = T[\rho] + V_{\text{ext}}[\rho] + V_{\text{H}}[\rho] + E_{\text{XC}}[\rho] \tag{1}$$

$T : \mathcal{D}_{N_{\text{el}}} \to \mathbb{R}$ is the kinetic energy functional, $V_{\text{ext}} : \mathcal{D}_{N_{\text{el}}} \to \mathbb{R}$ the external potential due to positively charged nuclei, $V_{\text{H}} : \mathcal{D}_{N_{\text{el}}} \to \mathbb{R}_+$ the Coulomb energy between electrons, and, finally, $E_{\text{XC}} : \mathcal{D}_{N_{\text{el}}} \to \mathbb{R}_-$ the *exchange-correlation* (XC) functional (Cramer, 2004). While T, V_{ext} and V_{H} permit analytical computations, the exact form of E_{XC} remains unknown, and its approximation frequently dominates DFT's error (Kim et al., 2013).

Machine learning has emerged as a promising data-driven approach to approximate E_{XC} (Kulik et al., 2022; Zhang et al., 2023). Many such ML functionals adopt the classical approach and define E_{XC} as the integral of a learnable XC energy density $\epsilon_{\text{XC}} : \mathbb{R}^{d_{\text{mGGA}}} \to \mathbb{R}$ (Perdew, 2001):

$$E_{\text{XC}}[\rho] = \int_{\mathbb{R}^3} \rho(r) \epsilon_{\text{XC}}\left(\boldsymbol{g}(r)\right) dr \tag{2}$$

where $\boldsymbol{g} : \mathbb{R}^3 \to \mathbb{R}^d$ are properties of the electron density ρ (Dick & Fernandez-Serra, 2021; Nagai et al., 2022). If $\boldsymbol{g}$ only depends on density quantities from its infinitesimal neighborhood, e.g., $\rho(r), \nabla\rho(r), \ldots$, the resulting functionals are called *semi-local*. While this integrates well into existing quantum chemistry code, long-range interactions exceed the functional class (Kaplan et al., 2023). In contrast, existing *non-local* functionals either scale poorly computationally (Zhou et al., 2019) or require costly reference data (Margraf & Reuter, 2021; Bystrom & Kozinsky, 2022). The critical challenge in *non-local* XC functionals lies in efficiently capturing long-range interactions.

To this end, we present Global Graph Exchange Correlation (GG-XC) transferring the success of equivariant graph neural networks (GNNs) (Batzner et al., 2022) to learning the non-local XC

functional[1]. In GG-XC, we compress the electron density to equivariant point embeddings via a convolution of the electron density with an equivariant kernel at the nuclear positions. On this point cloud, the equivariant GNN efficiently captures long-range information. Given the final embeddings, we define a non-local feature density $\boldsymbol{g}_{\text{NL}} : \mathbb{R}^3 \to \mathbb{R}^d$ that we feed into the XC energy density ϵ_{XC} from Equation 2. Compared to previous approaches, our finite point cloud embeddings are neither dependent on the nuclear charge nor the basis set but purely derived from the density ρ. To train on energies alone, we differentiate through the minimization of Equation 1 (Li et al., 2021).

In our experimental evaluation, GG-XC improves upon the learnable semi-local XC-functional on molecular dynamic trajectories, extrapolation to both out-of-distribution conformations and increasingly larger molecules. In particular, we find GG-XC to reduce errors of the semi-local functional by a factor of 2 to 3, on par with accurate ML force fields combined with DFT calculations (Δ-ML). In extrapolation to unseen structures, we find GG-XC to accurately reproduce out-of-distribution potential energy surfaces unlike force fields (incl. Δ-ML) while reducing the MAE by $35\,\%$ to $50\,\%$ compared to the next-best tested method. Finally, we find GG-XC to have an excellent data efficiency, achieving similar accuracies to the best force fields with 5 times less data. To summarize, we demonstrate that GNN-driven functionals push the frontier of non-local XC functionals and provide a promising path towards accurate and scalable DFT calculations.

2 Background

Notation. To denote functionals, i.e., functions mapping from functions to scalars, we write $F[f]$ where F is the functional and f the function. We use superscripts in brackets$^{(t)}$ to indicate sequences and regular superscripts of bold vectors, e.g., $\boldsymbol{x}^l$, to indicate the l-th irreducible representation of the SO(3) group. We generally use $r \in \mathbb{R}^3$ to denote points in the 3D Euclidean space. A notable exception is the finite set of nuclei positions $\{R_1, \ldots, R_{N_{\text{nuc}}}\}$ for which we will use capital letters. The norm of a vector is given by $\|r\| = \|r\|_2$. For directions, i.e., unit length vectors, we use $\widehat{r} = \frac{r}{\|r\|}$.

Kohn-Sham density functional theory is the foundation of our work. Here, we provide a very brief introduction. For more details, we refer the reader to Appendix A or Lehtola et al. (2020). In KS-DFT, the electron density ρ is represented by a set of orthogonal orbitals $\phi_i : \mathbb{R}^3 \to \mathbb{R}$, $\phi_i(x) = C_i^T \boldsymbol{\chi}(x)$, that are defined as linear combinations of a basis set of atomic orbitals $\chi_\mu : \mathbb{R}^3 \to \mathbb{R}$:

$$\rho(r) = \boldsymbol{\phi}(r)^T \boldsymbol{\phi}(r) = \boldsymbol{\chi}(r)^T C C^T \boldsymbol{\chi}(r) = \boldsymbol{\chi}(r)^T P \boldsymbol{\chi}(r) \tag{3}$$

where $P = CC^T \in \mathbb{R}^{N_{\text{bas}} \times N_{\text{bas}}}$ is the so-called density matrix. Given the representation in a finite basis set, one can compute the kinetic energy T, external potential V_{ext}, and electron-electron repulsion energies V_{H} analytically. Unfortunately, such analytical expressions are generally unavailable for $E_{\text{XC}}[\rho]$ as in Equation 2. Thus, one relies on numerical integration (Lehtola et al., 2020).

To minimize Equation 1, one typically uses the self-consistent field (SCF) method that we outline in greater detail in Appendix B. The SCF method is an iterative two-step optimization where one first computes the so-called Fock matrix $F = \frac{\partial E}{\partial P}$ based on the current coefficients C. Then, the optimal coefficients $C \in \mathbb{R}^{N_{\text{bas}} \times N_{\text{el}}}$ are obtained by solving the generalized eigenvalue problem

$$FC = SCE \tag{4}$$

with $S_{\mu\nu} = \int_{\mathbb{R}^3} \chi_\mu(r)\chi_\nu(r)dr$ being the overlap matrix of the atomic orbitals and E being a diagonal matrix. The procedure is repeated until convergence of P.

Equivariance allows defining symmetries of functions. A function $f : \mathcal{X} \to \mathcal{Y}$ is equivariant to a group $\mathcal{G}$ iff $\forall g \in \mathcal{G}, x \in \mathcal{X}, f\left(G_g^{\mathcal{X}} x\right) = G_g^{\mathcal{Y}} f(x)$ where $G_g^{\mathcal{X}}, G_g^{\mathcal{Y}}$ are the representations of g in the domain $\mathcal{X}$ and codomain $\mathcal{Y}$, respectively. Invariance is a special case of equivariance where $G_g^{\mathcal{Y}} = 1$, i.e., $f\left(G_g^{\mathcal{X}} x\right) = f(x)$, for all $g \in \mathcal{G}, x \in \mathcal{X}$.

The real spherical harmonics $Y^l : \mathbb{R}^3 \to \mathbb{R}^{2l+1}, l \in \mathbb{N}_+$, are a well-known example of SO(3)-equivariant functions as they transform under rotation according to the Wigner D matrices $D_R^l \in \mathbb{R}^{(2l+1)\times(2l+1)}, \forall R \in SO(3)$, with $D_R^0 = 1$ being the identity and D_R^1 being 3D-rotation matrices:

$$Y^l(D_R^1 x) = D_R^l Y^l(x). \tag{5}$$

[1]We provide the source code for review on https://figshare.com/s/656a3c16b0b1e881ff95

Such $2l+1$-dimensional equivariant functions $h_i : \mathbb{R}^3 \to \mathbb{R}^{2l+1}$, e.g., $h_i = Y^l$, can be recombined using the tensor product

$$(h_1 \otimes h_2)^{l_o}_{m_o} = \sum_{l_1,l_2} \sum_{m_1,m_2} C^{l_o,l_1,l_2}_{m_o,m_1,m_2} h^{l_1}_{m_1} h^{l_2}_{m_2}, \tag{6}$$

where $C \in \mathbb{R}^{(2l_o+1)\times(2l_i+1)\times(2l_j+1)}$ are the Clebsch-Gordan coefficients. As the Wigner D-matrices are orthogonal, it follows that the inner product of two l-equivariant functions yields an invariant one

$$h^l_i \left(D^1_R r\right)^T h^l_j \left(D^1_R r\right) = h^l_i(r)^T D^{lT}_R D^l_R h^l_j(r) = h^l_i(r)^T h^l_j(r). \tag{7}$$

While, we are interested in $E(3)$-invariant functions, using such equivariant intermediate representations improves accuracy, expands the function class, and improves data efficiency (Schütt et al., 2017; Gasteiger et al., 2021; Batatia et al., 2022).

3 Related work

Learnable density functional approximations (DFAs) of the XC functional have a long history. In their seminal paper, Kohn & Sham (1965) proposed an ϵ_{XC} fitted to a reference calculation. In this first parameterization ϵ_{XC} only depends on $\boldsymbol{g}(r) = \rho(r)$. To improve accuracy, the density features $\boldsymbol{g}$ have been successively refined by including gradients and physical quantities such as the kinetic energy density $\tau(r) = \frac{1}{2}\sum_i^{N_{\text{el}}} |\nabla\phi_i(r)|^2$. The resulting *semi-local* functionals are also known as meta-GGAs with $\boldsymbol{g}(r) = \boldsymbol{g}_{\text{mGGA}}(r) := [\rho(r), |\nabla\rho(r)|, \tau(r)]$ (Perdew, 2001). Many ML functionals use this parameterization with ϵ_{XC} frequently being a product of an MLP with a classical functional (Nagai et al., 2022; Kulik et al., 2022; Zhang et al., 2023). An appealing aspect of learning semi-local energy densities rather than $E_{\text{XC}}[\rho]$ directly is the known implementation of physical constraints (Sun et al., 2015; Kaplan et al., 2023), which inspired works by Dick & Fernandez-Serra (2021); Nagai et al. (2022). While augmenting $\boldsymbol{g}_{\text{mGGA}}$ with hand-crafted non-local features improves the accuracy, this comes at the cost of exact constraints (Nagai et al., 2020; Kirkpatrick et al., 2021). Similarly, such physical biasses are absent in kernel methods Margraf & Reuter (2021); Bystrom & Kozinsky (2022). Additionally, these methods require reference densities that are rarely available (Szabo & Ostlund, 2012). Alternative grid-based CNNs converge slowly in the number of grid points compared to specialized spherical integration grids (Zhou et al., 2019; Treutler & Ahlrichs, 1995). Lastly, machine-learned functionals have been used in other DFT contexts as well, e.g., for energy corrections on SCF-converged energies (Chen & Yang, 2024) or to model the kinetic energy functional in orbital-free DFT (Snyder et al., 2012; Mi et al., 2023; Zhang et al., 2024). Here, we propose to extend semi-local models with expressive fully learnable non-local features obtained from standard integration grids. This allows us to leverage the physical biases of meta-GGAs while allowing efficient non-local interactions with fast integration. Although this work focuses on the XC functional, the methods we present can be transferred to these applications as well.

Machine learning force fields have long functioned as a cheap but inaccurate alternative to quantum mechanical calculations by directly approximating the potential energy surface from data without solving the electronic structure problem (Unke et al., 2021). Contemporary force fields generally rely on graph neural networks (GNNs) (Schütt et al., 2017) representing molecules in terms of graphs with atoms as nodes. The advent of SO(3)-equivariant models led to significant improvements in accuracy and data efficiency, closing the gap to full-fidelity quantum mechanical calcualtions (Batzner et al., 2022; Batatia et al., 2022). While they effectively aim to accomplish the same goal as DFT at a fraction of the cost, their out-of-distribution accuracy is a common problem (Stocker et al., 2022). In GG-XC, we transfer the success of equivariant GNNs to DFT, by combining their approximation power with the physical nature of DFT, resulting in an accurate and data-efficient method.

4 Global Graph Exchange Correlation

With Global Graph Exchange Correlation (GG-XC), we propose an efficient *non-local* approximation to the unknown exchange-correlation functional E_{XC} that maps electron densities (positive integrable functions in $\mathbb{R}^3$) to scalar energies. To this end, GG-XC consists of four components as illustrated in Figure 1: (1) *Nuclei-centered equivariant embeddings* $\mathbf{H}^{(0)}$ compress the electron density ρ into

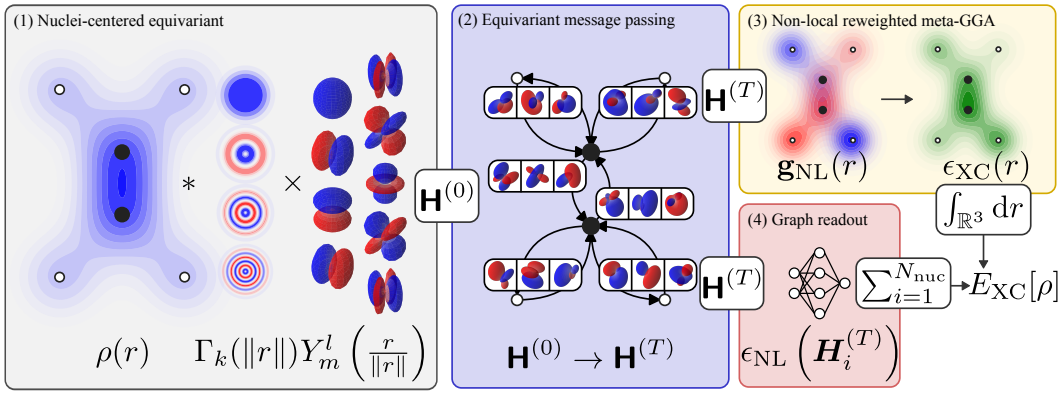

Figure 1: Illustration of Global Graph Exchange Correlation (GG-XC)'s four components: (1) We first compress the electron density into a finite point cloud representation $\mathbf{H}^{(0)}$ via a convolution with radial filters $\Gamma_k : \mathbb{R}_+ \to \mathbb{R}$ and spherical harmonics $Y_m^l : \mathbb{R}^3 \to \mathbb{R}$ evaluated at the nuclear position. (2) The embeddings $\mathbf{H}^{(0)}$ are updated using equivariant message passing to obtain long-range effects in $\mathbf{H}^{(T)}$. (3) We define a non-local feature density $\boldsymbol{g}_{\mathrm{NL}} : \mathbb{R}^3 \to \mathbb{R}^d$ from which we derive the exchange correlation density $\epsilon_{\mathrm{XC}} : \mathbb{R}^3 \to \mathbb{R}$. (4) We add a graph readout of the final embeddings to learn additional corrections. To obtain $E_{\mathrm{XC}}[\rho]$, we integrate ϵ_{XC} and add it to the global graph readout.

a finite point cloud representation by equivariantly integrating the density around the nuclei. (2) *Equivariant message passing* on this point cloud includes long-range information in $\mathbf{H}^{(T)}$. Based on our final point cloud embeddings $\mathbf{H}^{(T)}$, we define E_{XC} as the sum of two readouts: (3) A *non-local reweighted meta-GGA* based on our non-local feature density $\boldsymbol{g}_{\mathrm{NL}} : \mathbb{R}^3 \to \mathbb{R}^d$ and (4) a *graph readout* of the invariant nuclei-centered features $\mathbf{H}^{(T)0}$:

$$E_{\mathrm{XC}}[\rho] = \underbrace{\int_{\mathbb{R}^3} \rho(r) \underbrace{\gamma_{\mathrm{NL}}\left(\boldsymbol{g}_{\mathrm{NL}}(r), \boldsymbol{g}_{\mathrm{mGGA}}(r)\right)}_{\text{non-local weights}} \underbrace{\epsilon_{\mathrm{mGGA}}\left(\boldsymbol{g}_{\mathrm{mGGA}}(r)\right)}_{\text{meta-GGA}} dr}_{\text{non-local reweighted meta-GGA}} + \underbrace{\sum_{i=1}^{N_{\mathrm{nuc}}} \epsilon_{\mathrm{NL}}\left(\boldsymbol{H}_i^{(T)0}\right)}_{\text{graph readout}}. \tag{8}$$

(1) Nuclei-centered equivariant embeddings $\mathbf{H}^{(0)} \in \left[\mathbb{R}^1 \times \ldots \times \mathbb{R}^{2l_{\max}+1}\right]^{N_{\mathrm{nuc}} \times d} := \mathcal{F}_{l_{\max}}^{N_{\mathrm{nuc}} \times d}$ reduce the computational scaling by mapping the continuous density to a finite per-nucleus representation. Further, these enable using equivariant GNNs to learn long-range interactions efficiently. The positions of the nuclei $R_i \in \mathbb{R}^3$ naturally lend themselves as centroids for such embeddings as they represent peaks of the electron density (Kato, 1957). To perform this reduction, we evaluate a convolution of the electron density ρ with equivariant filters $S_k^l : \mathbb{R}^3 \to \mathbb{R}^{2l+1}$ at R_i:

$$H_{ik}^{(0)l} = \left(\rho_i * S_k^l\right)(R_i) \tag{9}$$

where $\rho_i : \mathbb{R}^3 \to \mathbb{R}_+$ is the soft-partitioned electron density associated with the i-th embedding

$$\rho_i(r) = \rho(r) \frac{\alpha_i(r)}{\sum_{j=1}^{N_{\mathrm{nuc}}} \alpha_j(r)}, \tag{10}$$

$$\alpha_i(r) = \exp\left(-\frac{\|r - R_i\|^2}{\lambda^2}\right) \tag{11}$$

with $\lambda \in \mathbb{R}_+$ being a free parameter. Such partitioning prevents overcounting if nuclei are close similar to quadrature methods (Becke, 1988). Like equivariant GNNs (Batzner et al., 2022), we define the filters S_k^l as a product of spherical harmonics $Y^l : \mathbb{R}^3 \to \mathbb{R}^{2l+1}$ and radial filters $\Gamma_k : \mathbb{R}_+ \to \mathbb{R}$:

$$S_k^l(r) = \Gamma_k(\|r\|) Y^l(\widehat{r}) \tag{12}$$

where $0 \leq l \leq l_{\max}$ indexes the real spherical harmonics. The spherical harmonics allow us to capture angular changes in the electron density that would otherwise be lost in a purely radial representation.

Put together, we compute the nuclei-centered equivariant embeddings

$$\begin{aligned} H_{ik}^{(0)l} &= \int_{\mathbb{R}^3} \rho_i(r)\Gamma_k^l(\|r-R_i\|)Y^l\left(\widehat{r-R_i}\right)dr \\ &\approx \sum_{j=1}^{N_{\text{quad}}} w_j\rho(r_j)\alpha_i(r_j)\Gamma_k(\|r_j-R_i\|)Y^l\left(\widehat{r_j-R_i}\right) \end{aligned} \tag{13}$$

with a set of N_{quad} standard integration points and weights $\{(r_j, w_j)\}_{j=1}^{N_{\text{quad}}} \in \left[\mathbb{R}^3 \times \mathbb{R}\right]^{N_{\text{quad}}}$ (Treutler & Ahlrichs, 1995). Importantly, while these embeddings are centered at the nuclei, they do not embed the nuclear charges, as in ML force fields (Schütt et al., 2017), but the electron density around them. This is important as the derivative with respect to the electron density affects the SCF procedure; see Appendix B. A nuclear charge embedding's derivative would not exist and, thus, not alter the DFT calculation, effectively yielding a force field.

We follow Schütt et al. (2021) and parametrize the radial filters Γ_k as a combination of $\sin, \cos$ with a fixed polynomial envelope function $u : \mathbb{R}_+ \to \mathbb{R}$ from Gasteiger et al. (2022) for a cutoff $c \in \mathbb{R}_+$:

$$\Gamma_k(r) = \begin{cases} \sin\left(\frac{r\pi k}{2c}\right) u\left(\frac{r}{c}\right) & \text{if } k \text{ even,} \\ \cos\left(\frac{r\pi(k+1)}{2c}\right) u\left(\frac{r}{c}\right) & \text{if } k \text{ odd.} \end{cases} \tag{14}$$

We found such frequency filters to yield more stable results than Bessel functions as the latter ones are very sensitive close to the centroids (Gasteiger et al., 2022). This sensitivity might be beneficial in force fields but is implicitly given in DFT due to the higher density close to the nuclei.

(2) Equivariant message passing allows for the propagation and updating of the equivariant electron density features $\mathbf{H}^{(0)}$ to capture long-range dependencies within the electron density like van-der-Waals forces (Frank et al., 2022). The equivariance of these features is important as it allows us to define a non-radial-symmetric feature density $\boldsymbol{g}_{\text{NL}}$ in the next step. We use Batzner et al. (2022)'s NequIP to perform message passing with the SO(3)-equivariant convolution from Thomas et al. (2018) to iteratively update the embeddings $\mathbf{H}^{(t)}$ over T steps:

$$\boldsymbol{H}_i^{(t+1)} = \text{EquiMLP}\left(\text{EquiLin}\left(\boldsymbol{H}_i^{(t)}\right) + \text{EquiConv}\left(\mathbf{H}^{(t)}, \boldsymbol{R}\right)_i\right) + \text{EquiLin}\left(\boldsymbol{H}_i^{(t)}\right), \tag{15}$$

where $\text{EquiLin} : \mathcal{F}_{l_{\text{max}}}^d \to \mathcal{F}_{l_{\text{max}}}^d$ is an equivariant dense layer that mixes features of the same order l

$$\text{EquiLin}(\boldsymbol{X})_k = \text{Concat}\left(\left[\sum_{k'=1}^{d} W_{k'k}^l X_{k'}^l\right]_{l=0}^{l_{\text{max}}}\right). \tag{16}$$

with each $\boldsymbol{W}^l \in \mathbb{R}^{d\times d}$ being a learnable weight matrix. $\text{EquiMLP} : \mathcal{F}_{l_{\text{max}}}^d \to \mathcal{F}_{l_{\text{max}}}^d$ is the equivariant equivalent of a standard MLP, where we use the invariant embeddings $l = 0$ and an activation function $\sigma : \mathbb{R} \to \mathbb{R}$ to gate the equivariant parts $l > 0$:

$$\text{EquiMLP}(\boldsymbol{X}) = \text{EquiLin}(\text{ActEquiLin}(\boldsymbol{X})), \tag{17}$$

$$\text{ActEquiLin}(\boldsymbol{X}) = \text{Concat}\left(\sigma(W^0\boldsymbol{X}^0), \left[\text{EquiLin}(\boldsymbol{X})^l \circ \sigma\left(W^l\boldsymbol{X}^0\right)\right]_{l=1}^{l_{\text{max}}}\right). \tag{18}$$

Lastly, we define the convolution $\text{EquiConv} : \mathcal{F}_{l_{\text{max}}}^{N_{\text{nuc}}\times d} \times \mathbb{R}^{N_{\text{nuc}}\times 3} \to \mathcal{F}_{l_{\text{max}}}^{N_{\text{nuc}}\times d}$ via the tensor product:

$$\begin{aligned} \text{EquiConv}(\boldsymbol{X}, \mathbf{R})_t^{l_o} = \sum_{s=1}^{N_{\text{nuc}}} \sum_{l_i, l_f=0}^{l_{\text{max}}} &\text{MLP}\left(\Gamma(\|R_s - R_t\|)\right)_{l_o, l_i, l_f} \\ &\cdot \left(\text{EquiMLP}(\boldsymbol{x}_s)^{l_i} \otimes Y^{l_f}\left(\widehat{R_s - R_t}\right)\right)^{l_0}. \end{aligned} \tag{19}$$

To accelerate the computation of the tensor product, we use Passaro & Zitnick (2023)'s efficient equivariant convolutions. For more details, we refer the reader to Batzner et al. (2022).

(3) Non-local reweighted meta-GGA. As the majority of the XC energy can be captured by semi-local functionals (Goerigk et al., 2017), starting with a machine-learned meta-GGA $\epsilon_{\text{mGGA}} : \mathbb{R}^{d_{\text{mGGA}}} \to$

$\mathbb{R}$ accounts for most of the XC energy. To correct the meta-GGA for non-local effects, we first define a non-local feature density $\boldsymbol{g}_{\text{NL}} : \mathbb{R}^3 \to \mathbb{R}^d$ based on the non-local embeddings $\left\{\mathbf{H}^{(t)}\right\}_{t=0}^{T}$:

$$\boldsymbol{g}_{\text{NL}}(r)_k = \sum_{i=1}^{N_{\text{nuc}}} \alpha_i(r) \sum_{t=0}^{T} \sum_{l=0}^{l_{\text{max}}} \underbrace{Y^l\left(\widehat{r - R_i}\right)^T H_{ik}^{(t)l}}_{\text{angular}} \underbrace{\Gamma(\|r - R_i\|)^T \boldsymbol{w}_k^{(t)l}\left(\boldsymbol{H}_i^{(t)0}\right)}_{\text{radial}} \tag{20}$$

where $\boldsymbol{w}_k^{(t)l} : \mathbb{R}^d \to \mathbb{R}^d$ are MLPs mapping to radial weights. While the inner product between the spherical harmonics Y^l and our equivariant embeddings $\mathbf{H}^{(t)l}$ expresses angular changes, the inner product of radial basis functions Γ and $\boldsymbol{w}_k^{(t)l}$ allows for radial changes. From this feature density $\boldsymbol{g}_{\text{NL}}$ and standard meta-GGA inputs $\boldsymbol{g}_{\text{mGGA}} : \mathbb{R}^3 \to \mathbb{R}^{d_{\text{mGGA}}}$, we derive the non-local correction $\gamma_{\text{NL}} : \mathbb{R}^{d+d_{\text{mGGA}}} \to \mathbb{R}$ to the exchange energy density ϵ_{mGGA}:

$$\epsilon_{\text{XC}}(r) = \gamma_{\text{NL}}\left(\boldsymbol{g}_{\text{NL}}(r), \boldsymbol{g}_{\text{mGGA}}(r)\right) \cdot \epsilon_{\text{mGGA}}\left(\boldsymbol{g}_{\text{mGGA}}(r)\right). \tag{21}$$

In practice, we implement γ_{NL} as an MLP. To obtain the final readout, we integrate the exchange-correlation density over the electron density $\int_{\mathbb{R}^3} \epsilon_{\text{XC}}(r)\rho(r)dr$. Like Equation 13, we evaluate the integral with standard integration grids.

(4) Graph readout. In addition to the meta-GGA-based readout, we add global graph readout of the embeddings $\left\{\mathbf{H}^{(t)}\right\}_{t=0}^{T}$ to capture the remaining non-local effects. We use an MLP $\epsilon_{\text{NL}} : \mathbb{R}^d \to \mathbb{R}$ on top of the invariant embeddings ($l = 0$) to obtain the final exchange-correlation energy:

$$E_{\text{XC}}[\rho] = \int_{\mathbb{R}^3} \epsilon_{\text{XC}}(r)\rho(r)dr + \sum_{t=0}^{T}\sum_{i=1}^{M} \epsilon_{\text{NL}}\left(\boldsymbol{H}_i^{(t)0}\right). \tag{22}$$

Limitations. While GG-XC demonstrates significant improvements over a semi-local ML functional, s. Section 5, it is not free of limitations. First, as we rely on the nuclear positions to represent the electronic density, it is not truly universal, i.e., independent of the external potential V_{ext} (Kohn & Sham, 1965). Second, the non-local nature of our functional permits no known way to enforce most physical constraints (Kaplan et al., 2023). These missing constraints enable the correction of basis set errors through the XC functional. While this may lead to unphysical matches between densities and energies, our experiments suggest that this does not lead to overfitting energies. Third, systems without nuclei, e.g., the homogenous electron gas, cannot be modeled with our approach. In such cases, one may want to replace the real-space point cloud with a frequency representation (Kosmala et al., 2023). Lastly, while more data efficient and better in extrapolation, running KS-DFT is more expensive than a surrogate force field.

5 Experiments

In the following, we compare GG-XC to various alternative methods of learning potential energy surfaces across several settings. In particular, we focus on the following tasks: interpolating accurate energy surfaces, extrapolation to unseen conformations, and extrapolation to larger molecules.

Methods. To accurately position GG-XC, we compare to methods of varying computational costs. In particular, we evaluate force fields, Δ-ML, i.e., combining force fields with KS-DFT calculations, and a learnable XC-functional (Dick & Fernandez-Serra, 2021). While force fields are orders of magnitude cheaper as they bypass quantum mechanical calculations altogether, they lack prior physical knowledge. For the Δ-ML methods, we shift all energies by DFT energies with LDA in the STO-6G basis. This reduces the learning problem to the difference between the KS-DFT and the target energy (Wengert et al., 2021). As force fields, we test increasingly expressive models based on their use of SO(3) irreducible representations: SchNet ($l = 0$) (Schütt et al., 2017), PaiNN ($l = 1$) (Schütt et al., 2021), and NequIP ($l = 2$) (Batzner et al., 2022). Finally, we compare with the same learnable semi-local XC-functional (Dick & Fernandez-Serra, 2021) used in GG-XC. As Δ-ML methods and learnable XC-functionals require a DFT computation for each structure, they are at the same computational cost as GG-XC.

Table 1: Test set MAE on the CCSD(T) MD17 dataset in $\mathrm{m}E_\mathrm{h}$. (**best**, second)

	Force field			Δ-ML			KS-DFT	
Molecule	SchNet	PaiNN	NequIP	SchNet	PaiNN	NequIP	Dick	GG-XC
Aspirin	7.01	2.82	5.52	2.02	1.20	1.04	1.94	**0.69**
Benzene	0.40	0.16	0.09	0.13	0.11	**0.02**	0.39	0.10
Ethanol	1.41	0.89	0.95	0.93	0.42	0.25	0.85	**0.21**
Malonaldehyde	2.10	1.00	2.32	0.61	0.44	0.29	0.73	**0.27**
Toluene	1.80	1.10	1.87	0.44	0.31	**0.13**	0.38	0.20

Table 2: Structural extrapolation on the 3BPA dataset. All methods were trained on the 300K training set. All numbers are relative MAE in $\mathrm{m}E_\mathrm{h}$. The last three rows refer to the potential energy surfaces.

	Force field			Δ-ML			KS-DFT	
Test set	SchNet	PaiNN	NequIP	SchNet	PaiNN	NequIP	Dick	GG-XC
300K	5.15	2.91	3.81	2.38	1.14	0.81	0.96	**0.42**
600K	9.06	5.81	7.55	3.96	2.13	1.56	1.36	**0.73**
1200K	18.33	14.14	17.30	6.84	5.97	3.30	2.27	**1.39**
$\beta = 120°$	3.84	1.78	2.25	2.53	1.25	1.09	0.75	**0.35**
$\beta = 150°$	4.64	1.89	2.64	2.03	0.84	0.88	0.61	**0.23**
$\beta = 180°$	4.97	1.92	3.03	1.79	1.06	0.73	0.56	**0.20**

Setup. To train the XC functionals, we follow Li et al. (2021) and implement the SCF method differentiably. This allows us to match the converged SCF energies directly to the target energies without needing ground truth electron densities. We provide implementation details in Appendix C. All methods are trained on energy labels only. Force fields are trained with hyperparameters from their respective works with modifications to learning rate, batch size, and initialization to improve performance; we outline the changes in Appendix G. We list GG-XC's hyperparameters in Appendix F.

Reproducing gold-standard accuracies. The objective behind learning XC-functionals or ML force fields is to facilitate access to accurate potential energy surfaces by distilling highly accurate reference data into a faster method. Hence, we compare these methods on the revised MD17 dataset, which contains precise 'gold-standard' CCSD(T) (CCSD for aspirin) reference energies for conformations of five molecules along the trajectory of a molecular dynamic (MD) simulation (Chmiela et al., 2018). Each molecule has a training set of 1000 structures, which we split into 950 training and 50 validation structures. Each test set contains an additional 500 structures (1000 for ethanol). Following Schütt et al. (2017), a separate model is fitted per molecule. Since CCSD(T) calculations inherently account for non-local effects, these datasets are well suited for investigating the ability to accurately interpolate multi-dimensional energy surfaces from a limited number of reference structures. In Appendix I, we provide additional Δ-ML data with a more accurate DFT functional and basis sets.

Table 1 lists the mean absolute error (MAE) for all methods on the MD17 test sets. We find that force fields generally struggle to reconstruct the energy surface of the MD17 dataset accurately when trained solely on energies. In contrast, Δ-ML methods and learnable XC-functionals generally achieve chemical accuracy at $1\,\mathrm{kcal\,mol}^{-1} \approx 1.6\,\mathrm{m}E_\mathrm{h}$. The DFT reference calculations systematically reduce the to-be-learned energy surface delta from $6.2\,\mathrm{m}E_\mathrm{h}$ to $4.9\,\mathrm{m}E_\mathrm{h}$. Compared to the GG-XC's base semi-local Dick & Fernandez-Serra (2021), GG-XC reduces errors by a factor of 2 to 4. Among all methods, GG-XC yields the lowest error on 3 of the 5 molecules. In Appendix H, we test the performance on a reduced training set of just 50 samples.

Extrapolating structures. In MD simulations, one often encounters structures far outside the training set, e.g., due to higher temperatures or environmental changes (Stocker et al., 2022). To investigate the extrapolation to unseen structures, we use the 3BPA dataset (Kovács et al., 2021). 3BPA contains various geometric configurations of the molecule 3-(benzyloxy)pyridin-2-amine. The training set consists of 500 structures sampled from an MD simulation at room temperature (300K). The test sets consist of MD trajectories at 300K, 600K, and 1200K to test in and out-of-distribution (OOD)

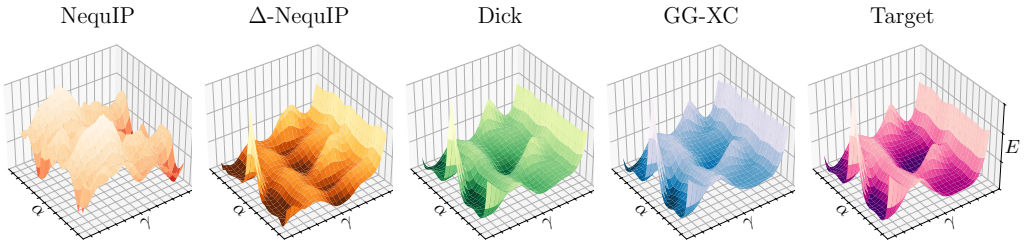

Figure 2: Two-dimensional slice of the potential energy surface of 3BPA with the dihedral $\beta = 120°$. Pure force fields like NequIP struggle to recover the shape of this out-of-distribution energy surface. When paired with DFT calculations, one can see that the energy surface moves closer to the target shape but introduces additional extrema. Learnable XC functionals like Dick & Fernandez-Serra (2021) and GG-XC demonstrate significantly better reproduction of the target energy surface.

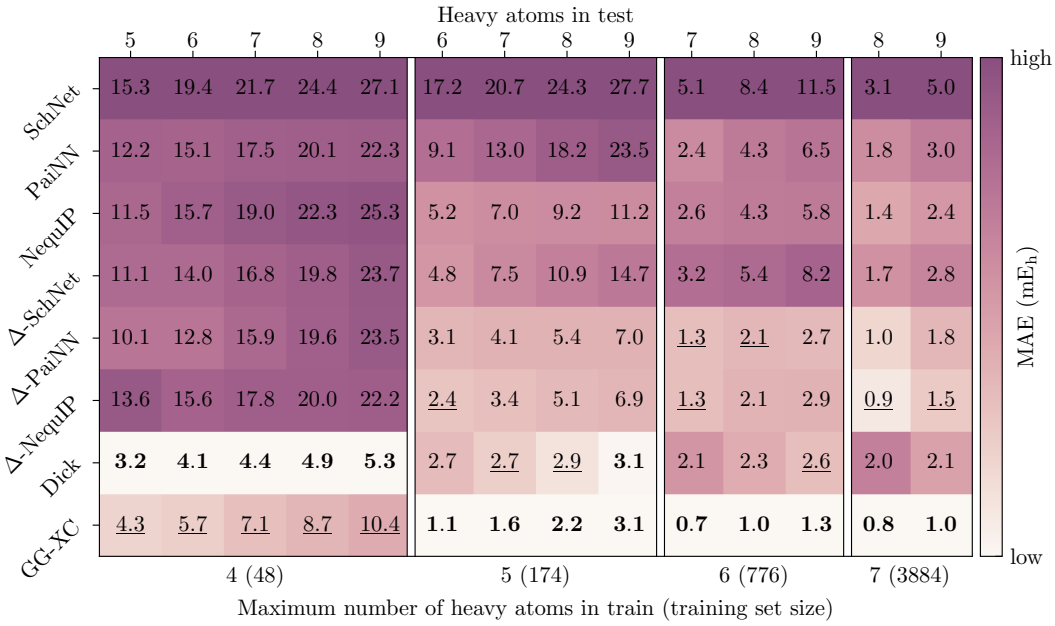

Figure 3: MAE in $\mathrm{m}E_\mathrm{h}$ on QM9 size extrapolation. Each row represents a different method. The four groups indicate the maximum number of heavy atoms in the training set and its size. Each column represents the test subset with the number of heavy atoms listed above.

performance. Additionally, the dataset contains three 2-dimensional potential energy surface slices where two dihedral angles are varied while keeping one fixed. The labels have been computed with the hybrid ωB97X XC functional and the 6-31G(d) basis set and, thus, include non-local interactions.

As constant offsets of the potential energy surface do not affect the system dynamics, we evaluate the relative MAE, i.e., $\mathbb{E}\left[|(E_\text{pred} - E_\text{target} - \text{median}(E_\text{pred} - E_\text{target})|\right]$. We list the relative MAE across all methods and test sets in Table 2; for absolute MAE numbers, we refer the reader to Appendix J. Across all test sets, GG-XC results in $35\,\%$ to $51\,\%$ lower errors than the next best-tested force field, Δ-ML, or XC functionals. On the far OOD 1200K samples, GG-XC is the only method achieving chemical accuracy $1.6\,\mathrm{m}E_\mathrm{h}$ at an relative MAE of $1.40\,\mathrm{m}E_\mathrm{h}$. To illustrate the qualitative improvement of GG-XC's energy surfaces, we plot the most OOD potential energy slice at $\beta = 120°$ in Figure 2 for NequIP, Δ-NequIP, Dick & Fernandez-Serra (2021), GG-XC and the target. The remaining methods and corresponding energy surfaces are plotted in Appendix K. It is evident that force fields fail to reproduce the target energy surface with no resemblance to the target surface. While the reference DFT calculations move Δ-NequIP closer to the target surface, the energy shape includes additional extrema not present in the target surface. In contrast, both XC functionals accurately reproduce the energy surface.

Extrapolation to larger molecules. Gathering reference data for large compounds is costly and expensive with accurate quantum chemical calculations like CCSD(T) scaling $O({N_{\text{el}}}^7)$ in the number of electrons N_{el}. Thus, extrapolation from small or medium-sized molecules to larger ones is critical. Here, we simulate this setting by splitting the QM9 dataset (Ramakrishnan et al., 2014) into subsets of increasing size based on the number of heavy atoms, i.e., QM9(S) are all QM9 molecules with at most S heavy atoms. For each $S \in \{4, 5, 6, 7\}$, we train a separate model and test on the remaining structures, i.e., QM9\QM9(S). For each training set, we split the structures 90%/10% into training and validation sets. The QM9 dataset comprises 134k stable organic molecules with up to 9 heavy atoms. The energies have been computed with the hybrid B3LYP XC functional with the 6-31G(2df,p) basis set and, thus, contain non-local interactions through the exact Hartree exchange. As the dataset contains only few molecules with fluorine, force fields (including Δ-ML) could not yield accurate energies if none or few are in the training set. Thus, we omitted all molecules that include fluorine.

We visualize the MAE for each combination of the number of heavy atoms in the test set and the maximum number of heavy atoms in the training set in Figure 3. Across all combinations of training and test sets, we find learnable XC functionals yielding the lowest errors with a preference for Dick & Fernandez-Serra (2021) on the smallest QM9(4). We hypothesize that this is due to the physical constraints enforced by the learnable XC functional, which are not present in the other methods, including GG-XC. Starting with QM9(5), GG-XC's errors are consistently the lowest. Notably, GG-XC trained on QM9(6) yields lower MAE on the largest structures than the best alternative on QM9(7) with 5 times more samples and one-atom larger molecules. On QM9(7), GG-XC is at least $33\,\%$ more accurate than the competing methods on test molecules with 9 heavy atoms. Compared to Dick & Fernandez-Serra (2021), GG-XC reduces the MAE by at least $2\times$ on QM9(6) and QM9(7).

6 Discussion

We tackled the problem of learning efficient non-local exchange-correlation functionals to enhance the accuracy of density functional theory. To this end, we have presented Global Graph Exchange Correlation (GG-XC), an equivariant graph neural network approach. We have introduced a basis-independent reduction of the electron density into a finite point cloud representation, enabling equivariant graph neural networks to capture long-range information. Based on this point cloud embedding, we have parameterized a non-local feature density used for the reweighing of a semi-local exchange-correlation energy density. Unlike other learnable non-local functionals, GG-XC can be trained by differentiating through the SCF solver such that the training requires only energy targets.

In our empirical evaluation, GG-XC improves upon previous machine learning-based methods of modeling potential energy surfaces. On MD17, GG-XC accurately reconstructs 'gold-standard' potential energy surfaces, namely CCSD(T), within the DFT framework. On the 3BPA, GG-XC reduces errors by $35\,\%$ to $50\,\%$ compared to the best-performing baseline. On QM9, GG-XC demonstrates remarkable data efficiency, achieving similar accuracies with 5 times less data or up to $33\,\%$ lower errors at the same amount of data compared to the best baseline. Finally, GG-XC extrapolates well to unseen conformations and larger molecules on 3BPA and QM9, respectively.

Overall, these results strongly underline the data efficiency of learning exchange-correlation functionals compared to machine-learned force fields. We hypothesize that a significant portion of GG-XC's generalization is thanks to including the physically constrained semi-local model that captures most of the exchange-correlation energy and biases GG-XC to approximately fulfill the same constraints. On top of these, our non-local contributions transfer the accuracies of graph neural network-based force fields to functionals while maintaining a similar data efficiency.

Future work. The simple implementation of GG-XC through standard deep learning libraries and the integration with the self-consistent field method open the door to a range of future research. Thanks to the independence on the basis set, GG-XC can transfer to non-atom-centered basis sets like plane waves, as they are common in periodic systems. Further, while we trained GG-XC solely on energies, multimodal training with other DFT-computable observables like electron densities or atomic forces could further improve accuracy and generalization. Another path to improving generalization may be integrating known physical constraints to the non-local part of GG-XC. Thanks to the basis-set independence, one may also attempt to learn the kinetic energy functional in orbital-free DFT. Finally, accurate exchange-correlation functionals like GG-XC may serve as a foundation for bridging the gap between sparse, accurate quantum mechanical calculations and fast force field.

Acknowledgements. We thank Jan Schuchardt and Leo Schwinn for their invaluable feedback on the manuscript and Johannes Margraf for our great discussions on density functional theory. Funded by the Federal Ministry of Education and Research (BMBF) and the Free State of Bavaria under the Excellence Strategy of the Federal Government and the Länder.

REFERENCES

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. *Advances in Neural Information Processing Systems*, 35:11423–11436, December 2022.

Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P. Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess E. Smidt, and Boris Kozinsky. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature Communications*, 13(1):2453, May 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-29939-5.

A. D. Becke. A multicenter numerical integration scheme for polyatomic molecules. *The Journal of Chemical Physics*, 88(4):2547–2553, February 1988. ISSN 0021-9606, 1089-7690. doi: 10.1063/1.454033.

James Bradbury, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, George Necula, Adam Paszke, Jake VanderPlas, Skye Wanderman-Milne, and Qiao Zhang. JAX: Composable transformations of Python+NumPy programs, 2018.

Kyle Bystrom and Boris Kozinsky. CIDER: An Expressive, Nonlocal Feature Set for Machine Learning Density Functionals with Exact Constraints. *Journal of Chemical Theory and Computation*, 18 (4):2180–2192, April 2022. ISSN 1549-9618. doi: 10.1021/acs.jctc.1c00904.

Zehua Chen and Weitao Yang. Development of a machine learning finite-range nonlocal density functional. *The Journal of Chemical Physics*, 160(1):014105, January 2024. ISSN 0021-9606. doi: 10.1063/5.0179149.

Stefan Chmiela, Huziel E. Sauceda, Klaus-Robert Müller, and Alexandre Tkatchenko. Towards exact molecular dynamics simulations with machine-learned force fields. *Nature Communications*, 9(1): 3887, September 2018. ISSN 2041-1723. doi: 10.1038/s41467-018-06169-2.

Christopher J. Cramer. *Essentials of Computational Chemistry: Theories and Models*. Wiley, Chichester, West Sussex, England ; Hoboken, NJ, 2nd ed edition, 2004. ISBN 978-0-470-09182-1 978-0-470-09181-4.

Sebastian Dick and Marivi Fernandez-Serra. Highly accurate and constrained density functional obtained with differentiable programming. *Physical Review B*, 104(16):L161109, October 2021. doi: 10.1103/PhysRevB.104.L161109.

Thorben Frank, Oliver Unke, and Klaus-Robert Müller. So3krates: Equivariant attention for interactions on arbitrary length-scales in molecular systems. *Advances in Neural Information Processing Systems*, 35:29400–29413, 2022.

Johannes Gasteiger, Florian Becker, and Stephan Günnemann. GemNet: Universal Directional Graph Neural Networks for Molecules. In *Advances in Neural Information Processing Systems*, May 2021.

Johannes Gasteiger, Chandan Yeshwanth, and Stephan Günnemann. Directional Message Passing on Molecular Graphs via Synthetic Coordinates, April 2022.

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 1463-9076, 1463-9084. doi: 10.1039/C7CP04913G.

R. O. Jones. Density functional theory: Its origins, rise to prominence, and future. *Reviews of Modern Physics*, 87(3):897–923, August 2015. ISSN 0034-6861, 1539-0756. doi: 10.1103/RevModPhys.87.897.

Aaron D. Kaplan, Mel Levy, and John P. Perdew. The Predictive Power of Exact Constraints and Appropriate Norms in Density Functional Theory. *Annual Review of Physical Chemistry*, 74(1): 193–218, 2023. doi: 10.1146/annurev-physchem-062422-013259.

Tosio Kato. On the eigenfunctions of many-particle systems in quantum mechanics. *Communications on Pure and Applied Mathematics*, 10(2):151–177, 1957. ISSN 1097-0312. doi: 10.1002/cpa.3160100201.

Min-Cheol Kim, Eunji Sim, and Kieron Burke. Understanding and Reducing Errors in Density Functional Calculations. *Physical Review Letters*, 111(7):073003, August 2013. ISSN 0031-9007, 1079-7114. doi: 10.1103/PhysRevLett.111.073003.

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. doi: 10.1126/science.abj6511.

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.

Arthur Kosmala, Johannes Gasteiger, Nicholas Gao, and Stephan Günnemann. Ewald-based Long-Range Message Passing for Molecular Graphs. In *Proceedings of the 40th International Conference on Machine Learning*, pp. 17544–17563. PMLR, July 2023.

Dávid Péter Kovács, Cas van der Oord, Jiri Kucera, Alice E. A. Allen, Daniel J. Cole, Christoph Ortner, and Gábor Csányi. Linear Atomic Cluster Expansion Force Fields for Organic Molecules: Beyond RMSE. *Journal of Chemical Theory and Computation*, 17(12):7696–7711, December 2021. ISSN 1549-9618. doi: 10.1021/acs.jctc.1c00647.

David B. Krisiloff, Caroline M. Krauter, Francis J. Ricci, and Emily A. Carter. Density Fitting and Cholesky Decomposition of the Two-Electron Integrals in Local Multireference Configuration Interaction Theory. *Journal of Chemical Theory and Computation*, 11(11):5242–5251, November 2015. ISSN 1549-9618. doi: 10.1021/acs.jctc.5b00762.

H. J. Kulik, T. Hammerschmidt, J. Schmidt, S. Botti, M. A. L. Marques, M. Boley, M. Scheffler, M. Todorović, P. Rinke, C. Oses, A. Smolyanyuk, S. Curtarolo, A. Tkatchenko, A. P. Bartók, S. Manzhos, M. Ihara, T. Carrington, J. Behler, O. Isayev, M. Veit, A. Grisafi, J. Nigam, M. Ceriotti, K. T. Schütt, J. Westermayr, M. Gastegger, R. J. Maurer, B. Kalita, K. Burke, R. Nagai, R. Akashi, O. Sugino, J. Hermann, F. Noé, S. Pilati, C. Draxl, M. Kuban, S. Rigamonti, M. Scheidgen, M. Esters, D. Hicks, C. Toher, P. V. Balachandran, I. Tamblyn, S. Whitelam, C. Bellinger, and L. M. Ghiringhelli. Roadmap on Machine learning in electronic structure. *Electronic Structure*, 4 (2):023004, August 2022. ISSN 2516-1075. doi: 10.1088/2516-1075/ac572f.

Susi Lehtola, Frank Blockhuys, and Christian Van Alsenoy. An Overview of Self-Consistent Field Calculations Within Finite Basis Sets. *Molecules*, 25(5):1218, January 2020. ISSN 1420-3049. doi: 10.3390/molecules25051218.

Li Li, Stephan Hoyer, Ryan Pederson, Ruoxi Sun, Ekin D. Cubuk, Patrick Riley, and Kieron Burke. Kohn-Sham Equations as Regularizer: Building Prior Knowledge into Machine-Learned Physics. *Physical Review Letters*, 126(3):036401, January 2021. ISSN 0031-9007, 1079-7114. doi: 10.1103/PhysRevLett.126.036401.

Johannes T. Margraf and Karsten Reuter. Pure non-local machine-learned density functional theory for electron correlation. *Nature Communications*, 12(1):344, January 2021. ISSN 2041-1723. doi: 10.1038/s41467-020-20471-y.

Wenhui Mi, Kai Luo, S. B. Trickey, and Michele Pavanello. Orbital-Free Density Functional Theory: An Attractive Electronic Structure Method for Large-Scale First-Principles Simulations. *Chemical Reviews*, 123(21):12039–12104, November 2023. ISSN 0009-2665. doi: 10.1021/acs.chemrev.2c00758.

Ryo Nagai, Ryosuke Akashi, and Osamu Sugino. Completing density functional theory by machine learning hidden messages from molecules. *npj Computational Materials*, 6(1):1–8, May 2020. ISSN 2057-3960. doi: 10.1038/s41524-020-0310-0.

Ryo Nagai, Ryosuke Akashi, and Osamu Sugino. Machine-learning-based exchange correlation functional with physical asymptotic constraints. *Physical Review Research*, 4(1):013106, February 2022. doi: 10.1103/PhysRevResearch.4.013106.

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 *ICML'23*, pp. 27420–27438, Honolulu, Hawaii, USA, July 2023. JMLR.org.

John P. Perdew. Jacob's ladder of density functional approximations for the exchange-correlation energy. In *AIP Conference Proceedings*, volume 577, pp. 1–20, Antwerp (Belgium), 2001. AIP. doi: 10.1063/1.1390175.

P. Pulay. Improved SCF convergence acceleration. *Journal of Computational Chemistry*, 3(4): 556–560, 1982. ISSN 1096-987X. doi: 10.1002/jcc.540030413.

Raghunathan Ramakrishnan, Pavlo O. Dral, Matthias Rupp, and O. Anatole von Lilienfeld. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific Data*, 1(1):140022, December 2014. ISSN 2052-4463. doi: 10.1038/sdata.2014.22.

Kristof Schütt, Oliver Unke, and Michael Gastegger. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *Proceedings of the 38th International Conference on Machine Learning*, pp. 9377–9388. PMLR, July 2021.

Kristof T. Schütt, Pieter-Jan Kindermans, Huziel E. Sauceda, Stefan Chmiela, Alexandre Tkatchenko, and Klaus-Robert Müller. SchNet: A continuous-filter convolutional neural network for modeling quantum interactions. *arXiv:1706.08566 [physics, stat]*, December 2017.

John C. Snyder, Matthias Rupp, Katja Hansen, Klaus-Robert Müller, and Kieron Burke. Finding Density Functionals with Machine Learning. *Physical Review Letters*, 108(25):253002, June 2012. ISSN 0031-9007, 1079-7114. doi: 10.1103/PhysRevLett.108.253002.

Sina Stocker, Johannes Gasteiger, Florian Becker, Stephan Günnemann, and Johannes T. Margraf. How Robust are Modern Graph Neural Network Potentials in Long and Hot Molecular Dynamics Simulations? Preprint, Chemistry, April 2022.

Jianwei Sun, Adrienn Ruzsinszky, and John P. Perdew. Strongly Constrained and Appropriately Normed Semilocal Density Functional. *Physical Review Letters*, 115(3):036402, July 2015. ISSN 0031-9007, 1079-7114. doi: 10.1103/PhysRevLett.115.036402.

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. ISSN 1759-0884. doi: 10.1002/wcms.1340.

Attila Szabo and Neil S. Ostlund. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. Courier Corporation, 2012.

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:1802.08219 [cs]*, May 2018.

Oliver Treutler and Reinhart Ahlrichs. Efficient molecular numerical integration schemes. *The Journal of Chemical Physics*, 102(1):346–354, January 1995. ISSN 0021-9606. doi: 10.1063/1.469408.

Oliver T. Unke, Stefan Chmiela, Huziel E. Sauceda, Michael Gastegger, Igor Poltavsky, Kristof T. Schütt, Alexandre Tkatchenko, and Klaus-Robert Müller. Machine Learning Force Fields. *Chemical Reviews*, 121(16):10142–10186, August 2021. ISSN 0009-2665. doi: 10.1021/acs.chemrev.0c01111.

O. Vahtras, J. Almlöf, and M.W. Feyereisen. Integral approximations for LCAO-SCF calculations. *Chemical Physics Letters*, 213(5-6):514–518, October 1993. ISSN 00092614. doi: 10.1016/0009-2614(93)89151-7.

Simon Wengert, Gábor Csányi, Karsten Reuter, and Johannes T. Margraf. Data-efficient machine learning for molecular crystal structure prediction. *Chemical Science*, pp. 10.1039.D0SC05765G, 2021. ISSN 2041-6520, 2041-6539. doi: 10.1039/D0SC05765G.

He Zhang, Siyuan Liu, Jiacheng You, Chang Liu, Shuxin Zheng, Ziheng Lu, Tong Wang, Nanning Zheng, and Bin Shao. Overcoming the barrier of orbital-free density functional theory for molecular systems using deep learning. *Nature Computational Science*, 4(3):210–223, March 2024. ISSN 2662-8457. doi: 10.1038/s43588-024-00605-8.

Xuan Zhang, Limei Wang, Jacob Helwig, Youzhi Luo, Cong Fu, Yaochen Xie, Meng Liu, Yuchao Lin, Zhao Xu, Keqiang Yan, Keir Adams, Maurice Weiler, Xiner Li, Tianfan Fu, Yucheng Wang, Haiyang Yu, YuQing Xie, Xiang Fu, Alex Strasser, Shenglong Xu, Yi Liu, Yuanqi Du, Alexandra Saxton, Hongyi Ling, Hannah Lawrence, Hannes Stärk, Shurui Gui, Carl Edwards, Nicholas Gao, Adriana Ladera, Tailin Wu, Elyssa F. Hofgard, Aria Mansouri Tehrani, Rui Wang, Ameya Daigavane, Montgomery Bohde, Jerry Kurtin, Qian Huang, Tuong Phung, Minkai Xu, Chaitanya K. Joshi, Simon V. Mathis, Kamyar Azizzadenesheli, Ada Fang, Alán Aspuru-Guzik, Erik Bekkers, Michael Bronstein, Marinka Zitnik, Anima Anandkumar, Stefano Ermon, Pietro Liò, Rose Yu, Stephan Günnemann, Jure Leskovec, Heng Ji, Jimeng Sun, Regina Barzilay, Tommi Jaakkola, Connor W. Coley, Xiaoning Qian, Xiaofeng Qian, Tess Smidt, and Shuiwang Ji. Artificial Intelligence for Science in Quantum, Atomistic, and Continuum Systems, November 2023.

Yi Zhou, Jiang Wu, Shuguang Chen, and GuanHua Chen. Toward the Exact Exchange–Correlation Potential: A Three-Dimensional Convolutional Neural Network Construct. *The Journal of Physical Chemistry Letters*, 10(22):7264–7269, November 2019. ISSN 1948-7185, 1948-7185. doi: 10.1021/acs.jpclett.9b02838.

A KOHN-SHAM DENSITY FUNCTIONAL THEORY

In Kohn-Sham density functional theory (KS-DFT) (Kohn & Sham, 1965), one attempts to approximate the ground state energy of a molecular system by finding the electron density that minimizes the energy functional Equation 1. Here, we largely compress Lehtola et al. (2020) to provide a brief introduction to the topic. We will neglect electron spin for simplicity and focus on a discretized DFT scheme based on atom-centered basis sets. Operators, i.e., functions mapping from functions to functions, are denoted by acting upon the ket, e.g., $\nabla|f\rangle$ is the derivative of f. Further, we use the braket notation to denote the inner product of two functions f and g as $\langle f|g\rangle = \int g(x)f(x)dx$. We will use Einstein's summation convention, where indices that only appear on one side of an equation are summed over.

The electron density ρ of an N_{el} system is associated with an antisymmetric wave function of a fictitious non-interacting system consisting of N_{el} orthonormal orbitals $\phi_i : \mathbb{R}^3 \to \mathbb{R}$. These orbitals are constructed as linear combinations of N_{bas} a finite basis set $\chi_\mu : \mathbb{R}^3 \to \mathbb{R}$ such that $\phi_i(r) = C_{\mu i}\chi_\mu(r)$. For systems without periodic boundary conditions, these basis sets typically consist of nuclei-centered functions called atomic orbitals. The electron density and pseudo-wave-function are given by

$$\Psi(\mathbf{r}) = \frac{1}{\sqrt{N!}} \det\left[\phi_i(r_j)\right]_{i,j=1}^{N}, \tag{23}$$

$$\begin{aligned}\rho(r) &= \int \Psi(r, r_2, ..., r_N)^* \Psi(r, r_2, ..., r_N) dr_2 ... dr_N \\ &= \boldsymbol{\phi}(r)^T \boldsymbol{\phi}(r) = \boldsymbol{\chi}(r)^T C^T C \boldsymbol{\chi}(r) = \boldsymbol{\chi}(r)^T P \boldsymbol{\chi}(r),\end{aligned} \tag{24}$$

where $C \in \mathbb{R}^{N_{\text{bas}} \times N_{\text{el}}}$ are the orbital coefficients and $P = C^T C \in \mathbb{R}^{N_{\text{bas}} \times N_{\text{bas}}}$ is the so-called density matrix. This construction ensures that the electron density belongs to the class of densities expressible by antisymmetric wave functions and enables the exact computation of the kinetic energy (Kohn & Sham, 1965)

$$T[\rho] = -\frac{1}{2}\langle\phi_i|\nabla^2|\phi_i\rangle = -\frac{1}{2}C_{\mu i}C_{\nu i}\langle\chi_\mu|\nabla^2|\chi_\nu\rangle = C_{\mu i}C_{\nu i}T_{\mu\nu} = P_{\mu\nu}T_{\mu\nu}, \tag{25}$$

making KS-DFT significantly more accurate than contemporary orbital-free DFT.

The external potential in molecular systems is defined by the nuclei positions $\mathbf{R}$ and charges $\mathbf{Z}$

$$V_{\text{ext}}[\rho] = -\int \frac{\rho(r)Z_n}{|r - R_n|}dr = -C_{\mu i}C_{\nu i}\int \frac{\chi_\mu(r)Z_n\chi_\nu(r)}{|r - R_n|}dr = C_{\mu i}C_{\nu i}V_{\mu\nu} = P_{\mu\nu}V_{\mu\nu}. \tag{26}$$

The Hartree energy is a mean-field approximation to the electron-electron interaction

$$\begin{aligned}V_{\text{H}}[\rho] &= \frac{1}{2}\iint \frac{\rho(r)\rho(r')}{|r - r'|}drdr' = C_{\mu i}C_{\nu i}C_{\lambda j}C_{\kappa j}\iint \chi_\mu(r)\chi_\lambda(r')\frac{1}{|r - r'|}\chi_\nu(r)\chi_\kappa(r')drdr' \\ &= C_{\mu i}C_{\nu i}C_{\lambda j}C_{\kappa j}J_{\mu\nu\lambda\kappa} = P_{\mu\nu}P_{\lambda\kappa}J_{\mu\nu\lambda\kappa}.\end{aligned} \tag{27}$$

Finally, the exchange-correlation energy is a functional of the electron density, which is unknown and has to be approximated. This is the main problem we are addressing in this work.

To find the ground state, the parameters C are optimized to minimize Equation 1. This is typically achieved iteratively in a self-consistent fashion; see Appendix B. The construction via Gaussian-type basis functions (GTOs) greatly simplifies the integrals appearing in the calculation allowing us to analytically precompute the so-called core Hamiltonian $H_{\text{core}\,\mu\nu} = T_{\mu\nu} + V_{\mu\nu}$ and the electron-electron repulsion tensor $J_{\mu\nu\lambda\kappa}$ once and reuse them throughout the optimization.

B SELF-CONSISTENT FIELD METHOD

In this work, we use the Self-Consistent Field (SCF) method (Kohn & Sham, 1965; Lehtola et al., 2020) to find the coefficients $\hat{C}$ that minimize the energy functional $E[\rho(r; C)]$ in Eq. 1

$$\hat{C} = \arg\min_C E = \arg\min_C P_{\mu\nu}H_{\text{core}\,\mu\nu} + P_{\mu\nu}P_{\lambda\kappa}J_{\mu\nu\lambda\kappa} + E_{\text{XC}} \tag{28}$$

$$\text{s.t. } \forall_{ij}\langle\phi_i|\phi_j\rangle = \delta_{ij}. \tag{29}$$

We can express the orthogonality constraint above via the overlap matrix $S \in \mathbb{R}^{N_{\text{bas}} \times N_{\text{bas}}}$

$$\forall_{ij} \langle \phi_i | \phi_j \rangle = \delta_{ij} \iff C^T S C = I, \tag{30}$$

$$S_{\mu\nu} \coloneqq \langle \chi_\mu | \chi_\nu \rangle. \tag{31}$$

This constraint is enforced by introducing the Lagrange multipliers E_{ij}, and we define the loss

$$\mathcal{L} = E + (C^T S C - I)_{ij} E_{ij}. \tag{32}$$

Since the matrix E is symmetric and the rotation of the basis does not change the energy, we can choose E to be diagonal. Using the Fock matrix $F \coloneqq \frac{\partial E}{\partial P}$, we can write the gradient of the loss w.r.t. C as

$$\frac{\partial \mathcal{L}}{\partial C} = 2FC - 2SCE \stackrel{!}{=} 0 \tag{33}$$

$$\implies FC = SCE. \tag{34}$$

This is the generalized eigenvalue problem we need to solve to find the coefficients C. We solve this by diagonalizing $S = V \Lambda V^T$, substituting $C = X\tilde{C}$ with $X = V\Lambda^{-1/2}V^T$, and multiplying both sides from the left with X^T to obtain the ordinary eigenvalue problem

$$\begin{aligned} \underbrace{X^T F X}_{\tilde{F}} \tilde{C} &= \underbrace{X^T S X}_{I} \tilde{C} E \\ \implies \tilde{F} &= \tilde{C} E. \end{aligned} \tag{35}$$

After solving this for $\tilde{C}$, we recover the original coefficients $C = X^T \tilde{C}$ via the eigenvectors associated with the N_{el} lowest eigenvalues E. Since F depends on $P = CC^T$, this problem is linearized by alternating optimization steps of $F(C)$, C, s.t. we need to iterate until convergence commonly referred to as self-consistence in this context.

C Self-consistent field implementation

We implement the SCF method from Appendix B fully differentiably in JAX (Bradbury et al., 2018) by following Lehtola et al. (2020). In particular, we precompute $\boldsymbol{\chi}(r)$ on the integration grid points (Treutler & Ahlrichs, 1995), and the core-Hamiltonian $H_{\mu\nu}$. Since the electron-repulsion integrals $J_{\mu\nu\lambda\kappa}$ scale as $O({N_{\text{bas}}}^4)$ in compute and memory, we use the density-fitting approximation (Vahtras et al., 1993) to reduce the memory and compute scaling to $O({N_{\text{bas}}}^2 N_{\text{aux}})$ where $N_{\text{aux}} \ll {N_{\text{bas}}}^2$ is the size of the auxiliary basis set. We explain the procedure in more detail in Appendix D. Additionally, we improve the convergence by implementing the direct inversion of the iterative subspace (DIIS) method, which we briefly summarize in Appendix E (Pulay, 1982). For precomputing the integrals and obtaining grid points, we use the PySCF (Sun et al., 2018). For the evaluation of the atomic orbitals on the integration grid points, we use our own JAX implementation.

Like Dick & Fernandez-Serra (2021), we perform several SCF iterations with a pre-defined XC-functional to obtain a good initial guess for the density. During training, we randomly interpolate between the precycled initial guess and a standard initial guess via

$$P_0 = \frac{1+\alpha}{2} P_{\text{precycle}} + \left(1 - \frac{1+\alpha}{2}\right) P_{\text{standard}} \tag{36}$$

where $\alpha \sim \mathcal{U}(0,1)$. At inference time, we fix $\alpha = 1$. This interpolation leads to varying initial densities and functions as a regularizer (Dick & Fernandez-Serra, 2021).

D Density fitting

Here, we largely follow Krisiloff et al. (2015) and give a brief introduction to the density-fitting approximation. For more details, we refer the reader to the original work. We want to approximate the two-electron integral tensor

$$J_{\mu\nu\lambda\kappa} = \int \chi_\mu(r)\chi_\nu(r') \frac{1}{|r-r'|} \chi_\lambda(r)\chi_\kappa(r') dr dr'. \tag{37}$$

To reduce the memory and compute scaling from $O({N_\text{bas}}^4)$ to $O({N_\text{bas}}^2 N_\text{aux})$, we introduce an auxiliary basis set $\{\chi_\mu^\text{aux}\}_{\mu=1}^{N_\text{aux}}$ and approximate the two-electron integrals as

$$J_{\mu\nu\lambda\kappa} \approx \hat{J}_{(\mu\nu)I} \tilde{J}_{IJ}^{-1} \hat{J}_{(\lambda\kappa)J} \tag{38}$$

where

$$\hat{J}_{(\mu\nu)I} = \int \chi_\mu(r)\chi_\nu(r) \frac{1}{|r-r'|} \chi_I^\text{aux}(r) dr dr', \tag{39}$$

$$\tilde{J}_{IJ} = \int \chi_I^\text{aux}(r) \frac{1}{|r-r'|} \chi_J^\text{aux}(r') dr dr', \tag{40}$$

and the indices I, J run over the auxiliary basis set $\{\chi_I^\text{aux}\}_{I=1}^{N_\text{aux}}$. Note that choosing $\{\chi_I^\text{aux}\}_{I=1}^{N_\text{aux}} = \{\chi_\mu \cdot \chi_\nu\}_{\mu,\nu=1}^{N_\text{bas}}$ yields the exact two-electron integrals. One can further simplify by computing

$$B_{(\mu\nu)I} = \hat{J}_{(\mu\nu)J} \tilde{J}_{IJ}^{-\frac{1}{2}} \tag{41}$$

where $\tilde{J}^{-\frac{1}{2}} = \tilde{J}^{-1}\tilde{J}^{\frac{1}{2}}$ with $\tilde{J}^{\frac{1}{2}}$ being the Choleksy decomposition of $\tilde{J}$. This reduces the contraction of the two-electron integrals with the density matrix to

$$V_\text{H}[\rho] = P_{\mu\nu} J_{\mu\nu\lambda\kappa} P_{\lambda\kappa} \approx P_{\mu\nu} B_{(\mu\nu)I} B_{(\lambda\kappa)I} P_{\lambda\kappa}. \tag{42}$$

E Direct inversion of iterative subspace

The direct inversion of the iterative subspace (DIIS) method is a common technique to accelerate the convergence of the SCF method. Here, we briefly summarize the method; for a derivation and more details, we refer the reader to Pulay (1982). The DIIS method aims at finding a linear combination of previous Fock matrices that minimizes the norm of the error matrix. This accelerates the convergence of the SCF method by extrapolating a new Fock matrix from previous ones. Let F_k be the Fock matrix in the k-th iteration and P_k the density matrix. Then the error matrix is defined as

$$e_k = \left(S^{-\frac{1}{2}}\right)^T (F_k P_k S - S P_k F_k) S^{-\frac{1}{2}} \tag{43}$$

where S is the overlap matrix. We aim to find coefficients c_i that minimize the norm of a linear combination of error matrices

$$\min_{\{c_i\}} \left\| \sum_{i=1}^m c_i e_i \right\|^2 \tag{44}$$

subject to the constraint $\sum_{i=1}^m c_i = 1$. This minimization problem can be solved using Lagrange multipliers, leading to the linear system

$$\begin{pmatrix} \boldsymbol{B} & \mathbf{1} \\ \mathbf{1}^T & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{c} \\ \lambda \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ 1 \end{pmatrix} \tag{45}$$

where $B_{ij} = \text{Tr}(e_i^T e_j)$, 1 is a vector of ones, and λ is a Lagrange multiplier. Solving this system gives us the optimal coefficients c_i which we use to obtain the extrapolated Fock matrix that is solved in the next SCF iteration.

F Hyperparameters

We list the hyperparameters for GG-XC in Table 3.

G Baseline changes

For reference methods, we used the model hyperparameters from their respective works. For NequIP, we used the default $l = 2$. We extended patience schedules to ensure full convergence for all baselines. On the QM9(S) datasets, we first fitted a linear model on the training set with the number of atoms and shifted all labels by this prediction. Further, we initialized all models with zero in the last layer. These steps improved generalization between 10 and 100 times on the smaller datasets.

Table 3: Hyperparameters for GG-XC.

Hyperparameter	Value
d number of features per irrep	32
l_{max} number of irreps	2
T number of layers	3
Radial filters	32
ϵ_{mGGA} Base semilocal functional	Dick & Fernandez-Serra (2021)
Batch size	1
Number of steps to compute loss	3
Parameter EMA	0.995
Optimizer	Adam
β_1	0.9
β_2	0.999
Basis set	6-31G(d)
Density fitting basis set	weigend
SCF iterations	15
Precycle XC functional	LDA
Precycle iterations	15
Learning rate	
MD17	$\frac{0.01}{1+\frac{1}{1000}}$
3BPA	$\frac{0.01}{1+\frac{1}{1000}}$
QM9	$\frac{0.001}{1+\frac{1}{1000}}$

Table 4: MD17 with only 50 samples.

	Force field			Δ-ML			KS-DFT	
	SchNet	PaiNN	NequIP	SchNet	PaiNN	NequIP	Dick	GG-XC
Aspirin	8.94	9.46	8.75	9.55	3.27	3.19	1.98	**1.23**
Benzene	2.11	2.28	2.84	0.25	0.26	**0.21**	0.53	0.59
Ethanol	4.15	3.30	4.44	2.99	1.60	1.32	0.99	**0.35**
Malonaldehyde	5.47	3.99	5.05	2.81	1.77	1.36	0.78	**0.54**
Toluene	6.65	4.40	5.15	1.15	0.79	0.80	**0.70**	0.99

H MD17 WITH 50 SAMPLES

We imitate the setting of Batatia et al. (2022) and train all models on only 50 samples for each of the MD17 trajectories. We reduce the validation set to 10 samples as well. As we found KS-DFT methods to be more learning rate sensitive here, we ablated the initial learning rate with 0.01 and 0.001 and report the lower ones for Dick & Fernandez-Serra (2021) and GG-XC. The resulting test set MAEs are listed in Table 4. KS-DFT methods and Δ-ML are more successful learning from such few samples. Especially on the challenging aspirin structures, GG-XC is the only method yielding accuracies within chemical accuracy.

I DELTA ML WITH SCAN (6-31G) DFT

In Table 5, we present additional MD17 data with the SCAN functional (Sun et al., 2015) and the 6-31G(d) basis set. The results show that improving the reference DFT functional significantly improves Δ-ML approaches. It should be noted that the pure SCAN is already more accurate than the Dick & Fernandez-Serra (2021) specifically fitted functional on Aspirin. We argue that this setting is an ideal case of Δ-ML as SCAN is well-suited for such close-to-equilibrium structures. However, it

Table 5: MAE for Δ-ML with SCAN and the 6-31G(d) basis set on MD17. For SCAN, we report the relative MAE instead to account for mean shifts.

Molecule	SchNet	PaiNN	NequIP	SCAN
Aspirin	0.57	0.46	0.34	1.84
Benzene	0.08	0.07	0.02	1.27
Ethanol	0.26	0.17	0.08	1.04
Malonaldehyde	0.32	0.25	0.13	1.26
Toluene	0.21	0.17	0.08	1.92

Table 6: Absolute MAE in in $\mathrm{m}E_\mathrm{h}$ for structural extrapolation on the 3BPA dataset. All methods are trained on the 300K training set.

	Force field			Δ-ML			KS-DFT	
Test set	SchNet	PaiNN	NequIP	SchNet	PaiNN	NequIP	Dick	GG-XC
300K	5.15	2.91	3.85	2.38	1.14	<u>0.81</u>	0.96	**0.42**
600K	28.14	13.85	24.56	4.09	2.13	<u>1.56</u>	2.19	**1.43**
1200K	93.50	49.14	81.00	7.05	6.01	<u>3.30</u>	6.32	<u>4.39</u>
$\beta = 120°$	19.98	4.16	13.77	2.54	1.26	<u>1.09</u>	1.58	**1.04**
$\beta = 150°$	20.13	6.64	15.53	2.59	<u>0.89</u>	**0.88**	1.60	1.02
$\beta = 180°$	20.25	9.01	17.35	1.93	1.15	**0.77**	1.70	<u>1.04</u>

should be mentioned that such improvements are orthogonal as one may also apply Δ-learning on top of GG-XC.

J 3BPA MEAN ABSOLUTE ERROR

In Table 6, we present the absolute MAE on the 3BPA dataset for all methods. It is apparent that while GG-XC performs well, Δ-NequIP achieves the lowest MAE. Thus, demonstrating that GG-XC's error is largly a constant offset that would not affect actual MD simulations.

K 3BPA POTENTIAL ENERGY SURFACES

We plot the potential energy surfaces for $120°$, $150°$, and $180°$ in Figure 4, Figure 5, and Figure 6, respectively. It is apparent that force fields struggle on all energy surfaces while XC-functionals yield a similar surface structure to the target.

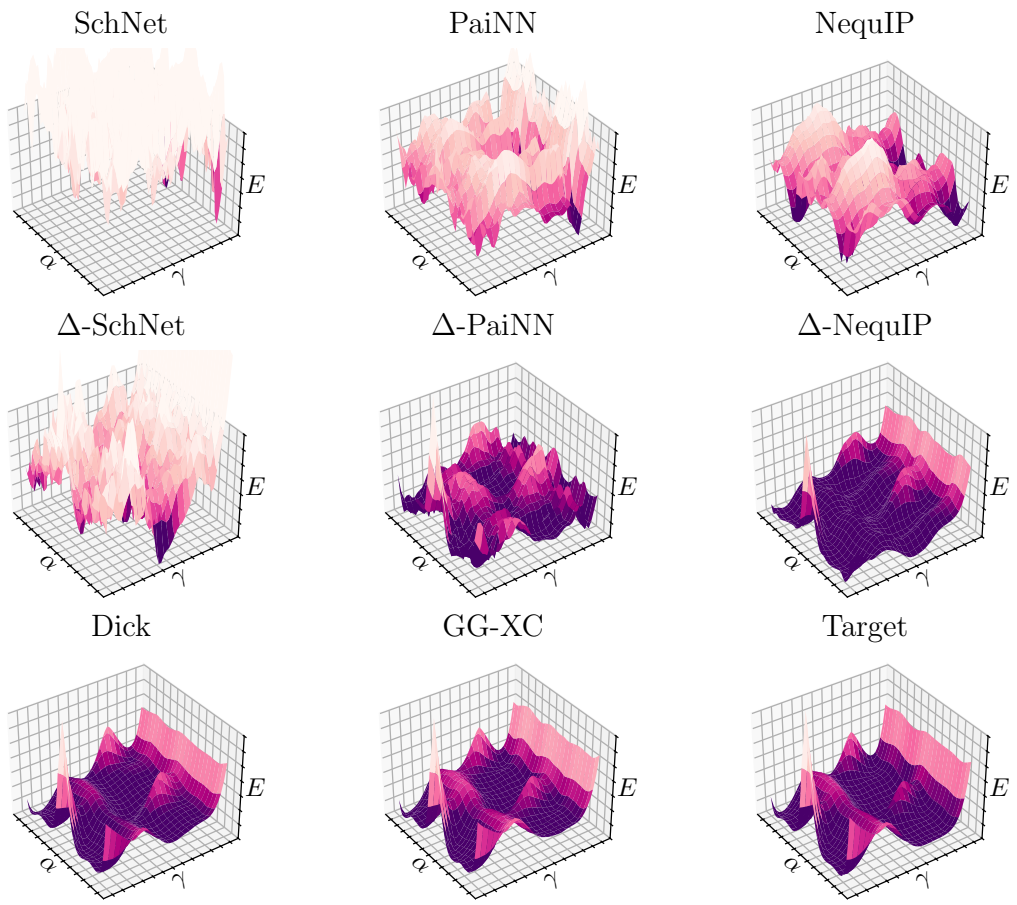

Figure 4: Energy surfaces of the 3BPA dataset at $\beta = 120°$.

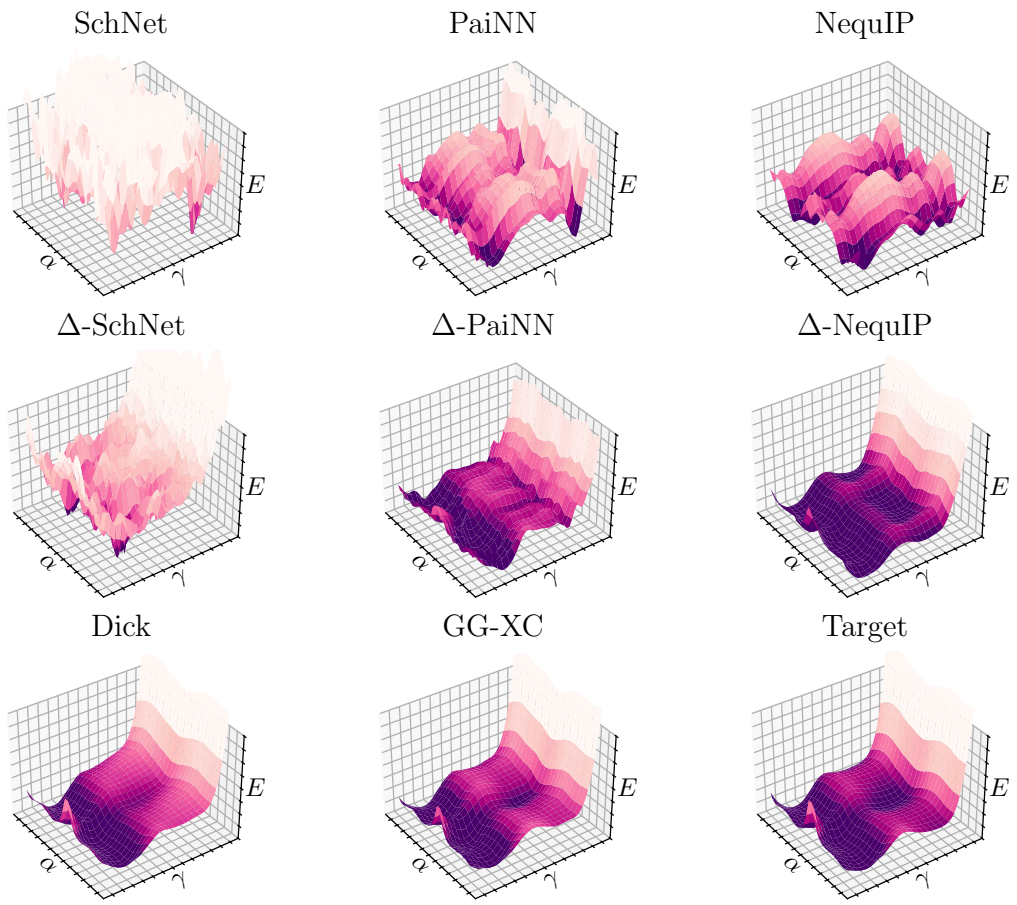

Figure 5: Energy surfaces of the 3BPA dataset at $\beta = 150°$.

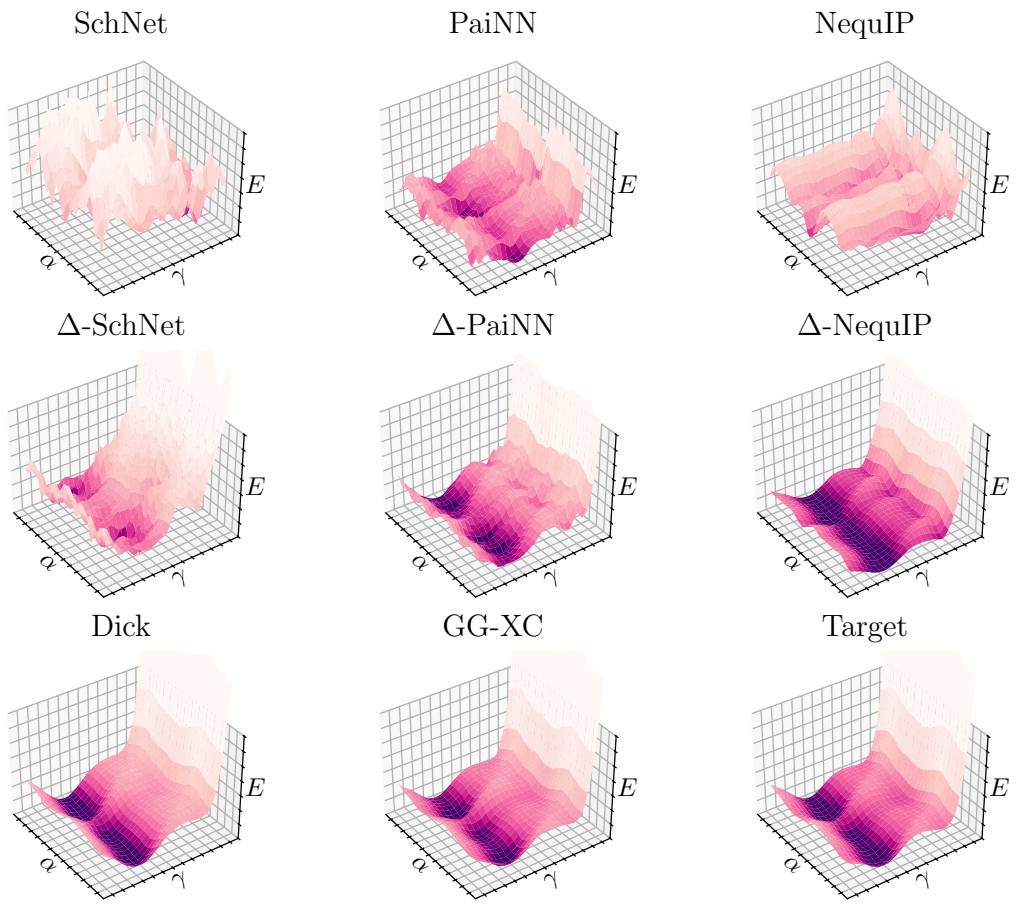

Figure 6: Energy surfaces of the 3BPA dataset at $\beta = 180°$.