BONDING-AWARE REPRESENTATION LEARNING FOR MACHINE-LEARNING FORCE FIELDS

Anonymous authors
Paper under double-blind review

ABSTRACT

Machine-learning force fields (MLFFs) provide a promising framework for modeling molecular energies and forces from atomic configurations. However, most existing MLFFs construct representations from local geometric neighborhoods defined by a fixed cutoff, treating nearby atom pairs primarily through their relative positions. Such representations do not explicitly encode bond connectivity or continuous changes in bond state during dynamics, e.g., reactive dynamics. Consequently, they may struggle to capture the associated variations in energy and force, particularly when the accompanying charge redistribution must satisfy global charge conservation. To address this challenge, we introduce a bond-aware representation framework that models local atom pairs as nodes in a bipartite bond-incidence graph. Each pair is assigned a structured state with orientation-invariant and orientation-equivariant components: the former captures latent bonding information, while the latter encodes signed charge-transfer flow. Aggregation through a signed incidence matrix produces orientation-independent atomic charges while exactly conserving the prescribed total charge. We evaluate our method on rMD17, MD22, Chignolin, Transition1x, and HORM, covering both non-equilibrium and reactive molecular dynamics. Across these benchmarks, our bond-aware model improves energy and force prediction accuracy. Stability analyses and assessments of energy drift further demonstrate the performance of our model in molecular dynamics simulations.

1 INTRODUCTION

Machine-learned force fields (MLFFs) offer a promising paradigm for atomistic simulation by learning energies and forces from molecular data. In particular, equivariant graph neural networks encode Euclidean symmetries directly and have shown favorable accuracy and data efficiency on a range of molecular and materials benchmarks (Schütt et al., 2021; Batzner et al., 2022; Musaelian et al., 2023; Batatia et al., 2022). A broader challenge arises when local bonding states vary continuously: an atom pair may be strongly bonded in one configuration, weakly bonded in another, and nonbonded in a third. Such variations occur in diverse settings, including chemical reactions (Van Duin et al., 2001), conformational distortions (Pophristic & Goodman, 2001), and thermal expansion (Kovács et al., 2021; Hong et al., 2026), as illustrated in Figure 1(a). A suitable force field should therefore capture the variations in energy and force associated with continuously changing bonding states [1].

Classical force fields address related issues by making bond order and charge explicit, but their accuracy and transferability are limited by hand-designed functional forms (Van Duin et al., 2001; Rappe & Goddard III, 1991). Standard MLFFs typically represent atom pairs within a geometric cutoff as edges in a message-passing graph (Batzner et al., 2022; Batatia et al., 2022; Liao & Smidt, 2023). Although these edges provide effective channels for communicating interatomic information, they do not maintain dedicated representations of pairwise bonding states, causing the representational bottleneck illustrated in Figure 1(b). This bottleneck can hinder accurate modeling of energy

[1] Here, the physical bonding state of an atom pair refers to its configuration- and electronic-state-dependent interaction, including bond strength, electron sharing, polarization, and charge-transfer behavior. Our primary objective, however, remains the accurate prediction of energies and forces rather than the direct recovery of these bond-state properties.

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

and force variations during changes in the bonding state, particularly when the associated charge redistribution must also conserve the system's total charge (Rappe & Goddard III, 1991).

Meanwhile, electronic- and bond-aware models enrich the force-field representation with additional electronic or interaction-level variables, often introducing auxiliary inputs, supervision, or physical assumptions (Unke et al., 2021a; Ko et al., 2021; Kalita et al., 2026; Cools-Ceuppens et al., 2022; Deng et al., 2023). Dataset- and sampling-based approaches instead improve coverage of difficult configurations through targeted off-equilibrium data generation and active learning, which can require substantial additional quantum-mechanical data (Zhang et al., 2025; 2024b; Young et al., 2021; Vitartas et al., 2026; Vandermause et al., 2022; Zhang et al., 2024a; Liu et al., 2025; Cui et al., 2025; Song et al., 2025). However, these strategies do not generally provide explicit and effective bond state representations of atomic pairs. Here raises a research problem: *How to incorporate the modeling of bond states to enhance the performance of energy and force prediction?*

We address this problem with BONDFF, a bond-aware MLFF framework built upon two complementary graphs: a *cutoff-based geometric graph* and a *bipartite bond-incidence graph*. The cutoff-based geometric graph captures spatial communication among atoms within a fixed cutoff. The bipartite bond-incidence graph models latent bond states that contribute to the potential energy while ensuring charge conservation, as illustrated in Figure 1(c). In particular, our bipartite bond-incidence graph comprises two types of nodes: atom nodes and virtual nodes that model the bond state, which we dubbed *bond-state nodes*. We define each bond-state node as an atom pair within an element-dependent cutoff, since bond formation depends on the elements and their distance. In the bipartite bond-incidence graph, each bond-state node is connected to the two corresponding atom nodes.

Together, the two graphs enable geometric information exchange and explicit bond-aware state modeling, providing the following advantages for accurate force-field construction. 1) The bond states encode latent bonding information and orientation-aware charge transfer, enabling the model to represent bond-dependent energetics. 2) The latent charge transfers are modeled as orientation-equivariant signed flows to enforce the total charge conservation for the molecular system. 3) As additional benefits of this formulation, bond states can exchange information through shared atoms and neighboring bond-state nodes, potentially capturing complex bonding environments, such as aromatic rings (Reuter & Lüchow, 2021). 4) Besides, the latent charge representation also supports the modeling of electrostatics that contributes to long-range interactions (Kosmala et al., 2023).

We evaluate BONDFF on six benchmarks to test the effectiveness of proposed bond-state framework in various application scenarios. rMD17 (Christensen & Von Lilienfeld, 2020) tests energy and force prediction for molecules under non-equilibrium environments. MD22 (Chmiela et al., 2023) aims at large and diverse systems with more complex bonding environments. The 3BPA (Kovács et al., 2021) contains molecules with different bond states sampled from different temperatures. The Chignolin benchmark (Wang et al., 2023a) tests whether the model captures interactions beyond local atomic neighborhoods in a protein-scale system. Finally, Transition1x (Schreiner et al., 2022) samples configurations along reaction pathways to assess changing connectivity, whereas HORM (Cui et al., 2025) provides reactive geometries and Hessian labels to test the prediction of local potential-energy curvature. Across these benchmarks, our bond-aware model improves the accuracy of both energy and force predictions. The stability analyses and energy-drift measurements further demonstrate its reliability in molecular dynamics simulations.

2 PRELIMINARIES

Problem formulation. We consider a system of N atoms with atomic index $Z = (Z_1, \ldots, Z_N)$ and Cartesian coordinates $R = (r_1, \ldots, r_N) \in \mathbb{R}^{3N}$. Given reference energies $E^\star$ and forces $F^\star$, an MLFF learns a differentiable scalar potential $E_\theta(Z, R)$ that is invariant to translations, rotations, and permutations of equivalent atoms. Conservative forces and the Cartesian Hessian follow from differentiation, $F_i = -\nabla_{r_i} E_\theta$ and $H = \nabla_R^2 E_\theta = -\nabla_R F \in \mathbb{R}^{3N\times 3N}$, respectively. Thus, energy–force learning fits the potential and its first derivatives, whereas Hessian information constrains local curvature—including vibrational modes and the unstable direction near a transition state.

Related works. Equivariant MLFFs encode rotational symmetry for efficient atomistic simulation (Schütt et al., 2021; Batzner et al., 2022; Musaelian et al., 2023; Batatia et al., 2022), while directional graph networks highlight the importance of edge geometry and angular interactions

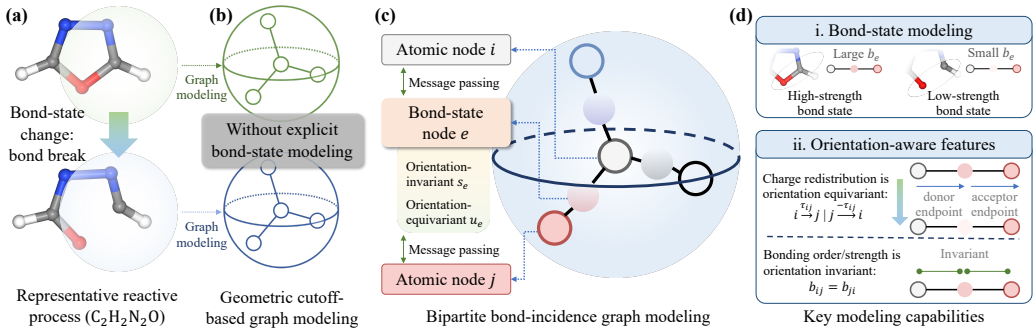

Figure 1: **Motivation, idea, and benefits of our bond-state graph modeling.** (a) Bonding states evolve continuously during a reactive transformation. (b) Conventional geometric graphs connect atoms within a distance cutoff but do not explicitly model bond states, making distinct interaction strengths difficult to discern. (c) BONDFF introduces bond-state nodes with orientation-invariant features s_e and orientation-equivariant features u_e, yielding symmetric latent bond strengths $b_{ij} = b_{ji}$ and antisymmetric charge flows $\tau_{ij} = -\tau_{ji}$. These nodes represent latent interactions rather than discrete chemical bonds. (d) Two key modeling capabilities: explicit bond-state modeling and orientation-aware representation.

(Gasteiger et al., 2020b; 2021). Reactive modeling also requires bond rearrangement and charge redistribution. ReaxFF uses empirical bond orders and charge equilibration (Van Duin et al., 2001; Senftle et al., 2016); SpookyNet and CHGNet incorporate electronic-state information (Unke et al., 2021a; Deng et al., 2023). Other work broadens reactive training data and sampling, including Hessian data for transition states (Zhang et al., 2025; 2024b; Vitartas et al., 2026; Zhang et al., 2024a; Song et al., 2025; Cui et al., 2025). BONDFF instead represents evolving bond states as pair charges, capturing charge transfer while conserving total charge. We provide more discussion in Appendix A.

3 METHOD

Our BONDFF is built upon two graphs: a cutoff-based geometric graph for atom–atom communication and a bipartite bond-state incidence graph for bond-state modeling. Each unordered atomic pair within a prescribed, element-dependent cutoff is assigned a bond-state node connected to its two atomic endpoints. These nodes represent latent interactions between atomic pairs without requiring a discrete chemical bond assignment. Figure 2 illustrates the overall architecture.

3.1 BOND-STATE INCIDENCE GRAPH

Given atoms $\mathcal{V} = \{1, \ldots, N\}$, where N is the number of atoms and i, j are atom indices, the cutoff-based geometric graph operates on graph $\mathcal{G}_{\text{geo}} = (\mathcal{V}, \mathcal{E}_{\text{geo}})$. Here $\mathcal{E}_{\text{geo}}$ is the geometric edge set, where edges are defined by a fixed cutoff; its edges do not carry a chemical bonding representation. To represent changing connectivity without the discrete chemical bond labels, we design a bond-state incidence graph G_{BI} with a set of atomic nodes $\mathcal{V}$ and a set of virtual pairwise bond-state nodes $\mathcal{B}$ as:

$$\mathcal{B} = \left\{ e = \{i, j\} : i < j,\ r_{ij} < R_{\text{bond}}^{(Z_i, Z_j)} \right\}, \qquad M = |\mathcal{B}|, \tag{1}$$

and the bipartite bond-incidence graph $\mathcal{G}_{\text{BI}} = (\mathcal{V} \cup \mathcal{B}, \mathcal{I})$, where $\mathcal{I} = \{(i, e), (j, e) : e = \{i, j\} \in \mathcal{B}\}$ is the atom–bond incidence set. A bond-state node is a computational representation of a latent interaction, not an asserted chemical bond.

Before bond-state message passing, each pair $e = \{i, j\} \in \mathcal{B}$ is assigned its distance $r_e = \|r_i - r_j\|_2$ and a cutoff weight $c_e = c(r_e / R_{\text{bond}}^{(Z_i, Z_j)})$, where $R_{\text{bond}}^{(Z_i, Z_j)}$ is the element-dependent cutoff; c_e scales the bond-strength, charge-flow, and bond-energy readouts. For signed charge flow, we arbitrarily choose one endpoint as the tail $t(e)$ and the other as the head $h(e)$, then form $B \in \mathbb{R}^{N \times M}$ with -1 at each tail and $+1$ at each head, and $C = |B|$, where N is the number of atoms and $M = |\mathcal{B}|$; the cutoff function and reference-direction rules are detailed in Appendix C.1.

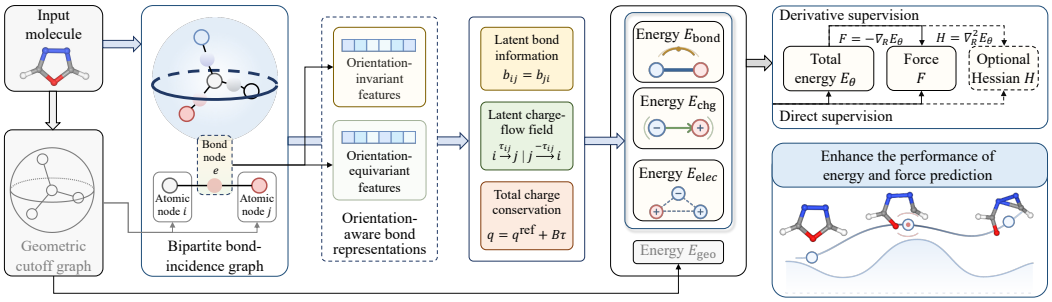

Figure 2: **Overview of the BONDFF architecture to model bond-states.** The geometric cutoff-based MLFF part provides atomic representations and a baseline energy, while the bond-incidence branch learns orientation-aware bond features, bond strengths, and charge-conserving charge flows. Three energy terms are inspired by the Morse potential, local charge response, and Coulomb's law, respectively. Their combination yields the total energy, conservative force, and Hessian derivatives.

Orientation-aware bond-state node. To separate orientation-invariant information (energy) from signed transfer information (transfer flow), bond state e carries two components at layer ℓ:

$$H_e^{(\ell)} = \left(s_e^{(\ell)}, u_e^{(\ell)}\right), \qquad s_e^{(\ell)} \in \mathbb{R}^{d_s}, \quad u_e^{(\ell)} \in \mathbb{R}^{d_u}. \tag{2}$$

Here d_s and d_u are the orientation-invariant and equivariant channel widths. Let $S^{(\ell)} \in \mathbb{R}^{M\times d_s}$ and $U^{(\ell)} \in \mathbb{R}^{M\times d_u}$ stack these features. Reorientation acts as

$$S'^{(\ell)} = S^{(\ell)}, \qquad U'^{(\ell)} = \Delta U^{(\ell)}, \tag{3}$$

or, equivalently, $s_{ij} = s_{ji}$ and $u_{ij} = -u_{ji}$. We call s_e orientation-invariance and u_e orientation-equivariance. This orientation symmetry acts in bond-index space and is distinct from spatial $E(3)$ equivariance: neither s_e nor u_e is necessarily a three-dimensional vector.

To initialize both components without imposing a hand-designed chemical rule, we use the geometric graph outputs $h_i^{\text{geo}} = h_{i,\text{geo}}^{(L)}$ as inputs to a shared initialization encoder ϕ_{init} in both endpoints,

$$d_{i\to j}^{(0)} = \phi_{\text{init}}\left(h_i^{\text{geo}}, h_j^{\text{geo}}, \text{RBF}(r_{ij}), c_e\right), d_{j\to i}^{(0)} = \phi_{\text{init}}\left(h_j^{\text{geo}}, h_i^{\text{geo}}, \text{RBF}(r_{ij}), c_e\right). \tag{4}$$

Projecting these encodings gives

$$s_{ij}^{(0)} = \frac{d_{i\to j}^{(0)} + d_{j\to i}^{(0)}}{2}, \qquad u_{ij}^{(0)} = \frac{d_{i\to j}^{(0)} - d_{j\to i}^{(0)}}{2}, \tag{5}$$

which satisfies Eq. 3 without constraining ϕ_{init}. Initializing $u_e^{(0)} = 0$ is also valid, provided that subsequent layers contain an invariant-to-equivariant map.

Atomic node features. The bond-state incidence graph takes the final-layer atomic representations learned by the cutoff-based geometric graph as the initialized atomic node features.

In summary, at layer ℓ, the inputs of our graph are the stacked atomic nodes $\mathbf{H}^{(\ell)}$ and the bond-state nodes $(S^{(\ell)}, U^{(\ell)})$. The bond-state incidence graph updates bond-state nodes and atomic atoms through message passing. We provide more details for feature initialization in Appendix C.2.

3.2 INCIDENCE PROPAGATION

Message passing. To let neighboring bond-state nodes coordinate with atomic nodes, each propagation map has the same two-step form: aggregate features of bond-state nodes to atomic nodes, apply an atom-domain map, and scatter the result back to bond-state nodes. Compactly,

$$\mathcal{T}_{a\to b}^{(\ell)}(X) = R_b^\top \Phi_{ab}^{(\ell)}(A_a X, \mathbf{H}^{(\ell)}), \qquad a, b \in \{s, u\}, \tag{6}$$

where X is the source-component bond matrix and $A_s = R_s = C, A_u = R_u = B$. Thus, C is used for the orientation-invariant component and B for the orientation-equivariant component, both when

aggregating and when scattering. Here $\Phi_{ab}^{(\ell)}$ is a learnable atom-domain map: it takes the atom-wise aggregate from component a, conditions it on the backbone features $\mathbf{H}^{(\ell)}$, and returns an atom-wise message for component b. The four expanded definitions and their symmetry check are given in Appendix C.3.

Interaction between the orientation-invariant and equivariant components. We let the invariant and equivariant components interact to enrich the bond-state representation while preserving reorientation symmetry. For this interaction, we define $\rho_\epsilon(x) = \sqrt{x^2 + \epsilon^2} - \epsilon$ with smoothing constant $\epsilon > 0$, so that $\rho_\epsilon(-x) = \rho_\epsilon(x)$. For each bond-state node, the interaction terms between its two components are

$$z_{u\to s,e}^{(\ell)} = W_{us}^{(\ell)} \rho_\epsilon(u_e^{(\ell)}), \qquad z_{s\to u,e}^{(\ell)} = W_u^{(\ell)} u_e^{(\ell)} \odot \sigma\Big(W_g^{(\ell)} s_e^{(\ell)} + b_g^{(\ell)}\Big). \tag{7}$$

The first term is orientation-invariant because ρ_ϵ is unchanged by sign reversal. The second term remains orientation-equivariant because the equivariant state $u_e^{(\ell)}$ is modulated only by a gate computed from the orientation-invariant state $s_e^{(\ell)}$.

Bond-state update. We combine these signals with a gated residual update. Define

$$m_{s,e}^{(\ell)} = \mathcal{T}_{s\to s,e}^{(\ell)} + \mathcal{T}_{u\to s,e}^{(\ell)} + z_{u\to s,e}^{(\ell)}, m_{u,e}^{(\ell)} = \mathcal{T}_{u\to u,e}^{(\ell)} + \mathcal{T}_{s\to u,e}^{(\ell)} + z_{s\to u,e}^{(\ell)}. \tag{8}$$

The updates can then be written compactly as

$$s_e^{(\ell+1)} = s_e^{(\ell)} + \mathcal{U}_s^{(\ell)}\Big(s_e^{(\ell)}, m_{s,e}^{(\ell)}, r_e, c_e\Big), u_e^{(\ell+1)} = u_e^{(\ell)} + \mathcal{U}_u^{(\ell)}\Big(u_e^{(\ell)}, m_{u,e}^{(\ell)}, s_e^{(\ell+1)}\Big), \tag{9}$$

where $\mathcal{U}_s$ updates the invariant component and $\mathcal{U}_u$ updates the equivariant component using gates computed from invariant features. Their expanded definitions are given in Appendix C.4. The invariant update uses only quantities unchanged by reorientation. The equivariant update uses only quantities that change sign under reorientation, multiplied by invariant gates; consequently, affine maps and normalization on a path containing only equivariant features must be bias-free, and its nonlinearity must preserve sign reversal. Besides, the bond branch forms an invariant atom-domain aggregate update for atomic features denoted as $a^{(\ell)} = \Psi_a^{(\ell)}\big(CS^{(\ell)}, BU^{(\ell)}, \mathbf{H}^{(\ell)}\big)$.

3.3 Latent bond state and charge readouts

To approximate a chemically interpretable, distance-dependent quantity, we define the invariant, nonnegative *latent bond-strength* readout

$$b_e = c_e \,\mathrm{softplus}\Big[f_b\Big(s_e^{(L)}, \rho_\epsilon(u_e^{(L)}), \mathrm{RBF}(r_e)\Big)\Big]. \tag{10}$$

Consequently, $b_{ij} = b_{ji} \geq 0$ and $b_e \to 0$ as $r_e \to R_{\mathrm{bond}}$. We call b_e a *latent continuous bond-strength field*. Energy and force supervision alone does not identify it with a Wiberg, Mayer, or other population-analysis bond order. To model the charge transfer that has directions between endpoints, we represent charge redistribution as an orientation-equivariant scalar flow. Let τ_e denote the signed flow on bond e and let $\tau_{\max} > 0$ bound its magnitude:

$$\tau_e = c_e \tau_{\max} \tanh\Big[w_\tau\Big(s_e^{(L)}, r_e\Big)^\top u_e^{(L)}\Big]. \tag{11}$$

Here w_τ is invariant, so reorientation gives $\tau' = \Delta\tau$, or $\tau_{ji} = -\tau_{ij}$. The flow changes endpoint charges through the signed incidence matrix, while a reference allocation handles the prescribed net charge. To represent a system whose prescribed total charge is Q, we add a reference charge vector $q^{\mathrm{ref}} \in \mathbb{R}^N$, where q_i^{ref} is the baseline charge assigned to atom i before bond-mediated redistribution. The predicted atomic charges are

$$q_i = q_i^{\mathrm{ref}} + \sum_{e\in\mathcal{B}} B_{ie}\tau_e, i = 1, \ldots, N. \tag{12}$$

Here $q, q^{\mathrm{ref}} \in \mathbb{R}^N$ are vectors indexed by atoms and $\tau \in \mathbb{R}^M$ collects one signed flow per bond-state node. For systems with varying total charge, we obtain the reference allocation from an invariant atomic score f_Q:

$$\pi_i = \frac{\exp f_Q(h_i^{(L)})}{\sum_{k=1}^{N} \exp f_Q(h_k^{(L)})}, \qquad q_i^{\mathrm{ref}} = Q\pi_i. \tag{13}$$

The softmax weights satisfy $\pi_i > 0$ and $\sum_i \pi_i = 1$, so they provide a geometry-dependent, learned baseline partition of the prescribed charge Q.

Proposition 3.1 (Orientation independence and exact charge conservation)**.** *For charges defined by Eqs. 12 and 13, q is independent of the arbitrary bond-state orientations and satisfies $\sum_i q_i = Q$ for every geometry and every parameter value. In the neutral-only construction $q^{\mathrm{ref}} = 0$, the same statement holds with $Q = 0$.*

The result follows from $B\Delta\Delta\tau = B\tau$ and $\mathbf{1}^\top B = 0$; a proof is given in Appendix C.7.

3.4 Invariant energy readout

The bond branch provides a residual potential that complements, rather than replaces, the geometric backbone. Let R collect all Cartesian coordinates and let α control the correction scale:

$$E_\theta(Z, R, Q) = E_{\mathrm{geo}}(Z, R, Q) + \alpha\, \Delta E_{\textsc{BondFF}}(Z, R, Q), \tag{14}$$

where α is a hyperparameter. We decompose the correction into bond, charge-response, and screened electrostatic contributions with $\Delta E_{\textsc{BondFF}} = E_{\mathrm{bond}} + E_{\mathrm{charge}} + E_{\mathrm{elec}}$.

Bond energy. The bond term gives each bond-state node a direct scalar energy readout $E_{\mathrm{bond}} = \sum_{e\in\mathcal{B}} c_e f_E\Big(s_e^{(L)}, \rho_\epsilon(u_e^{(L)}), b_e, \mathrm{RBF}(r_e)\Big)$. The readout uses only orientation-invariant inputs to ensure the bond energy is orientation invariant. In particular, ρ_ϵ allows the equivariant features to be supervised without making the energy depend on arbitrary reference orientations.

Charge-response energy. The charge vector in Eq. 12 represents a conserved redistribution over the bond-state graph. We assign this redistribution an environment-dependent quadratic response energy: $E_{\mathrm{charge}} = \sum_{i=1}^{N} \Big[f_\chi\Big(h_i^{(L)}\Big)\, q_i + \frac{1}{2}\Big(\epsilon_J + \mathrm{softplus}\Big(f_J\Big(h_i^{(L)}\Big)\Big)\Big)\, q_i^2\Big]$. The first learned map gives an environment-dependent linear preference for charge, while the positive quadratic coefficient penalizes excessive polarization.

Electrostatic energy. Latent charges provide a readout for interactions beyond the local bond environment: R_{LR}: $E_{\mathrm{elec}} = 1/2 \sum_{i\neq j} q_i q_j K_{ij}^{\mathrm{LR}}, K_{ij}^{\mathrm{LR}} = \kappa\, s_{\mathrm{geo}}(r_{ij}) s_{\mathrm{LR}}(r_{ij}) / \sqrt{r_{ij}^2 + \eta_{ij}^2}$. Here, the smooth switching functions suppress short-range contributions and impose a long-range cutoff; $\eta_{ij} > 0$ provides shielding, and κ converts units. Their definitions and the choice of cutoff are given in Appendix C.8.

Our model design follows mainstream equivariant MLFFs design to ensure the invariance of the energy readout and enables gradient force supervision. We provide a detailed discussion on geometric symmetry and energy conservation in Appendix C.9.

3.5 Learning objective and optional supervision

We train against energies and conservative forces, adding curvature supervision when available. For a minibatch $\mathcal{D}$ and its Hessian-labeled subset $\mathcal{D}_H \subseteq \mathcal{D}$, the basic energy–force objective is

$$\mathcal{L}_{\mathrm{EF}} = \frac{w_E}{|\mathcal{D}|} \sum_{a\in\mathcal{D}} \left(\frac{E_\theta^{(a)} - E^{\star(a)}}{N_a}\right)^2 + \frac{w_F}{3\sum_{a\in\mathcal{D}} N_a} \sum_{a\in\mathcal{D}} \sum_{i=1}^{N_a} \left\| F_i^{(a)} - F_i^{\star(a)} \right\|_2^2. \tag{15}$$

If reference Hessians are available for $\mathcal{D}_H \subseteq \mathcal{D}$, curvature can be supervised by $\mathcal{L}_H = w_H/|\mathcal{D}_H| \sum_{a\in\mathcal{D}_H} \left\|H^{(a)} - H^{\star(a)}\right\|_F^2/(3N_a)^2$. Directional Hessian-vector products may replace the full matrix when second-order supervision is too expensive. The complete objective is $\mathcal{L} = \mathcal{L}_{\mathrm{EF}} + \mathcal{L}_H$ with unavailable supervision terms set to zero.

4 EXPERIMENTS

4.1 EXPERIMENTAL SETTINGS

Datasets. We evaluate BONDFF on six complementary molecular benchmarks. rMD17 contains molecules from nonequilibrium configurations (Christensen & Von Lilienfeld, 2020). MD22 tests larger molecular systems with complex bonding structures (Chmiela et al., 2023), while AIMD-Chig (Chignolin) assesses long-range interactions of a 166-atom mini-protein (Wang et al., 2023a). 3BPA tests generalization across temperatures by training on 300 K configurations and evaluating at 300, 600, and 1200 K (Kovács et al., 2021). For reactive configurations, Transition1x contains structures sampled on and around reaction pathways (Schreiner et al., 2022), whereas HORM adds reference Hessians for reactive geometries (Cui et al., 2025). We use the established benchmark splits when available and describe the details in Appendix D.1.

Baselines. The comparison set spans general equivariant geometric MLFFs (NequIP (Batzner et al., 2022), Allegro (Musaelian et al., 2023), MACE (Batatia et al., 2022), ViSNet (Wang et al., 2024b), EquiformerV2 (Liao et al., 2024), LEFTNet (Du et al., 2023), QuinNet (Wang et al., 2023b), and FreeCG (Shao et al., 2025)), edge/directional enhanced MLFFs (DimeNet++ (Gasteiger et al., 2020a), GemNet-OC (Gasteiger et al., 2021), PaiNN (Schütt et al., 2021)), long-range extensions (LSR-MP (Li et al., 2024), Ewald message passing (Kosmala et al., 2023), and Neural P^3M (Wang et al., 2024a)), and models that expose electronic or charge-related information (SpookyNet (Unke et al., 2021a), fourth-generation HDNNP (Ko et al., 2021), CHGNet (Deng et al., 2023), AIMNet2-NSE (Kalita et al., 2026), and eMLP (Cools-Ceuppens et al., 2022)).

Without loss of generality, we use MACE as the default cutoff-based geometric backbone in most experiments for our method. For the backbone ablation study, we evaluate MACE, ViSNet, EquiformerV2, and LEFTNet with and without the BONDFF branch, using the same data split and training budget within each pair. Backbones adapted to our model are removed with two layers to standardize backbone depth across the paired comparisons. Architectural summaries and detailed adaptations appear in Appendix D.2.

Task and metrics. The primary task is supervised potential learning from atomic numbers and Cartesian coordinates: models predict the total energy and atomic forces, and the HORM study additionally evaluates the Cartesian Hessian. We report mean absolute error (MAE) for the within-dataset experiments, using kcal/mol and kcal/(mol·Å) for rMD17, MD22, and Chignolin, and eV, eV/Å, and eV/Å^2 for energy, force, and Hessian results on Transition1x and HORM. We also report inference time per structure, peak accelerator memory, and serialized model size on the dataset Transition1x. Lower values are better for every reported error and efficiency metric.

4.2 RESULTS AND ANALYSIS

Reactive benchmarks and 3BPA temperature shift. We evaluate BONDFF on the reactive HORM and Transition1x benchmarks and on 3BPA, where models are trained on 300 K structures and evaluated on test sets at 300, 600, and 1200 K. Among the models in Table 1, BONDFF has the lowest reported MAE for every HORM and Transition1x target and for 3BPA energy at each test temperature. The 3BPA error increases with test temperature for every model shown, suggesting that the higher-temperature test sets are more challenging under this training protocol. Comparisons across model families are descriptive because implementations and training protocols may differ.

We report paired backbone comparisons in Appendix E.1. The baselines use the default settings from their original publications. For each +BONDFF variant, we replace two backbone layers with two BONDFF layers to keep training cost and model capacity roughly comparable. As shown in Table 4, adding BONDFF lowers energy, force, and Hessian MAE for MACE, EquiformerV2, and LEFTNet on HORM, with and without Hessian supervision. It also lowers energy and force MAE on Transition1x. The gains vary by backbone and target: for example, MACE's HORM Hessian MAE without Hessian supervision decreases only from 1.119 to 1.092 eV/Å^2. Vanilla ViSNet produced NaN errors in these experiments, whereas its augmented variants yielded finite errors, precluding a paired accuracy comparison. Peak memory usage and model size decrease for all four augmented backbones. Inference time increases for three backbones and decreases slightly for EquiformerV2.

Table 1: Reactive benchmark and temperature-shift comparison. Energy, force, and Hessian mean absolute errors (MAE) are reported on HORM; energy and force MAE on Transition1x; and energy MAE on 3BPA. Energy, force, and Hessian are in eV, eV/Å, and eV/Å^2, respectively. For 3BPA, models are trained on 300 K configurations and evaluated at 300, 600, and 1200 K. Boldface marks the lowest value in each column; lower is better.

Baseline category	Model	**HORM** with Hessian			**Transition1x**		**3BPA** (Energy ↓)		
		Energy ↓	Force ↓	Hessian ↓	Energy ↓	Force ↓	300 K	600 K	1200 K
Edge/ directional	GemNet-OC	0.497	0.140	0.544	0.429	0.287	0.135	1.348	3.340
	PaiNN	0.463	0.097	0.401	0.462	0.723	0.209	1.020	3.954
	DimeNet++	0.714	0.081	0.593	0.091	0.561	0.736	2.626	3.943
Electronic/ charge-aware	SpookyNet	0.806	0.514	0.912	0.369	0.492	0.551	2.498	4.719
	4G-HDNNP	0.354	0.485	0.272	0.256	0.453	0.676	1.917	5.325
	CHGNet	0.823	0.802	0.732	0.643	0.476	0.460	1.306	4.410
	AIMNet2-NSE	0.904	0.952	0.488	0.794	0.488	0.628	2.227	3.142
	eMLP	1.297	0.738	1.776	1.215	0.994	1.759	2.506	5.031
BONDFF		**0.129**	**0.051**	**0.097**	**0.063**	**0.081**	**0.129**	**0.629**	**1.903**

Table 2: MD22 and Chignolin comparison. Energy (kcal/mol) and force (kcal/(mol·Å)) mean absolute errors (MAE) are reported on six MD22 systems and Chignolin for the listed MLFFs. Boldface marks the lowest value in each row; lower is better.

System	General MLFF baselines					Long-range baselines			Proposed
	ViSNet	EquiformerV2	MACE	QuinNet	FreeCG	LSR-MP	Ewald	Neural P^{3}M	BONDFF
Energy									
Ac-Ala3-NHMe	0.064	0.055	0.063	0.084	0.051	0.067	0.075	0.072	**0.044**
AT-AT	0.071	**0.067**	0.108	0.144	**0.067**	0.078	0.083	0.071	0.115
AT-AT-CG-CG	0.196	0.197	0.154	0.379	0.254	0.118	0.235	**0.112**	**0.112**
DHA	0.074	0.070	0.135	0.118	0.076	0.090	0.092	0.071	**0.052**
Buckyball catcher	0.508	0.537	0.489	0.563	0.512	**0.319**	0.416	0.354	0.335
Double-walled nanotube	0.800	0.601	1.670	1.810	**0.543**	1.810	1.563	0.775	0.710
Chignolin	2.436	0.669	1.027	1.294	1.092	1.227	0.695	0.924	**0.564**
Force									
Ac-Ala3-NHMe	0.083	0.071	0.088	0.068	0.053	0.094	0.074	0.079	**0.049**
AT-AT	0.081	0.078	0.099	0.069	0.063	0.078	0.082	0.074	**0.031**
AT-AT-CG-CG	0.148	0.139	0.115	0.127	0.125	0.106	0.129	0.099	**0.088**
DHA	0.060	0.055	0.065	0.052	0.051	0.060	0.074	0.068	**0.045**
Buckyball catcher	0.184	0.201	0.085	0.109	0.178	0.103	**0.077**	0.085	0.120
Double-walled nanotube	0.362	0.292	0.277	0.247	0.245	0.339	0.375	0.256	**0.222**
Chignolin	0.372	0.187	0.389	0.362	0.328	0.278	0.261	0.272	**0.153**

Non-equilibrium benchmark: rMD17. We evaluate BONDFF on the nonequilibrium rMD17 benchmark of ten small-molecule trajectories; full results are reported in Appendix E.2. BONDFF achieves the lowest or tied-lowest energy MAE for seven molecules and the lowest force MAE for six molecules. The results verified the effectiveness of our method.

Complex bonding structure and long-range interactions: MD22 and Chignolin. We conducted experiments on six MD22 systems, which span peptides, lipids, nucleic acids, and supramolecular structures, and presented results in Table 2. Across the six MD22 systems, BONDFF achieves the lowest or tied-lowest energy MAE on three and the lowest force MAE on five. Compared with the strongest of the three long-range baselines on each system, it has lower energy MAE on four of the seven systems, ties on AT-AT-CG-CG, and has lower force MAE on six. AT-AT and the buckyball catcher are the energy exceptions; the latter is also the force exception. On Chignolin, the long-range baselines provide a complementary comparison. BONDFF achieves the lowest energy and force MAE among all listed models (0.564 kcal/mol and 0.153 kcal/(mol·Å)).

Cross-dataset generalization from Transition1x. Table 3 evaluates transfer from Transition1x to 3BPA and HORM without target-dataset training. BONDFF has the lowest energy and force RMSE on all seven test sets. On the 3BPA constrained-dihedral sets, its energy RMSE remains between 0.180 and 0.189 eV across $\beta = 120°$–$180°$, while the best baseline rises from 0.411 to 0.498 eV. The thermal force RMSE increases from 300 to 1200 K for every model, but BONDFF retains the lowest error at 1200 K (0.341 eV/Å versus 0.496 eV/Å for the best baseline). The advantage is narrower on HORM: BONDFF reaches 0.511 eV energy RMSE versus 0.541 eV for PaiNN, and 0.263 eV/Å force RMSE versus 0.286 eV/Å for MACE. These results support the cross-dataset generalization capacity for BONDFF among the evaluated models.

Table 3: Cross-dataset generalization. Energy and force root mean squared errors (RMSE) are reported on the 3BPA thermal and constrained-dihedral test sets and the HORM test set for models trained on Transition1x. Energy and force errors are in eV and eV/Å, respectively. Boldface marks the lowest value in each row for each metric; lower is better.

Dataset	sub-dataset	Energy				Force			
		MACE	PaiNN	4G-HDNNP	BONDFF	MACE	PaiNN	4G-HDNNP	BONDFF
3BPA temperature	300 K	0.258	0.707	0.690	**0.104**	0.264	0.267	0.272	**0.199**
	600 K	0.188	1.435	1.959	**0.079**	0.366	0.424	0.454	**0.263**
	1200 K	0.282	1.291	1.635	**0.239**	0.496	0.564	0.545	**0.341**
3BPA Dihedral β	120	0.411	0.461	0.500	**0.183**	0.173	0.217	0.263	**0.133**
	150	0.475	0.523	0.539	**0.180**	0.167	0.209	0.237	**0.112**
	180	0.498	0.504	0.565	**0.189**	0.168	0.237	0.221	**0.105**
HORM		0.609	0.541	0.631	**0.511**	0.286	0.343	0.395	**0.263**

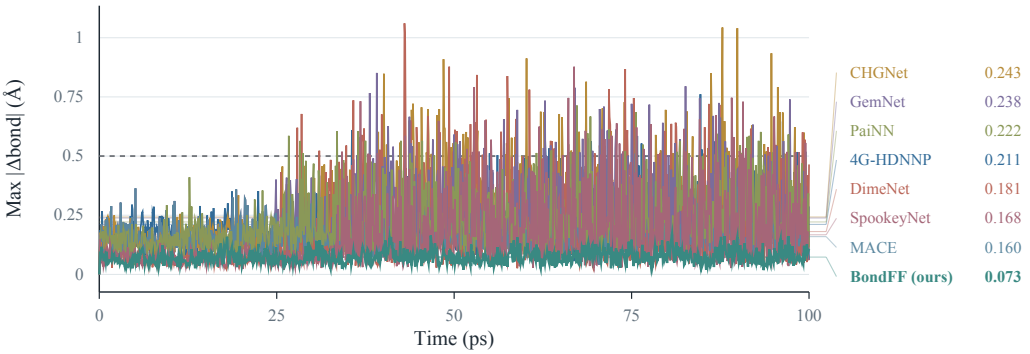

Figure 3: Maximum absolute deviation of a bond length from its initial value during a 100 ps NVT trajectory initialized at 300 K on a held-out Transition1x molecule. The dashed line marks the 0.5 Å bond-deviation threshold; values at right are the means of maximum bond-deviation.

Molecular dynamics stability and energy drift. To assess dynamical stability, we simulate a held-out Transition1x molecule with a five-membered O_0–C_1–N_2–C_3–N_4–O_0 ring and H atoms bonded to C_1 and C_3. Over 100 ps of NVT dynamics at a target temperature of 300 K, the curve labeled BondFF remains below the 0.5 Å bond-deviation threshold and has the lowest mean maximum deviation (0.073 Å, compared with 0.160 Å for MACE); several comparison trajectories cross the threshold (Figure 3). We also examine the total-energy trace supplied as a 300 K NVE run: BondFF has the mean change closest to zero (-0.086 eV) and no clear sustained drift over 100 ps, despite transient fluctuations. In the NVT run, the mean instantaneous temperature (395.5 K) is closest to the 300 K target among the plotted models. The temperature and total-energy traces are shown in Appendix E.3.

Parameter sensitive analysis and ablation study. We conduct a parameter sensitivity analysis of the residual-energy scale α, bond-state depth ℓ_b, and bond-state width $d_u + d_s$, together with an ablation study of the three energy terms and two bond-state structure variants. The parameter study in Appendix E.4 shows little change across the tested α values, whereas reducing bond-state depth to one layer or width to 64 features increases errors; larger settings yield mixed, generally small changes relative to the default. The ablations in Appendix E.5 increase all reported MAEs when any one energy term is omitted or either tested structure variant is used.

5 CONCLUSION

In this paper, we investigate how representations of continuously varying bond states can improve MLFFs. We introduce BONDFF, which comprises a geometric graph with a bipartite bond-incidence graph. Bond-state nodes capture latent pairwise bonding information, while orientation-aware signed flows model charge redistribution and conserve the total charge. We evaluate BONDFF on rMD17, MD22, 3BPA, Chignolin, Transition1x, and HORM, covering non-equilibrium configurations, complex molecular structures, temperature shifts, and reactive geometries. The results demonstrate the value of explicit bond-state modeling for energy and force prediction.

AI USE STATEMENT

We used generative AI tools to assist with coding and script checking, understand existing literature, provide feedback on research methodology and experiments, improve writing and presentation, brainstorm ideas, search for information, and suggest keywords for this paper. We did not use these tools to develop theoretical models or conceptual frameworks; formulate mathematical claims; provide key steps in mathematical proofs; design research methodology or experiments; draft proofs; propose or refine hypotheses; implement methods; translate text; clean or reformat datasets; support qualitative or thematic analysis; or interpret results. Generating synthetic datasets was not applicable to this work. We reviewed all AI-assisted work and manually surveyed the literature to check AI-generated research ideas for overlap with prior work. We take responsibility for the final content, including all text, claims, and artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work develops and evaluates a machine-learned force field using molecular benchmark datasets and computational simulations. It does not involve human participants, personal data, or animal experiments conducted by the authors. The model's predicted energies, forces, and trajectories are subject to the coverage and accuracy of its training data; results outside the evaluated molecular systems and conditions should be validated before use in scientific or safety-critical decisions.

REPRODUCIBILITY STATEMENT

The paper specifies the bond-incidence architecture and learning objective, identifies the molecular benchmarks and evaluation metrics, and describes the paired backbone comparisons and molecular-dynamics analyses. The benchmark datasets are cited in the experimental settings. Exact reproduction of the numerical results additionally requires the dataset partitions, preprocessing, training configurations, random seeds, and implementation details, which are not all provided in the current manuscript.

REFERENCES

Albert P Bartók, Sandip De, Carl Poelking, Noam Bernstein, James R Kermode, Gábor Csányi, and Michele Ceriotti. Machine learning unifies the modeling of materials and molecules. *Science advances*, 3(12):e1701816, 2017.

Ilyes Batatia, David P Kovacs, Gregor Simm, Christoph Ortner, and Gábor Csányi. MACE: Higher order equivariant message passing neural networks for fast and accurate force fields. *Advances in Neural Information Processing Systems*, 35:11423–11436, 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, 2022.

Chi Chen and Shyue Ping Ong. A universal graph deep learning interatomic potential for the periodic table. *Nature Computational Science*, 2(11):718–728, 2022.

Stefan Chmiela, Valentin Vassilev-Galindo, Oliver T Unke, Adil Kabylda, Huziel E Sauceda, Alexandre Tkatchenko, and Klaus-Robert Müller. Accurate Global Machine Learning Force Fields for Molecules with Hundreds of Atoms. *Science Advances*, 9(2):eadf0873, 2023.

Anders S Christensen and O Anatole Von Lilienfeld. On the role of gradients for machine learning of molecular energies and forces. *Machine Learning: Science and Technology*, 1(4):045018, 2020.

Maarten Cools-Ceuppens, Joni Dambre, and Toon Verstraelen. Modeling electronic response properties with an explicit-electron machine learning potential. *Journal of Chemical Theory and Computation*, 18(3):1672–1691, 2022.

Taoyong Cui, Yunhong Han, Haojun Jia, Chenru Duan, and Qiyuan Zhao. A large scale molecular hessian database for optimizing reactive machine learning interatomic potentials. *Scientific Data*, 13(1):37, 2025.

Bowen Deng, Peichen Zhong, KyuJung Jun, Janosh Riebesell, Kevin Han, Christopher J Bartel, and Gerbrand Ceder. Chgnet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nature Machine Intelligence*, 5(9):1031–1041, 2023.

weitao Du, Yuanqi Du, Limei Wang, Dieqiao Feng, Guifeng Wang, Shuiwang Ji, Carla Gomes, and Zhi-Ming Ma. A new perspective on building efficient and expressive 3d equivariant graph neural networks. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 66647–66674. Curran Associates, Inc., 2023. doi: 10.52202/075280-2910.

Johannes Gasteiger, Shankari Giri, Johannes T Margraf, and Stephan Günnemann. Fast and uncertainty-aware directional message passing for non-equilibrium molecules. *arXiv preprint arXiv:2011.14115*, 2020a.

Johannes Gasteiger, Janek Groß, and Stephan Günnemann. Directional message passing for molecular graphs. In *International Conference on Learning Representations*, 2020b.

Johannes Gasteiger, Florian Becker, and Stephan Günnemann. Gemnet: Universal directional graph neural networks for molecules. *Advances in neural information processing systems*, 34:6790–6802, 2021.

Johannes Gasteiger, Muhammed Shuaibi, Anuroop Sriram, Stephan Günnemann, Zachary Ward Ulissi, C. Lawrence Zitnick, and Abhishek Das. Gemnet-OC: Developing graph neural networks for large and diverse molecular simulation datasets. *Transactions on Machine Learning Research*, 2022. ISSN 2835-8856.

Haokai Hong, Wanyu Lin, Zhang Chusong, and Kay Chen Tan. Geometric graph neural diffusion for stable molecular dynamics simulations. In *The Fourteenth International Conference on Learning Representations*, 2026.

Bhupalee Kalita, Roman Zubatyuk, Dylan M Anstine, Maike Bergeler, Volker Settels, Conrad Stork, Sebastian Spicher, and Olexandr Isayev. Aimnet2-nse: A transferable reactive neural network potential for open-shell chemistry. *Angewandte Chemie International Edition*, 65(5):e16763, 2026.

Jisu Kim, Hyunsung Cho, Haekwan Jeon, Jisu Jung, and Seungwu Han. Reactive machine learning interatomic potentials for chemistry and materials science. *Chemical Reviews*, 126(8):4467–4510, 2026.

Tsz Wai Ko, Jonas A Finkler, Stefan Goedecker, and Jörg Behler. A fourth-generation high-dimensional neural network potential with accurate electrostatics including non-local charge transfer. *Nature communications*, 12(1):398, 2021.

Arthur Kosmala, Johannes Gasteiger, Nicholas Gao, and Stephan Günnemann. Ewald-Based Long-Range Message Passing for Molecular Graphs. In *International Conference on Machine Learning*, pp. 17544–17563, 2023.

Dávid Péter Kovács, Cas van der Oord, Jiri Kucera, Alice EA 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, 2021.

Yunyang Li, Yusong Wang, Lin Huang, Han Yang, Xinran Wei, Jia Zhang, Tong Wang, Zun Wang, Bin Shao, and Tie-Yan Liu. Long-Short-Range Message-Passing: A Physics-Informed Framework to Capture Non-Local Interaction for Scalable Molecular Dynamics Simulation. In *The Twelfth International Conference on Learning Representations*, 2024.

Yi-Lun Liao and Tess Smidt. Equiformer: Equivariant Graph Attention Transformer for 3D Atomistic Graphs. In *The Eleventh International Conference on Learning Representations*, 2023.

Yi-Lun Liao, Brandon M Wood, Abhishek Das, and Tess Smidt. EquiformerV2: Improved Equivariant Transformer for Scaling to Higher-Degree Representations. In *The Twelfth International Conference on Learning Representations*, 2024.

Yuanbin Liu, Joe D Morrow, Christina Ertural, Natascia L Fragapane, John LA Gardner, Aakash A Naik, Yuxing Zhou, Janine George, and Volker L Deringer. An automated framework for exploring and learning potential-energy surfaces. *Nature Communications*, 16(1):7666, 2025.

Albert Musaelian, Simon Batzner, Anders Johansson, Lixin Sun, Cameron J Owen, Mordechai Kornbluth, and Boris Kozinsky. Learning local equivariant representations for large-scale atomistic dynamics. *Nature Communications*, 14(1):579, 2023.

Ken Perlin. Improving noise. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, pp. 681–682, 2002.

Vojislava Pophristic and Lionel Goodman. Hyperconjugation Not Steric Repulsion Leads to The Staggered Structure of Ethane. *Nature*, 411(6837):565–568, 2001.

Anthony K Rappe and William A Goddard III. Charge equilibration for molecular dynamics simulations. *The Journal of Physical Chemistry*, 95(8):3358–3363, 1991.

Leonard Reuter and Arne Lüchow. Real space electron delocalization, resonance, and aromaticity in chemistry. *Nature Communications*, 12(1):4820, 2021.

Mathias Schreiner, Arghya Bhowmik, Tejs Vegge, Jonas Busk, and Ole Winther. Transition1x-a dataset for building generalizable reactive machine learning potentials. *Scientific Data*, 9(1):779, 2022.

Kristof Schütt, Oliver Unke, and Michael Gastegger. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *International conference on machine learning*, pp. 9377–9388. PMLR, 2021.

Kristof T Schütt, Farhad Arbabzadah, Stefan Chmiela, Klaus R Müller, and Alexandre Tkatchenko. Quantum-chemical insights from deep tensor neural networks. *Nature communications*, 8(1): 13890, 2017.

Thomas P Senftle, Sungwook Hong, Md Mahbubul Islam, Sudhir B Kylasa, Yuanxia Zheng, Yun Kyung Shin, Chad Junkermeier, Roman Engel-Herbert, Michael J Janik, Hasan Metin Aktulga, et al. The reaxff reactive force-field: development, applications and future directions. *npj Computational Materials*, 2(1):15011, 2016.

Shihao Shao, Haoran Geng, Zun Wang, and Qinghua Cui. Freecg: Free the design space of clebsch-gordan transform for machine learning force fields. In Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu (eds.), *International Conference on Learning Representations*, volume 2025, pp. 5627–5654, 2025.

Suhwan Song, Heejae Kim, Jaehee Jang, Hyuntae Cho, Gunhee Kim, and Geonu Kim. Scalable reactive atomistic dynamics with gaia. *arXiv preprint arXiv:2509.25798*, 2025.

Oliver T Unke, Stefan Chmiela, Michael Gastegger, Kristof T Schütt, Huziel E Sauceda, and Klaus-Robert Müller. Spookynet: Learning force fields with electronic degrees of freedom and nonlocal effects. *Nature communications*, 12(1):7273, 2021a.

Oliver T Unke, Stefan Chmiela, Huziel E Sauceda, Michael Gastegger, Igor Poltavsky, Kristof T Schutt, Alexandre Tkatchenko, and Klaus-Robert Muller. Machine learning force fields. *Chemical reviews*, 121(16):10142–10186, 2021b.

Adri CT Van Duin, Siddharth Dasgupta, Francois Lorant, and William A Goddard. Reaxff: a reactive force field for hydrocarbons. *The Journal of Physical Chemistry A*, 105(41):9396–9409, 2001.

Jonathan Vandermause, Yu Xie, Jin Soo Lim, Cameron J Owen, and Boris Kozinsky. Active learning of reactive bayesian force fields applied to heterogeneous catalysis dynamics of h/pt. *Nature Communications*, 13(1):5183, 2022.

Valdas Vitartas, Hanwen Zhang, Veronika Juraskova, Tristan Johnston-Wood, and Fernanda Duarte. Active learning meets metadynamics: automated workflow for reactive machine learning interatomic potentials. *Digital Discovery*, 5(1):108–122, 2026.

Tong Wang, Xinheng He, Mingyu Li, Bin Shao, and Tie-Yan Liu. AIMD-Chig: Exploring the conformational space of a 166-atom protein Chignolin with ab initio molecular dynamics. *Scientific Data*, 10(1):549, 2023a.

Yusong Wang, Chaoran Cheng, Shaoning Li, Yuxuan Ren, Bin Shao, Ge Liu, Pheng-Ann Heng, and Nanning Zheng. Neural pˆ3m: A long-range interaction modeling enhancer for geometric gnns. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 120336–120365. Curran Associates, Inc., 2024a. doi: 10.52202/079017-3824.

Yusong Wang, Tong Wang, Shaoning Li, Xinheng He, Mingyu Li, Zun Wang, Nanning Zheng, Bin Shao, and Tie-Yan Liu. Enhancing geometric representations for molecules with equivariant vector-scalar interactive message passing. *Nature Communications*, 15(1):313, 2024b.

Zun Wang, Guoqing Liu, Yichi Zhou, Tong Wang, and Bin Shao. Efficiently incorporating quintuple interactions into geometric deep learning force fields. *Advances in Neural Information Processing Systems*, 36:77043–77055, 2023b.

Changxi Yang, Chenyu Wu, Wenbo Xie, Daiqian Xie, and P Hu. General reactive element-based machine learning potentials for heterogeneous catalysis. *Nature Catalysis*, 8(9):891–904, 2025.

Tom A Young, Tristan Johnston-Wood, Volker L Deringer, and Fernanda Duarte. A transferable active-learning strategy for reactive molecular force fields. *Chemical science*, 12(32):10944–10955, 2021.

Hanwen Zhang, Veronika Juraskova, and Fernanda Duarte. Modelling chemical processes in explicit solvents with machine learning potentials. *Nature Communications*, 15(1):6114, 2024a.

Shuhao Zhang, Małgorzata Z Makoś, Ryan B Jadrich, Elfi Kraka, Kipton Barros, Benjamin T Nebgen, Sergei Tretiak, Olexandr Isayev, Nicholas Lubbers, Richard A Messerly, et al. Exploring the frontiers of condensed-phase chemistry with a general reactive machine learning potential. *Nature Chemistry*, 16(5):727–734, 2024b.

Shuhao Zhang, Roman Zubatyuk, Yinuo Yang, Adrian Roitberg, and Olexandr Isayev. Ani-1xbb: an ani-based reactive potential for small organic molecules. *Journal of chemical theory and computation*, 21(9):4365, 2025.

Appendix Contents

A Related Work **15**
A.1 Equivariant machine learning force fields . . . 15
A.2 Classical Reactive Force Fields . . . 16
A.3 Representing Chemical Bonds and Charges in Machine Learning Force Fields . . . 16

B Notation **17**

C Implementation Details **18**
C.1 Preprocess . . . 18
C.2 Directed endpoint encoder and radial basis . . . 18
C.3 Expanded incidence-propagation maps . . . 19
C.4 Expanded gated bond-state update . . . 19
C.5 Readout-head parameterization . . . 19
C.6 Physical intuition for the energy terms . . . 20
C.7 Charge conservation . . . 21
C.8 Electrostatic cutoff . . . 21
C.9 Symmetry and conservative . . . 21

D Experimental Details **21**
D.1 Datasets . . . 21
D.2 Baselines . . . 22

E Experiments **23**
E.1 General MLFFs comparison . . . 23
E.2 Results on rMD17 Dataset . . . 23
E.3 NVT Temperature and NVE Energy Diagnostics . . . 24
E.4 Parameter sensitivity analysis . . . 25
E.5 Ablation study . . . 25

F Broader impact. **26**

G Limitations. **26**

A Related Work

A.1 Equivariant machine learning force fields

Modern machine learning force fields (MLFFs) or machine learning interatomic potentials (MLIPs) have advanced rapidly by combining local atomic environments with symmetry-preserving neural architectures. Early representation-learning work showed that neural atomistic models can recover chemically meaningful structure-property relations from coordinates and elements alone (Schütt et al., 2017), while SOAP/GAP-style descriptors established a data-efficient route for unifying molecular and materials modeling (Bartók et al., 2017). Recent equivariant architectures make the

symmetry bias explicit: PaiNN couples scalar and vector channels for tensorial molecular properties (Schütt et al., 2021), NequIP demonstrates data-efficient E(3)-equivariant force fields (Batzner et al., 2022), Allegro reduces the cost of local equivariant message passing (Musaelian et al., 2023), and MACE uses higher-order equivariant messages to obtain fast and accurate many-body force fields (Batatia et al., 2022). Universal potentials such as M3GNet further show that graph-based MLIPs can cover broad regions of the periodic table when trained at scale (Chen & Ong, 2022).

BONDFF augments equivariant MLFF backbones rather than replacing their geometric message passing. In atom-centered models such as MACE, cutoff edges encode relative geometry for constructing atomic representations, while changes in bonding are learned implicitly from the training data (Batatia et al., 2022). BONDFF adds a separate bipartite bond-incidence graph in which candidate atom pairs are represented by bond-state nodes with orientation-invariant features and signed charge-transfer features. These latent states contribute to a bond-energy correction and to atomic charge redistribution that conserves the prescribed total charge by construction. Accordingly, the controlled comparison pairs each backbone with and without the BONDFF branch under the same data split and training budget.

A.2 CLASSICAL REACTIVE FORCE FIELDS

Classical force fields enable reactive molecular dynamics at scales inaccessible to direct electronic-structure calculations. Conventional force fields assume fixed connectivity, whereas ReaxFF uses distance-dependent bond orders together with coordination, angle, torsion, van der Waals, and Coulombic terms to describe bond breaking and formation across molecular, condensed, interfacial, and multiphase environments (Van Duin et al., 2001; Senftle et al., 2016). Its electrostatics commonly employs QEq, which obtains environment-dependent partial charges by minimizing an approximate electrostatic energy subject to total-charge conservation (Rappe & Goddard III, 1991). These models expose chemically meaningful bond orders and charges, but their empirical functional forms, charge partitioning, and system-specific parameterization can limit transferability, especially for unusual coordination, barriers, polarization, and charge states (Senftle et al., 2016).

A.3 REPRESENTING CHEMICAL BONDS AND CHARGES IN MACHINE LEARNING FORCE FIELDS

For systems in which local bonding changes, an MLFF must capture the associated variation in energy and force across configurations, including bond stretching, conformational distortion, and bond breaking or formation. Depending on the system, these changes may also involve charge redistribution, different spin states, or nonlocal interactions. Relevant work spans electronic-state representation, sampling of challenging configurations, and training strategies; both model design and coverage of the potential-energy surface matter (Unke et al., 2021b; Kim et al., 2026).

Representing electronic degrees of freedom. SpookyNet conditions on total charge and spin and incorporates nonlocal interactions and analytical corrections (Unke et al., 2021a). Fourth-generation HDNNPs combine environment-dependent atomic electronegativities with charge equilibration to model nonlocal charge transfer (Ko et al., 2021). AIMNet2-NSE uses spin–charge equilibration for molecules with varying charge and spin multiplicity, while eMLP represents electron pairs as explicit particles to model electronic response properties (Kalita et al., 2026; Cools-Ceuppens et al., 2022). These models make different choices about which electronic quantities to represent and how to couple them to energy prediction. In contrast, BONDFF introduces bond-state nodes for candidate atom pairs, with latent bonding features and signed charge-transfer flows; its incidence-based charge readout conserves the prescribed total charge by construction.

Active learning and dataset sampling. ANI-1xBB uses off-equilibrium configurations generated by a bond-breaking workflow, ANI-1xnr draws on automated sampling of condensed-phase reactions, and autoplex automates structure exploration and iterative potential fitting for materials (Zhang et al., 2025; 2024b; Liu et al., 2025). Hierarchical active learning with GAP and metadynamics-based active learning sample challenging molecular configurations, while Bayesian on-the-fly learning for H/Pt and descriptor-based selection address catalytic and explicit-solvent settings, respectively (Young et al., 2021; Vitartas et al., 2026; Vandermause et al., 2022; Zhang et al., 2024a). Element-based potentials trained on diverse interactions address heterogeneous catalysis,

GAIA automates the generation of varied atomic configurations, and HORM provides Hessian data and a training strategy for transition-state modeling (Yang et al., 2025; Song et al., 2025; Cui et al., 2025). These efforts expand training coverage or provide curvature information. They complement BONDFF's focus on representing continuously varying pairwise bond states within an MLFF.

B NOTATION

Symbol	Description
$N,\ i, j, k$	Number of atoms and atom indices.
$Z_i,\ Z$	Atomic number of atom i and the collection of atomic numbers.
$r_i,\ R$	Cartesian position of atom i and all coordinates.
$r_{ij},\ r_e,\ \hat{r}_{ij}$	Pairwise distance, the distance for $e = \{i, j\}$, and the unit vector from i to j.
$Q,\ \zeta$	Prescribed total charge and optional spin-state conditioning variable.
$E^\star,\ F_i^\star,\ H^\star$	Reference energy, force, and Cartesian Hessian.
$E_\theta,\ F_i,\ H$	Predicted energy, conservative force, and Cartesian Hessian.
$\Xi,\ \Xi^\star$	Predicted and reference virial or stress tensor.
θ	Trainable model parameters.
$O \in O(3)$	Orthogonal spatial transformation.
$\epsilon,\ \epsilon_J$	Positive smoothing and hardness constants.
$\mathcal{D},\ \mathcal{D}_H$	Training minibatch and its subset with Hessian labels.

Symbol	Description
$\mathcal{G}_{\text{geo}}$	Geometric message-passing graph used by the backbone.
$\mathcal{V},\ \mathcal{E}_{\text{geo}}$	Atom set and geometric edge set.
$\mathcal{B},\ M$	Unordered bond-state node set and its size $M = \lvert\mathcal{B}\rvert$.
$e = \{i, j\}$	An unordered bond-state node; membership does not assert chemical bonding.
$t(e),\ h(e)$	Tail and head under the arbitrary reference orientation of e.
$\mathcal{G}_{\text{BI}},\ \mathcal{I}$	Bipartite bond-incidence graph and its atom–bond incidences.
$R_{\text{bond}}^{(Z_i,Z_j)}$	Element-dependent bond cutoff.
$x_e,\ c_e$	Normalized bond distance and compact quintic C^2 support envelope.
$B,\ B_{ke}$	Signed atom–bond incidence matrix and one of its entries.
$C = \lvert B\rvert$	Unsigned atom–bond incidence matrix.
$\Delta,\ \delta_e$	Diagonal bond-reorientation matrix and its $\{-1, +1\}$ entries.
$\bar{B},\ \bar{C}$	Degree-normalized signed and unsigned incidence matrices.

Symbol	Description
$\ell,\ L$	Intermediate layer index and final bond-state layer.
$h_i^{(\ell)},\ \mathbf{H}^{(\ell)}$	Atomic node features and the stacked atomic features at layer ℓ.
$H_e^{(\ell)}$	Bond-state features $(s_e^{(\ell)}, u_e^{(\ell)})$.
$s_e^{(\ell)},\ S^{(\ell)}$	Orientation-invariant bond feature and its stacked matrix.
$u_e^{(\ell)},\ U^{(\ell)}$	Orientation-equivariant bond feature and its stacked matrix.
$d_a,\ d_s,\ d_u,\ d$	Atomic, invariant-bond, equivariant-bond, and generic hidden widths.
$d_{i\to j}^{(0)}$	Endpoint encoding used to initialize a bond-state node.
ϕ_{init}	Shared directed endpoint encoder.
$\Phi_{ss}, \Phi_{uu}, \Phi_{us}, \Phi_{su}$	Learned atom-domain maps for within- and cross-sector propagation.
$\mathcal{T}_{s\to s}, \mathcal{T}_{u\to u}$	Within-sector factored incidence maps.
$\mathcal{T}_{u\to s}, \mathcal{T}_{s\to u}$	Cross-sector factored incidence maps.
ρ_ϵ	Smooth elementwise even contraction of an equivariant feature.
$z_{u\to s,e},\ z_{s\to u,e}$	Same-bond invariant and equivariant fusion features.
$W_{(\cdot)},\ b_g$	Learned linear maps and the bias used only in the invariant gate.
$g_s,\ g_u,\ \text{Norm}_{s,u}$	Radial or invariant gates and parity-compatible normalizations.
$a_i^{(\ell)},\ \Psi_a,\ \phi_a$	Invariant atom-domain aggregate, aggregate map, and atomic update map.
σ, SiLU, tanh, $\odot$	Sigmoid, invariant-sector and odd-sector nonlinearities, and element-wise multiplication.
$\text{RBF}(r_e)$	Radial-basis expansion of a bond distance.

Symbol	Description
b_e	Latent continuous bond strength.
$\tau_e,\ \tau,\ \tau_{\max}$	Oriented bond charge flow, stacked flow vector, and scale bound.
$\tau_{\mathrm{cycle}},\ \ker(B)$	Unobservable cycle flow and the null space of the signed incidence matrix.
w_τ	Invariant vector-valued weight in the odd flow readout.
$q_i,\ q,\ q_i^{\mathrm{ref}},\ \mathbf{1}$	Atomic charge, stacked charge vector, reference charge allocation, and all-ones vector.
$\pi_i,\ f_Q$	Normalized total-charge allocation and its learned invariant score.
$\varphi_i,\ w_e$	Learned atomic potential and nonnegative invariant flow weight.
$\mu_e,\ \mu_{\mathrm{mol}},\ r_0$	Orientation-independent bond dipole, molecular dipole, and dipole origin.
$E_{\mathrm{geo}},\ \Delta E_{\mathrm{BondFF}}$	Cutoff-based geometric graph energy term and bond-state incidence graph energy term.
$E_{\mathrm{bond}}, E_{\mathrm{charge}}, E_{\mathrm{elec}}$	Local bond, charge response, and optional electrostatic energy terms.
α	Scale of the bond energy.
f_b, f_E, f_χ, f_J	Learned bond-strength, local-energy, charge-coefficient, and hardness maps.
$\widetilde{\chi}_i,\ J_i$	Learned linear charge coefficient and positive hardness.
$K(r)$	Smooth electrostatic interaction kernel.
$\mathcal{L}, \mathcal{L}_{\mathrm{EF}}, \mathcal{L}_H$	Total, energy–force, and Hessian objectives.
w_E, w_F, w_H	Weights of the energy, force, and Hessian loss terms.
λ_τ	Low-norm bond-flow regularization weight.

C Implementation Details

C.1 Preprocess

To keep the bond-state readouts local and smooth near the cutoff, we scale each pair contribution by a distance-dependent factor. For $e = \{i, j\}$, let $r_e = \|r_i - r_j\|_2$, $x_e = r_e / R_{\mathrm{bond}}^{(Z_i, Z_j)}$, and $c_e = c(x_e)$. We set $c(x) = 1 - 10x^3 + 15x^4 - 6x^5$ for $0 \le x < 1$ and $c(x) = 0$ for $x \ge 1$. The function and its first two derivatives reach zero at the cutoff, so the contributions vary smoothly for the following network readout(Perlin, 2002).

To represent signed charge flow without making predictions depend on an arbitrary pair direction, we assign each bond-state node e a reference direction from $t(e)$ to $h(e)$ with $t(e), h(e) = i, j$. This direction only fixes the sign convention. With $M = |\mathcal{B}|$, define $B, C \in \mathbb{R}^{N \times M}$ by $B_{ke} = -1$ when $k = t(e)$, $B_{ke} = +1$ when $k = h(e)$, and $B_{ke} = 0$ otherwise; the unsigned matrix is $C = |B|$. Reversing any set of reference directions gives $B' = B\Delta$ and $C' = C$, where $\Delta = \mathrm{diag}(\delta_1, \ldots, \delta_M)$ and each $\delta_e \in \{-1, +1\}$.

C.2 Directed endpoint encoder and radial basis

Our proposed bipartite bond-incidence graph consumes invariant atomic features so that its scalar bond states do not depend on the spatial frame. For a backbone with irreducible $E(3)$ features, we retain the scalar channels of its final atom representation and project them to the interface width d_a,

$$\bar{h}_i^{\mathrm{geo}} = \mathrm{SiLU}\big(W_a h_{i,0e}^{\mathrm{geo}} + b_a\big)\,. \tag{16}$$

Here $h_{i,0e}^{\mathrm{geo}}$ denotes the invariant scalar block.

We expand each bond distance with $K_{\mathrm{RBF}} = 32$ Gaussian radial functions, following the uniformly spaced Gaussian construction used for atomistic message passing in eMLP (Cools-Ceuppens et al., 2022). Because the bond-state support is element dependent, the centers and width are scaled by the corresponding support radius $R_e \equiv R_{\mathrm{bond}}^{(Z_i, Z_j)}$:

$$\begin{aligned} \mathrm{RBF}_n(r_e) &= \exp\left[-\frac{(r_e - \mu_{en})^2}{2\sigma_e^2}\right], \qquad n = 0, \ldots, K_{\mathrm{RBF}} - 1, \\ \mu_{en} &= \frac{n}{K_{\mathrm{RBF}} - 1} R_e, \qquad\qquad \sigma_e = \frac{R_e}{K_{\mathrm{RBF}}}. \end{aligned} \tag{17}$$

Thus, every element pair uses the same number of radial channels over its cutoff range, $r_e/R_e \in [0, 1]$. The radial basis encodes distance, while the cutoff factor c_e is also supplied to the encoder and scales the latent bond-strength, charge-flow, and direct bond-energy readouts.

For each endpoint order, the encoder concatenates the two atom-interface features, the radial basis, and the scalar envelope,

$$\begin{aligned} x_{i\to j} &= \left[\bar{h}_i^{\mathrm{geo}} \,\|\, \bar{h}_j^{\mathrm{geo}} \,\|\, \mathrm{RBF}(r_{ij}) \,\|\, c_e\right], \\ \phi_{\mathrm{init}}(x_{i\to j}) &= W_{\mathrm{init},2}\ \mathrm{SiLU}(W_{\mathrm{init},1} x_{i\to j} + b_{\mathrm{init},1}) + b_{\mathrm{init},2}. \end{aligned} \tag{18}$$

The same parameters are used for both endpoint orders and for every bond-state node. The output has width $d_0 = \max(d_s, d_u)$, and the hidden affine layer has width $2d_0$. No normalization or dropout is applied inside ϕ_{init}. The projections $P_s \in \mathbb{R}^{d_s \times d_0}$ and $P_u \in \mathbb{R}^{d_u \times d_0}$ in Eq. 5 are learned independently; P_u has no bias, whereas the even projection may use b_s. Consequently, unequal even and odd channel widths do not alter the endpoint-exchange symmetry.

C.3 Expanded incidence-propagation maps

For completeness, this appendix expands the compact propagation rule in Eq. 6. Let $S^{(\ell)}$ and $U^{(\ell)}$ be the stacked even and odd bond features, respectively. The four maps are

$$\mathcal{T}^{(\ell)}_{s\to s}(S) = C^\top \Phi^{(\ell)}_{ss}(CS, \mathbf{H}^{(\ell)}), \qquad \mathcal{T}^{(\ell)}_{u\to u}(U) = B^\top \Phi^{(\ell)}_{uu}(BU, \mathbf{H}^{(\ell)}), \tag{19}$$

$$\mathcal{T}^{(\ell)}_{u\to s}(U) = C^\top \Phi^{(\ell)}_{us}(BU, \mathbf{H}^{(\ell)}), \qquad \mathcal{T}^{(\ell)}_{s\to u}(S) = B^\top \Phi^{(\ell)}_{su}(CS, \mathbf{H}^{(\ell)}). \tag{20}$$

The notation $\Phi^{(\ell)}_{ab}$ means an atom-domain neural map whose input is the atom aggregate of sector a and whose output is a message in sector b. For example, $\Phi^{(\ell)}_{us}$ takes the atom aggregate BU of the odd bond features and produces an even atom message; the final multiplication by $C^\top$ scatters that message back to every incident bond. Thus, $\mathcal{T}_{s\to s}$ and $\mathcal{T}_{u\to u}$ are within-sector maps, whereas $\mathcal{T}_{u\to s}$ and $\mathcal{T}_{s\to u}$ are cross-sector maps. Under a reorientation $B' = B\Delta$, $U' = \Delta U$, and $S' = S$, the atom aggregates satisfy $B'U' = BU$ and $C'S' = CS$. Consequently, maps scattered with $C^\top$ are even, while maps scattered with $B'^\top = \Delta B^\top$ are odd. This proves that all four maps preserve the required parity, including the cross-sector maps.

C.4 Expanded gated bond-state update

The compact updates in Eq. 9 expand to

$$\begin{aligned} \widetilde{s}_e^{(\ell+1)} &= W_s^{(\ell)} s_e^{(\ell)} + m_{s,e}^{(\ell)} + g_s^{(\ell)}(\mathrm{RBF}(r_e), c_e), \\ s_e^{(\ell+1)} &= s_e^{(\ell)} + \mathrm{SiLU}\Big(\mathrm{Norm}_s(\widetilde{s}_e^{(\ell+1)})\Big), \end{aligned} \tag{21}$$

$$\begin{aligned} \widetilde{u}_e^{(\ell+1)} &= W_o^{(\ell)} u_e^{(\ell)} + m_{u,e}^{(\ell)}, \\ u_e^{(\ell+1)} &= u_e^{(\ell)} + \tanh\Big(\mathrm{Norm}_u(\widetilde{u}_e^{(\ell+1)})\Big) \odot \sigma\Big(g_u^{(\ell)}(s_e^{(\ell+1)})\Big). \end{aligned} \tag{22}$$

Here, the odd-sector affine maps and normalization are bias-free, and $\tanh$ is applied elementwise as an odd nonlinearity. The gate depends only on the invariant state, so the two updates preserve their respective parity.

C.5 Readout-head parameterization

We have three energy term readouts:

1. Bond energy:

$$E_{\mathrm{bond}} = \sum_{e\in\mathcal{B}} c_e f_E\Big(s_e^{(L)}, \rho_\epsilon(u_e^{(L)}), b_e, \mathrm{RBF}(r_e)\Big). \tag{23}$$

2. Charge-response energy:

$$E_{\mathrm{charge}} = \sum_{i=1}^{N} \left[f_\chi\Big(h_i^{(L)}\Big)\, q_i + \frac{1}{2}\left(\epsilon_J + \mathrm{softplus}\Big(f_J\Big(h_i^{(L)}\Big)\Big)\right) q_i^2 \right], \tag{24}$$

3. Electrostatic energy:

$$E_{\text{elec}} = \frac{1}{2}\sum_{i\neq j} q_i q_j K_{ij}^{\text{LR}}, \qquad K_{ij}^{\text{LR}} = \frac{\kappa\, s_{\text{geo}}(r_{ij}) s_{\text{LR}}(r_{ij})}{\sqrt{r_{ij}^2 + \eta_{ij}^2}}, \tag{25}$$

All readout networks use multilayer perceptrons with a SiLU hidden activation and a linear output layer. For an input $x \in \mathbb{R}^p$, we write

$$\text{MLP}_{p\to q}^{(h)}(x) = W_2\,\text{SiLU}(W_1 x + b_1) + b_2, \tag{26}$$

where h is the hidden width and q the output width. The bond heads use $h = 2\max(d_s, d_u)$, and the atom heads use $h = 2d_a$. Heads have separate parameters, no dropout, and no normalization layers. Biases are permitted in heads whose inputs and outputs are orientation invariant.

The bond-strength and local-energy heads in Eqs. 10 and 23 are

$$f_b = \text{MLP}_{d_s+d_u+K_{\text{RBF}}\to 1}\Big([s_e^{(L)} \| \rho_\epsilon(u_e^{(L)})\| \,\text{RBF}(r_e)]\Big), \tag{27}$$

$$f_E = \text{MLP}_{d_s+d_u+1+K_{\text{RBF}}\to 1}\Big([s_e^{(L)} \| \rho_\epsilon(u_e^{(L)})\| b_e \| \,\text{RBF}(r_e)]\Big). \tag{28}$$

The external softplus and envelope in Eq. 10, rather than the final layer of f_b, enforce nonnegativity and cutoff compatibility. The energy head has an unconstrained scalar output because its sign is learned from energy and force supervision.

The charge-flow weight is an invariant vector-valued head,

$$w_\tau = \text{MLP}_{d_s+K_{\text{RBF}}+1\to d_u}\Big([s_e^{(L)} \| \,\text{RBF}(r_e)\| c_e]\Big). \tag{29}$$

All inputs to w_τ are orientation invariant. Therefore $w_\tau^\top u_e^{(L)}$ changes sign only through $u_e^{(L)}$, and the odd flow law in Eq. 11 is preserved without imposing an odd nonlinearity on the weight network.

Finally, f_Q, f_χ, and f_J are independent instances of $\text{MLP}_{d_a\to 1}^{(2d_a)}$ applied to $\bar{h}_i^{\text{geo}}$. The charge-allocation softmax in Eq. 13 is evaluated separately for each molecular graph in a minibatch. The f_χ output is unconstrained, while the external transformation $\epsilon_J + \text{softplus}(f_J)$ enforces positive hardness. These choices ensure that every atomic readout is permutation equivariant and spatially invariant.

C.6 Physical intuition for the energy terms

The main text uses learned readouts rather than imposing the following analytic forms. The bond-energy readout is motivated by the bounded, saturating behavior of a Morse potential,

$$V_{\text{Morse}}(r) = D_e\big[1 - \exp\big(-a(r - r_e)\big)\big]^2 - D_e, \tag{30}$$

where D_e, a, and r_e denote a dissociation scale, inverse length scale, and reference distance. In terms of the direct parameterization in Eq. 24, the shorthand coefficients used below are $\widetilde{\chi}_i = f_\chi(h_i^{(L)})$ and $J_i = \epsilon_J + \text{softplus}(f_J(h_i^{(L)}))$. The charge-response term is consistent with a local electronegativity–hardness picture, for which

$$\frac{\partial E_{\text{charge}}}{\partial q_i} = \widetilde{\chi}_i + J_i q_i, \qquad q_i^{\text{eq}} = -\frac{\widetilde{\chi}_i}{J_i} \quad \text{(isolated on-site response)}, \tag{31}$$

and $\partial^2 E_{\text{charge}}/\partial q_i^2 = J_i > 0$. These relations provide intuition for the parameterization but do not make the predicted charges unique or turn the model into a complete QEq solver. The electrostatic readout directly motivated by Coulomb's law:

$$E_{\text{elec}}^{\text{Coul}} = \frac{1}{2}\sum_{i\neq j} \frac{\kappa q_i q_j}{r_{ij}}. \tag{32}$$

C.7 CHARGE CONSERVATION

Proof. The reference allocation depends only on invariant atomic features and is therefore unchanged by bond-state node reorientation. Under reorientation, $B' = B\Delta$ and $\tau' = \Delta\tau$, with $\Delta^2 = I$, so $B'\tau' = B\Delta\Delta\tau = B\tau$ and hence $q' = q$. Each column of B contains one -1 and one $+1$, so $\mathbf{1}^\top B = 0$. Moreover, $\sum_i \pi_i = 1$ implies $\mathbf{1}^\top q^{\mathrm{ref}} = Q$, and therefore $\mathbf{1}^\top q = \mathbf{1}^\top q^{\mathrm{ref}} + \mathbf{1}^\top B\tau = Q$. □

On a graph with cycles, $B\tau$ does not identify the individual flows: any cycle flow $\tau_{\mathrm{cycle}} \in \ker(B)$, where $\ker(B)$ is the null space of B, leaves q unchanged. A weak ℓ_2 penalty can select a low-norm representative. Alternatively, the more restrictive parameterization

$$\tau_e = c_e w_e\big(\varphi_{t(e)} - \varphi_{h(e)}\big), \qquad w_e = w_e(s_e, \rho_e(u_e), r_e) \geq 0, \tag{33}$$

reduces arbitrary circulation. We treat direct flow versus potential-difference flow as an ablation because the latter may reduce expressivity. Neither construction makes partial charges unique or guarantees long-range charge equilibration between disconnected fragments.

C.8 ELECTROSTATIC CUTOFF

The electrostatic correction uses a smooth inner switch s_{geo} to reduce overlap with the backbone's short-range interactions and a compact quintic outer switch,

$$s_{\mathrm{LR}}(r) = \begin{cases} 1 - 10y^3 + 15y^4 - 6y^5, & y = r/R_{\mathrm{LR}} < 1, \\ 0, & y \geq 1, \end{cases} \tag{34}$$

which makes the kernel and its first two derivatives vanish at R_{LR}. We treat R_{LR} as a fixed accuracy–cost hyperparameter, selected by validation convergence of energies and forces rather than as a learned chemical length scale. The resulting correction is finite-ranged; periodic systems requiring unscreened long-range electrostatics should instead use Ewald, particle-mesh, or a related long-range solver.

C.9 SYMMETRY AND CONSERVATIVE

Proposition C.1 (Symmetry and conservative derivatives)**.** *Assume that the backbone energy is invariant, all maps are shared over atomic and bond-state nodes, and the energy heads receive only the invariant quantities specified above. Then E_θ is invariant to translations, $O(3)$ transformations, permutations of equivalent atoms, and bond-state reorientation. Its derivatives*

$$F_i = -\nabla_{r_i} E_\theta, \qquad H = \nabla_R^2 E_\theta = -\nabla_R F \tag{35}$$

form a conservative, spatially equivariant force field and its symmetric Cartesian Hessian.

The symmetry result follows from shared maps and sum aggregation, spatially invariant inputs to the energy terms, and the fact that $B\tau$ is unchanged when a reference direction and its signed flow are reversed together. The cutoff c_e and its first two derivatives vanish at the bond-state boundary.

D EXPERIMENTAL DETAILS

D.1 DATASETS

We evaluate energy and force prediction on molecular dynamics and reactive-geometry benchmarks, and additionally evaluate Cartesian Hessians on HORM. The datasets probe different parts of the potential-energy surface: rMD17 contains trajectories of small molecules, MD22 and AIMD-Chig contain larger flexible systems, and Transition1x and HORM emphasize configurations relevant to chemical reactions. The dataset-specific protocols are given below.

rMD17 and MD22. The revised MD17 (rMD17) benchmark provides recalculated energies and forces for molecular-dynamics trajectories of ten small organic molecules (Christensen & Von Lilienfeld, 2020). We evaluate the ten molecules separately using the established rMD17 training and test partitions, with 1,000 configurations in each partition per molecule, as in the MACE benchmark (Batatia et al., 2022). MD22 contains ab initio trajectories of substantially larger, flexible molecular systems, including peptides, nucleic-acid assemblies, a lipid, and supramolecular complexes (Chmiela et al., 2023). We report results for the six MD22 systems shown in Table 2: Ac-Ala3-NHMe, AT-AT, AT-AT-CG-CG, DHA, the buckyball catcher, and the double-walled nanotube. For rMD17 and MD22, we follow the established per-system benchmark partitions used for the FreeCG comparisons, select checkpoints on validation data, and evaluate once on the held-out test data (Shao et al., 2025). Energy and force errors are mean absolute errors in kcal/mol and kcal/(mol·Å), respectively; the MD22 energy error is computed for the total molecular energy, without division by atom count (Shao et al., 2025).

AIMD-Chig. The AIMD-Chig dataset samples folded, unfolded, and metastable conformations of the 166-atom Chignolin mini-protein using ab initio molecular dynamics (Wang et al., 2023a). We use it as a separate large-system energy-and-force benchmark, denoted Chignolin in Table 2.

3BPA. Following the MACE 3BPA protocol (Batatia et al., 2022; Kovács et al., 2021), the training set contains 500 geometries of 3-(benzyloxy)pyridin-2-amine sampled from a 300 K molecular-dynamics trajectory. Separate thermal test sets sampled at 300, 600, and 1200 K assess accuracy within and beyond the training-temperature distribution. The dataset also provides constrained-dihedral scans through regions of the potential-energy surface away from the thermal training configurations. We report temperature-specific 3BPA energy errors in Table 1; the thermal and dihedral test sets are additionally used for the Transition1x-to-3BPA transfer evaluation in Table 3.

Transition1x and HORM. Transition1x contains approximately 9.6 million DFT energy-and-force calculations for geometries on and near reaction pathways, generated from nudged elastic band calculations on more than 10,000 organic reactions (Schreiner et al., 2022). We use its provided train, validation, and test partitions, which comprise approximately 90%, 5%, and 5% of the data and are disjoint by molecular formula. We select models on the validation partition and report energy and force errors on the test partition. HORM provides reactive geometries with reference energies, forces, and Cartesian Hessians recalculated at a consistent level of theory (Cui et al., 2025). Following its published benchmark protocol, we train on the HORM-Transition1x subset of 1,725,362 geometries from 9,000 reactions and report energy, force, and Hessian errors on the separate 60,000-geometry HORM-RGD1 subset.

Reproducibility. The global random seed is 123. We conduct ten independent runs for each experiment and report the arithmetic mean of the resulting metrics.

D.2 Baselines

All comparisons use atomic numbers and Cartesian coordinates as structural inputs and evaluate total energies and forces; the HORM comparisons additionally evaluate Cartesian Hessians. The baseline set covers several ways of representing molecular geometry and interactions. NequIP (Batzner et al., 2022) and Allegro (Musaelian et al., 2023) use local equivariant representations, while MACE (Batatia et al., 2022) builds higher-order equivariant features through many-body message passing. ViSNet (Wang et al., 2024b) combines vector and scalar geometric features, EquiformerV2 (Liao et al., 2024) uses equivariant attention with higher-degree representations, and LEFTNet (Du et al., 2023) uses local substructure and frame-transition encodings. QuinNet (Wang et al., 2023b) incorporates quintuple geometric interactions, and FreeCG (Shao et al., 2025) explores alternative Clebsch–Gordan feature transformations. They are used across the rMD17, MD22, and paired backbone comparisons as indicated in the main-text tables.

The reactive comparison includes three directional or edge-aware models. PaiNN (Schütt et al., 2021) propagates scalar and vector features, DimeNet++ (Gasteiger et al., 2020a) uses angular directional messages with a more efficient design than DimeNet, and GemNet-OC (Gasteiger et al., 2022) extends directional graph modeling to larger and more diverse atomistic datasets. For MD22 and Chignolin, LSR-MP (Li et al., 2024) couples short-range atomistic messages to a long-range

pathway, Ewald message passing (Kosmala et al., 2023) introduces reciprocal-space interactions, and Neural P^3M (Wang et al., 2024a) combines atom- and mesh-level interactions.

The electronic and charge-related comparisons comprise SpookyNet (Unke et al., 2021a), which conditions on molecular charge and spin and includes nonlocal interactions; fourth-generation HDNNP (Ko et al., 2021), which uses charge equilibration to represent nonlocal charge transfer; CHGNet (Deng et al., 2023), whose published training uses magnetic moments to regularize charge-sensitive features; AIMNet2-NSE (Kalita et al., 2026), which models open-shell systems with spin–charge equilibration; and eMLP (Cools-Ceuppens et al., 2022), which represents electron pairs explicitly. Their original formulations therefore do not all have the same information requirements as geometry-only energy–force learning.

For the reported adapted comparisons, we restrict supervision to the reference quantities available for the task: energies and forces, plus Hessians in the HORM experiments that use Hessian supervision. Auxiliary electronic labels, such as magnetic moments or explicit electron-pair positions, and optional conditioning on externally supplied electronic states are omitted when those data are unavailable; neutral systems use a fixed total charge rather than an additional learned label. We retain geometry-derived latent variables and charge-equilibration operations that can be computed without extra labels. In particular, an eMLP variant without explicit electron-pair information is a restricted adaptation, not a reproduction of the original explicit-electron potential. These modifications place the models on a common input–output task but can change their published capabilities. In the paired backbone ablations, the vanilla and BONDFF-augmented versions of each backbone use the same split and training budget; the augmented backbone has two fewer layers, as specified in the main text.

E EXPERIMENTS

E.1 GENERAL MLFFS COMPARISON

Table 4: Comparison with general-purpose MLFFs. Energy, force, and Hessian mean absolute errors (MAE) on HORM, and energy and force MAE on Transition1x, are reported for backbones with and without the BONDFF branch. Energy, force, and Hessian errors are in eV, eV/Å, and eV/Å^2, respectively. On Transition1x, time denotes inference time (ms/sample), memory denotes peak memory (GB), and size denotes model weight size (MB). Vanilla ViSNet produced NaN errors during training; dashes mark its unavailable accuracy results. Boldface marks the lower value within each available pair or the sole available ViSNet accuracy value. Lower is better.

		MACE	+BONDFF	ViSNet	+BONDFF	Equiformer V2	+BONDFF	LEFTNet	+BONDFF
HORM with Hessian	Energy	0.148	**0.129**	-	**0.386**	0.150	**0.127**	0.234	**0.225**
	Force	0.098	**0.051**	-	**0.288**	0.074	**0.059**	0.139	**0.132**
	Hessian	0.192	**0.097**	-	**0.257**	0.108	**0.094**	0.252	**0.246**
HORM without Hessian	Energy	0.554	**0.489**	-	**0.532**	0.449	**0.414**	0.524	**0.473**
	Force	0.074	**0.051**	-	**0.202**	0.115	**0.103**	0.137	**0.087**
	Hessian	1.119	**1.092**	-	**1.222**	1.232	**1.017**	1.432	**0.975**
Transition1x	Energy	0.071	**0.063**	-	**0.158**	0.034	**0.024**	0.102	**0.087**
	Force	0.083	**0.081**	-	**0.126**	0.076	**0.052**	0.096	**0.086**
	Time	**17.58**	19.14	**3.90**	4.29	34.18	**33.28**	**3.69**	4.27
	Memory	2.05	**1.92**	2.30	**2.19**	5.09	**3.70**	2.19	**2.10**
	Size	68.40	**62.20**	39.26	**38.05**	92.22	**66.40**	62.59	**36.29**

E.2 RESULTS ON RMD17 DATASET

BONDFF attains the lowest or tied-lowest energy MAE on seven of ten molecules and the lowest force MAE on six (Table 5). Other models lead on the remaining molecules, indicating that relative performance depends on the molecule and prediction target.

Table 5: rMD17 comparison. Energy and force mean absolute errors (MAE) are reported on ten rMD17 molecules for the listed MLFFs. Energy and force errors are in kcal/mol and kcal/(mol·Å), respectively. Boldface marks the lowest value in each row; lower is better.

Molecule	NequIP	MACE	Allegro	ViSNet	QuinNet	EquiformerV2	FreeCG	BondFF
Energy (kcal/mol)								
Aspirin	0.0530	0.0507	0.0530	0.0445	0.0486	0.0379	0.0530	**0.0376**
Azobenzene	0.0161	0.0277	0.0277	0.0156	0.0394	**0.0113**	0.0217	**0.0113**
Benzene	0.0009	0.0092	0.0069	**0.0007**	0.0096	0.0010	0.0107	0.0012
Ethanol	0.0092	0.0032	0.0092	0.0078	0.0096	0.0046	0.0087	**0.0030**
Malonaldehyde	0.0184	0.0185	0.0138	0.0132	0.0168	0.0120	0.0146	**0.0113**
Naphthalene	0.0046	0.1153	0.0046	0.0057	0.0174	**0.0043**	0.0118	0.0046
Paracetamol	0.0323	0.0300	0.0346	0.0258	0.0362	**0.0197**	0.0392	0.0200
Salicylic acid	0.0161	0.0208	0.0208	0.0161	0.0330	0.0117	0.0233	**0.0110**
Toluene	0.0069	0.0115	0.0092	0.0059	0.0139	0.0048	0.0334	**0.0042**
Uracil	0.0092	0.0115	0.0138	0.0069	0.0149	0.0045	0.0116	**0.0045**
Force kcal/(mol · Å)								
Aspirin	0.1891	0.1522	0.1684	0.1520	0.1429	0.1450	0.1212	**0.1119**
Azobenzene	0.0669	0.0692	0.0600	0.0585	0.0513	0.0532	**0.0486**	0.0522
Benzene	0.0069	0.0069	0.0046	0.0056	0.0047	**0.0022**	0.0056	0.0023
Ethanol	0.0646	0.0484	0.0484	0.0522	0.0516	**0.0340**	0.0438	0.0353
Malonaldehyde	**0.0118**	0.0946	0.0830	0.0893	0.0875	0.0661	0.0802	0.0911
Naphthalene	0.0300	0.0369	0.0208	0.0291	0.0242	0.0201	0.0228	**0.0188**
Paracetamol	0.1361	0.1107	0.1130	0.1029	0.0979	0.0846	0.0840	**0.0833**
Salicylic acid	0.0922	0.0715	0.0669	0.0795	0.0771	0.0648	0.0648	**0.0566**
Toluene	0.0369	0.0350	0.0415	0.0264	0.0244	0.0260	0.0239	**0.0232**
Uracil	0.0669	0.0484	0.0415	0.0495	0.0487	0.0373	0.0446	**0.0354**

E.3 NVT Temperature and NVE Energy Diagnostics

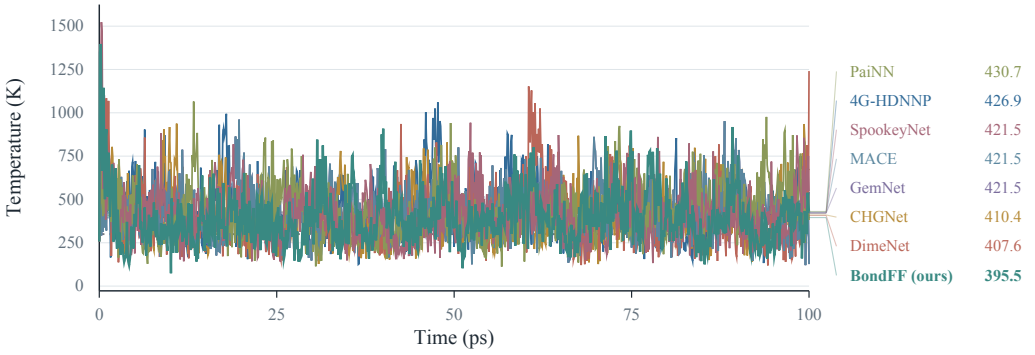

Figure 4: Instantaneous temperature in the corresponding 300 K NVT trajectories. The dashed line marks the target temperature; values at right are trajectory means.

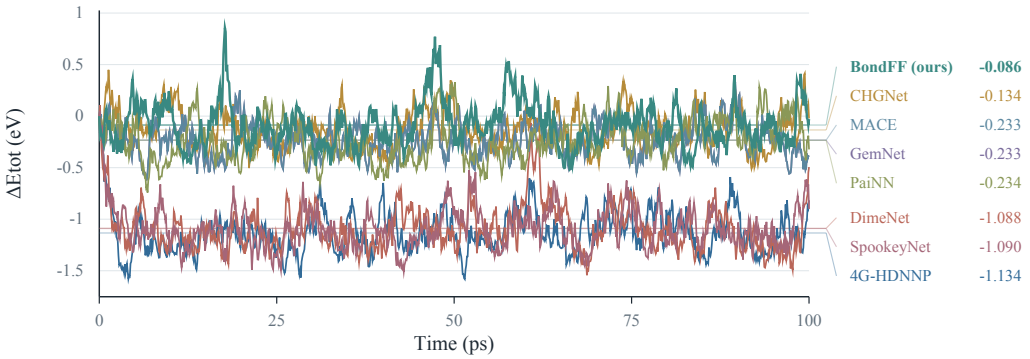

Figure 5: Change in total energy relative to the first saved frame over 100 ps for the Transition1x trajectories specified as 300 K NVE. Values at right are trajectory means in eV.

E.4 Parameter sensitivity analysis

Table 6: Parameter sensitivity analysis. Energy, force, and Hessian mean absolute errors (MAE) are reported on HORM, and energy and force MAE on Transition1x, for variations in the BondFF energy scale, bond-state depth, and bond-state width. Energy, force, and Hessian errors are in eV, eV/Å, and eV/Å^2, respectively. Each parameter is varied individually from the default configuration ($\alpha = 1$, $\ell_b = 2$, $d_u + d_s = 128$). Boldface marks the lowest value in each metric column; lower is better.

Dataset		HORM			Transition1x	
Metrics		Energy	Force	Hessian	Energy	Force
E_{BondFF}	0.25	0.129	0.051	0.097	**0.061**	0.081
scale	0.5	0.129	0.051	0.097	0.063	**0.080**
α	2	0.129	0.052	0.096	0.063	0.081
BondFF	1	0.167	0.100	0.118	0.090	0.108
layers	3	0.129	**0.050**	**0.096**	0.062	**0.080**
ℓ_b	4	0.129	**0.050**	**0.096**	0.062	0.081
Latent state	64	0.174	0.060	0.119	0.082	0.088
features $d_u + d_s$	256	**0.128**	0.051	0.099	**0.061**	0.085
Default BondFF	α=1, ℓ_b=2, $d_u + d_s$=128	0.129	0.051	0.097	0.063	0.081

We vary the BondFF energy scale (α), bond-state layer depth (ℓ_b), and bond-state width $d_u + d_s$ one at a time to assess their effect on HORM and Transition1x errors (Table 6).

Table 6 shows that changing α between 0.25 and 2 leaves most reported errors unchanged at the displayed precision; the largest variation is a 0.002 eV difference in Transition1x energy MAE. Reducing the bond-state depth to $\ell_b = 1$ or the width to $d_u + d_s = 64$ increases all five errors relative to the default. Increasing depth to three or four layers gives only small changes, while the width 256 slightly improves energy MAE but increases Transition1x force MAE. However, increasing feature depth and layer depth both increased the computational overhead. Therefore, the default setting provides a balanced result among the tested configurations, without being optimal for every metric.

E.5 Ablation study

We assess the three residual-energy terms and two bond-state structure variants against the default BondFF configuration (Table 7).

Table 7 shows higher MAE than the default for every reported metric after omitting any one of the three energy terms. The largest Chignolin increase occurs without E_{elec} (energy: 0.564 to 0.886 kcal/mol; force: 0.153 to 0.284 kcal/(mol·Å)), while omitting E_{charge} gives the largest HORM

Table 7: Ablation study. Energy, force, and Hessian mean absolute errors (MAE) are reported on HORM, and energy and force MAE on Transition1x and Chignolin, for BONDFF energy-term and bond-state structure ablations. HORM and Transition1x energy and force errors are in eV and eV/Å, and HORM Hessian errors are in eV/Å^2; Chignolin energy and force errors are in kcal/mol and kcal/(mol·Å). Each energy-term variant omits one term. Boldface marks the lowest value in each metric column; lower is better.

Dataset		HORM			Transition1x		Chignolin	
Metrics		Energy	Force	Hessian	Energy	Force	Energy	Force
Unit		eV	eV/Å	eV/Å^2	eV	eV/Å	kcal/mol	kcal/(mol·Å)
Energy term ablation	w.o. E_bond	0.134	0.097	0.124	0.085	0.082	0.585	0.201
	w.o. E_charge	0.138	0.076	0.172	0.099	0.085	0.578	0.191
	w.o. E_elec	0.148	0.061	0.149	0.089	0.091	0.886	0.284
Structure ablation	Orientation-agonistic[a]	0.283	0.084	0.104	0.154	0.194	0.725	0.283
	Full BondFF[b]	0.242	0.124	0.193	0.184	0.118	0.619	0.203
Default BONDFF		**0.129**	**0.051**	**0.097**	**0.063**	**0.081**	**0.564**	**0.153**

[a] Omits orientation initialization and all related orientation-aware design components.
[b] Uses BondFF for all layers without a geometric cutoff graph; atomic-node features are randomly initialized.

Hessian increase (0.097 to 0.172 eV/Å^2). Both reported structure variants also have higher errors than the default on all seven metrics. These comparisons support the contribution of the tested components within this configuration.

F BROADER IMPACT.

This work develops a foundational approach to machine-learning force fields rather than a system intended for immediate deployment. More accurate molecular energy and force predictions could accelerate simulation-driven research in drug discovery and materials science. The same capability could also be misused to aid the design of harmful chemical compounds. Applications in sensitive settings, therefore, warrant appropriate human oversight.

G LIMITATIONS.

The electrostatic correction has finite spatial support and does not replace a long-range solver for periodic systems. Although the model conserves total charge by construction, its learned atomic charges are not independently validated as physical charge estimates.