Beyond Scalar Electrostatics: Multipole Features for Long-Range Molecular Machine Learning

Anonymous Authors[1]

Abstract

Simulating long-range interactions remains a significant challenge for molecular machine learning potentials due to the need to accurately capture interactions over large spatial regions. In this work, we integrate the multipole expansion into equivariant ML potentials to model long-range interactions in QM/MM simulations more accurately. By incorporating the multipole expansion, we capture structured environmental effects beyond scalar electrostatic descriptors. Benchmark evaluations on solvated QM/MM systems demonstrate that including higher-order multipole features improves the accuracy of predicted energies and forces compared to scalar Coulomb descriptors and Ewald-based approaches, with low-order terms providing a favorable accuracy cost trade-off. Furthermore, we show that transfer learning from foundational models trained without explicit environmental information improves data efficiency in QM/MM settings. These results demonstrate the effectiveness of our approach for accurate and scalable simulations of complex molecular systems with long-range interactions.

1. Introduction

Modeling long-range interactions remains a central challenge for molecular machine learning. While graph neural networks (GNNs) and equivariant message-passing architectures have achieved strong performance on molecules and materials, most existing models rely primarily on local neighborhoods and therefore struggle to represent slowly decaying electrostatic effects (Sauceda et al., 2022; Deng et al., 2023; Gao et al., 2020; Ko et al., 2021; Batzner et al., 2022b; Schütt et al., 2021). This limitation is particularly important in settings where a small reactive region is strongly influenced by a large surrounding environment, such as solvated systems, biomolecules, and heterogeneous interfaces (Reichardt, 2007; Reichardt & Welton; Go & Taketomi, 1978; Sagui & Darden, 1999; Ambrosetti et al., 2014; Kang et al., 2024). In such systems, purely local representations can miss weak but cumulatively important long-range interactions, leading to systematic errors in energies and forces.

A key challenge in modeling long-range interactions is balancing expressiveness and scalability. On one end, several approaches augment local architectures with physics-inspired scalar descriptors, such as electrostatic potentials or related corrections, to expose the model to environmental information (Anstine & Isayev, 2023; Gao & Remsing, 2022; Gastegger et al., 2021; Song & Yang, 2025). These methods are computationally efficient, but a scalar potential provides only a compressed view of the external environment and cannot fully resolve anisotropic or higher-order electrostatic effects. On the other end, learned long-range coupling schemes, including message passing or attention over all atoms in the system, can represent richer interactions (Batatia et al., 2023; Frank et al., 2024; 2022; Maziarka et al., 2024). However, these approaches can become prohibitively expensive when the environment is large.

This trade-off is especially clear in electrostatically embedded QM/MM, where a small quantum-mechanical (QM) region interacts with a classical molecular-mechanics (MM) environment represented by fixed point charges (Tzeliou et al., 2022; Boereboom et al., 2018; Senn & Thiel, 2009; Chung et al., 2015). In this setting, scalar descriptors are often too restrictive, while full QM-MM learned coupling is unnecessarily expensive because MM atoms do not require the same high-dimensional learned representations as atoms in the quantum region. An effective ML architecture for QM/MM should therefore capture structured long-range information from the environment without introducing pairwise learned interactions over all QM/MM atom pairs.

In this work, we introduce Field-MACE, a long-range extension of the MACE architecture that incorporates the multipole expansion into an SO(3)-equivariant neural network. The central idea is to summarize the MM charge distribution through a hierarchy of multipole features and inject these features into the equivariant message-passing stack

---

[1] Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

of the QM region. Because multipole moments naturally decompose into angular channels, they are well aligned with spherical-harmonic-based architectures such as MACE and NequIP (Edmonds; Thorne, 1980; Batatia et al., 2022; Batzner et al., 2022a). This yields a representation that preserves directional and higher-order information beyond a scalar potential while avoiding full pairwise long-range message passing over the environment.

We evaluate Field-MACE on solvated QM/MM benchmarks and compare it against scalar Coulomb descriptors and Ewald-based long-range features. Across four benchmark systems, we find that low-order multipoles already provide substantial gains, with truncation at $\ell = 2$ or $\ell = 3$ yielding a strong accuracy-cost trade-off. We further show that initializing the short-range blocks of Field-MACE from a vacuum-trained foundation model substantially improves data efficiency in the embedded QM/MM setting. Our main contributions are:

- We introduce Field-MACE, a multipole-based long-range extension of an SO(3)-equivariant molecular architecture for QM/MM learning.
- We show that multipole features provide a favorable trade-off between expressiveness and scalability, outperforming scalar Coulomb descriptors and Ewald-based features on solvated benchmarks.
- We demonstrate that low-order multipoles capture most of the empirical gains, making the approach practical for large MM environments.
- We show that transfer learning from vacuum-trained foundation models improves data efficiency even when the downstream task includes explicit environmental effects.

2. Method

In this section, we first introduce background information on the multipole expansion and equivariant architectures constructed from a spherical-harmonic basis.

2.1. Message-passing

Atoms in a molecular system are characterized by their position vector $\mathbf{r}_i \in \mathbb{R}^3$, and atomic number, Z_i. These systems can be represented as a graph in which atoms correspond to nodes $\mathcal{V}_i$. Edges connect nodes within a cutoff neighborhood,

$$\mathcal{N}(i) = \{\mathcal{V}_j \mid \|\mathbf{r}_{ij}\|_2 = \|\mathbf{r}_i - \mathbf{r}_j\|_2 < r_{\text{cut}}\}\,, \tag{1}$$

where r_{cut} is a predefined cutoff radius. The categorical atomic numbers are mapped to learnable invariant embeddings $X_i \in \mathbb{R}^k$. During message passing (Gilmer et al., 2017; Jørgensen et al., 2018), each node i is associated with features $\mathbf{h}_i^{(t)} \in \mathbb{R}^k$ and each edge (i, j) is assigned the vector between the two atoms $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$. Features are initialized with the atomic embeddings $\mathbf{h}_i^{(0)} = X_i$ and updated iteratively:

$$\mathbf{m}_{ij}^{(t)} = \phi_{\text{msg}}\left(\mathbf{h}_i^{(t)}, \mathbf{h}_j^{(t)}, \mathbf{r}_{ij}\right) \tag{2}$$

$$\mathbf{m}_i^{(t)} = \sum_{j \in \mathcal{N}(i)} \mathbf{m}_{ij}^{(t)} \tag{3}$$

$$\mathbf{h}_i^{(t+1)} = \phi_{\text{update}}\left(\mathbf{h}_i^{(t)}, \mathbf{m}_i^{(t)}\right) \tag{4}$$

This three-step cycle (message construction, aggregation, update) is applied repeatedly. The messages $\mathbf{m}_{ij}^{(t)}$ are created for each edge using a learnable function ϕ_{msg}, the vector $\mathbf{r}_{ij}$ and the node features for atoms $\mathbf{h}_i$ and $\mathbf{h}_j$. Subsequently, they are aggregated over the neighborhood of each node. Finally, the node features $\mathbf{h}_i$ are updated using another learnable function ϕ_{update}, which combines the node features with the messages passed to it.

As illustrated in Fig. 1b, after one iteration $\mathbf{h}^{(t)}$ incorporates information from its immediate neighbors. After two iterations, it contains information from neighbors-of-neighbors, and so on. Depending on the architecture, $\mathbf{m}_{ij}^{(t)}$ may represent scalar and/or higher-order (tensorial) information.

2.2. Equivariance

Vectorial and tensorial molecular properties depend on the orientation of the system in space; *i.e.*, if the system is rotated globally, the associated properties must transform accordingly, which is known as equivariance (Esteves, 2020). Formally, the set of features $\mathbf{h}_i$ is equivariant under a group G if

$$D(R)\mathbf{h}_i(\mathbf{r}_1, ..., \mathbf{r}_n) = \mathbf{h}_i(R \circ (\mathbf{r}_1, ..., \mathbf{r}_n))\,, \tag{5}$$

where $R \in G$ and $D(R)$ is the equivalent operation acting on $\mathbf{h}_i$. The main focus of this paper will be on the special orthogonal group SO(3). In the case of SO(3), the equivariant features can be decomposed into irreducible representations (Thomas et al., 2018) that transform like spherical harmonics $Y_{\ell m}(\theta, \phi)$ with degree ℓ and order m ($m \in \{-\ell, ..., \ell\}$). Under a rotation $R \in SO(3)$, the spherical harmonics transform via Wigner-D matrices

$$Y_{\ell m'}(\theta', \phi') = \sum_m D^{\ell}_{m'm}(R) Y_{\ell m}(\theta, \phi)\,. \tag{6}$$

Using a spherical-harmonic basis, a generic two-body equivariant message construction for atom i and its neighbors can

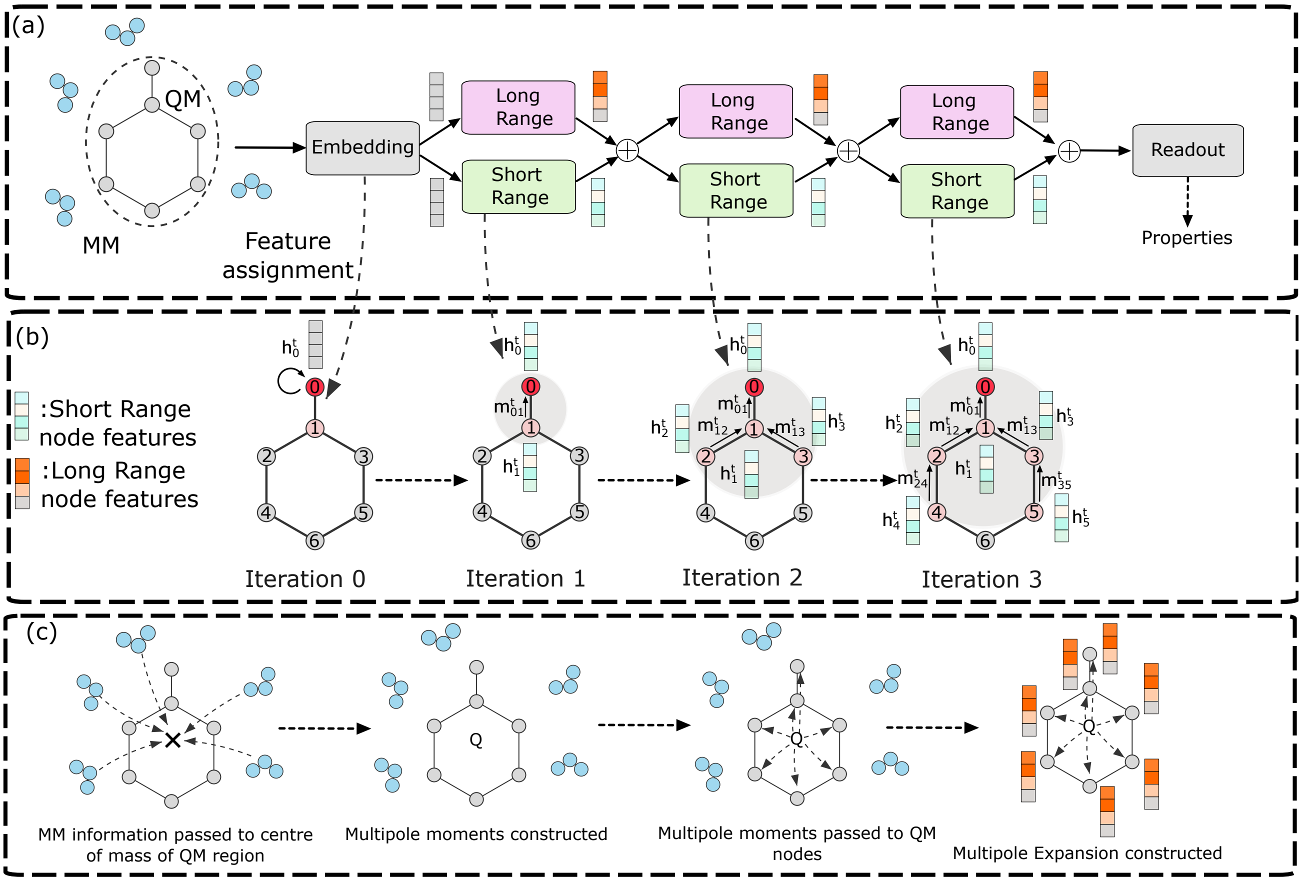

Figure 1. Overview of the model architecture and message-passing framework. (a) High-level overview of the model architecture. The system is divided into a quantum-mechanics (QM) region and a molecular-mechanics (MM) region. QM atoms receive initial node features through an embedding block, which are passed to the short-range and long-range interaction blocks. The MM region contributes only to the long-range block. Iterative message passing updates the QM features, which are then aggregated to predict the target properties. (b) Schematic illustration of the iterative message-passing procedure. Columns of blocks represent node features associated with each atom, and the shaded region indicates the receptive field from which atom 0 can obtain information at each iteration. (c) Construction of multipole features from the MM point-charge environment, which are used as long-range inputs in panel (a).

be written as

$$\mathbf{m}^{L,(t)}_{M,i,j} = R_{\ell_1 m_1}(r_{ij})\,Y_{\ell_1 m_1}(\theta,\phi)\,\mathbf{h}^{\ell_2,(t)}_{j,m_2} \tag{7}$$

$$\mathbf{m}^{\ell,(t)}_{m,i} = \sum_{L,M} C^{lm}_{L,M} \sum_{j\in\mathcal{N}(i)} \mathbf{m}^{L,(t)}_{M,i,j} \tag{8}$$

$$\mathbf{h}^{\ell,(t+1)}_{i,m} = w^{l}_{m}\,\mathbf{m}^{\ell,(t)}_{m,i} + \mathbf{h}^{\ell,(t)}_{i,m} \tag{9}$$

$$L = (\ell_1, \ell_2), \quad M = (m_1, m_2)\,,$$

where angles θ and ϕ correspond to the directionality of the vector $\mathbf{r}_{ij}$ between atoms i and j and r_{ij} is its length. To preserve the correct equivariance for each order of spherical harmonics, we must include the standard Clebsch Gordan coefficients $C^{lm}_{L,M}$ (Smorodinskiĭ & Shelepin, 1972). The values w^{l}_{m} correspond to learnable linear transformations across the channels of the representation corresponding to the degree and order of the spherical harmonics. An additional non-linear function can be placed alongside the linear transformation. The equations form the basis of the short-range interaction block in the following section. MACE extends this construction to many-body interactions, which is discussed in the Supporting Information (SI) in section S1.

2.3. Multipole Expansion

The multipole expansion decomposes complex electromagnetic fields into hierarchical terms. This approach is particularly useful for analyzing fields far from their sources, where higher-order contributions become negligible. Constructed in terms of spherical harmonic functions, the multipole expansion integrates naturally into equivariant neural networks that employ a spherical harmonic basis. In Field-MACE, its inclusion is motivated by the limited expressiveness of purely scalar electrostatic descriptors: Evaluating only a scalar electrostatic potential at the QM atoms does not provide a complete description of the MM environment, as different configurations of MM charges can produce identical scalar potentials at a given QM atom. In contrast, the multipole expansion resolves this problem by offering a more complete description of the surrounding environment.

A scalar potential $\Phi(\mathbf{r})$ generated by a charge distribution $\rho(\mathbf{r}')$, where $\mathbf{r}$ is the field point and $\mathbf{r}'$ is the source point, is defined by Coulomb's law as

$$\Phi(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0} \int \frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\, \mathrm{d}\mathbf{r}' \,. \tag{10}$$

The function $\frac{1}{|\mathbf{r}-\mathbf{r}'|}$ can be expressed as a sum of spherical harmonics assuming $r << r'$:

$$\frac{1}{|\mathbf{r}-\mathbf{r}'|} = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \frac{4\pi}{2\ell+1} \frac{r^{\ell}}{r'^{\ell+1}} Y_{\ell m}(\theta,\phi) Y_{\ell m}{}^{*}(\theta',\phi') \tag{11}$$

The condition $r << r'$ reflects the standard convergence requirement of the multipole expansion. In Field-MACE, the truncated series is used to construct equivariant feature tensors rather than to compute exact electrostatic potentials, so strict convergence is not required.

Substituting this expansion into Coulomb's law yields

$$\Phi(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \frac{r^{\ell}}{2\ell+1} Y_{\ell m}(\theta,\phi) Q_{\ell m}, \tag{12}$$

where the multipole moment $Q_{\ell m}$ is defined as

$$Q_{\ell m} = \int \rho(\mathbf{r}') \frac{Y^{*}_{\ell m}(\theta',\phi')}{r'^{\ell+1}}\, \mathrm{d}\mathbf{r}' \,. \tag{13}$$

Each term in this expansion corresponds to a specific multipole order, with higher-order terms decaying more rapidly with distance. Together, these terms provide a structured representation of how the MM charges influence the QM region. The monopole term ($\ell = 0$) represents the total charge contribution, the dipole term ($\ell = 1$) encodes the orientation and strength of the resulting electric field, and the quadrupole term ($\ell = 2$) captures the curvature in the field. Higher-order terms ($\ell > 2$) describe progressively finer topological variations of the electrostatic field generated by the charge distribution. These multipole moments centered at each QM atom can be used as feature vectors, as shown in the next section.

2.4. Multipole Moments

The QM/MM approach splits a chemical system into a QM and an MM region as shown in Fig. 1a, where the MM atoms are described by their positions $\mathbf{r}_j^{\mathrm{MM}}$ and partial charges q_j^{MM} as assigned by a classical force field. The corresponding charge distribution surrounding the QM region is

$$\rho(\mathbf{r}') = \begin{cases} q_j^{\mathrm{MM}}, & \text{if } \mathbf{r}' = \mathbf{r}_j^{\mathrm{MM}}, \\ 0, & \text{otherwise} \end{cases} \,. \tag{14}$$

This simplifies the multipole moments, eq. (13), into a sum over discrete charges:

$$Q_{\ell m} = \sum_j Q^j_{\ell m} = \sum_j q_j^{\mathrm{MM}} \frac{Y^{*}_{\ell m}(\theta_j^{\mathrm{MM}}, \phi_j^{\mathrm{MM}})}{(r_j^{\mathrm{MM}})^{l+1}} \,. \tag{15}$$

The values $\theta_j^{\mathrm{MM}}, \phi_j^{\mathrm{MM}}$ correspond to the angles of the MM charges in spherical coordinates. To maintain translational invariance, we must first shift all coordinates such that the center of mass of the QM region lies at the origin before computing the multipole moments. The electrostatic potential consists of a single scalar value, but the terms in the expansion are used to construct higher order tensors, which we use as features when building the long-range messages. These are defined as follows,

$$\Phi^i_{\ell m}(\mathbf{r}_i^{\mathrm{QM}}) = \frac{1}{\varepsilon_0} \sum_j \frac{Q^j_{\ell m}}{2\ell+1} \frac{\mathbf{h}^{i,\mathrm{QM}}_{\ell,m}}{(r_i^{\mathrm{QM}})^{\ell+1}}. \tag{16}$$

Additionally, it is worth noting that the multipole moments themselves can be used as feature vectors for each atom to avoid centering the QM region, but they need to be constructed for every QM atom in the region. These features would be given by,

$$Q^i_{\ell m}(\mathbf{r}_i^{\mathrm{QM}}) = \frac{1}{\varepsilon_0} \sum_j q_j^{\mathrm{MM}} \frac{Y^{*}_{\ell m}(\theta_j^{\mathrm{MM}}, \phi_j^{\mathrm{MM}})}{(r_j^{\mathrm{MM}} - r_i^{\mathrm{QM}})^{l+1}}. \tag{17}$$

In practice, Eq. (16) corresponds to a global multipole expansion of the MM charge distribution around a single reference point at the center of the QM region. This representation is computationally efficient and well suited for medium to long range electrostatic interactions, but it reduces the spatial resolution for very short range interactions at the QM/MM boundary, because the MM information is first aggregated at the QM center. To improve the description of such local interactions, we also consider the local variant in Eq. (17), where multipole moments are constructed with respect to each QM atom. In this per-atom expansion, an MM atom approaching a particular QM atom gives a dominant contribution to the moments associated with that atom, so its short range effect is not averaged out at the QM center of mass. Additionally, Eq. (17) is advantageous for larger QM regions or cases where the center of mass is not well defined, such as systems with periodic boundary conditions. Furthermore, since the number of terms in the multipole expansion is infinite, a user-defined cutoff must be introduced in order to limit the number of higher-order terms. For very large MM regions, the spatial resolution of a single global multipole representation may decrease, as more components are required to distinguish between large charge distributions. This limitation can be mitigated by defining multiple multipole expansions over a series of shell like regions around the QM subsystem (e.g., 0–5 Å, 5–10 Å), which might provide finer spatial resolution of the electrostatic environment. In the next Section, we will incorporate the multipole moments and terms from the multipole expansion terms in the message construction. A high-level illustration of the process that was described can be seen in Fig. 1c).

2.5. Message Construction

In Field-MACE, multipole-derived long-range information augments the short-range equivariant messages (Fig. 1b). Concretely, after the short-range message construction in Eq. (8), we add a long-range contribution, as can be seen in Fig. 1a) by incorporating the multipole expansion terms,

$$\mathbf{m}_{m,i}^{\ell,(t+1)} = \mathbf{m}_{m,i}^{\ell,(t)} + w_m^{\ell,(t)} \Phi_{\ell m}^{i,(t)}(\mathbf{r}_i^{\mathrm{QM}}). \tag{18}$$

Equivalently, with the multipole moments:

$$\mathbf{m}_{m,i}^{\ell,(t+1)} = \mathbf{m}_{m,i}^{\ell,(t)} + w_m^{\ell,(t)} Q_{\ell m}^{i,(t)}(\mathbf{r}_i^{\mathrm{QM}}). \tag{19}$$

The node features are then updated in the same way as in Eq. (9). Injecting the multipole information at the message level ensures that MM electrostatics can influence the learned many-body interactions among QM atoms, rather than entering only as a correction to each QM atom independently. While one could alternatively couple multipole tensors to equivariant features via explicit tensor products, this would scale less favorably with both the number of multipole components and feature dimensionality; we therefore use a direct sum in Field-MACE to preserve low-scaling computational cost.

3. Results

3.1. Effect of Higher Order Moments

We extend the established MACE framework to include multipole features while keeping the remaining hyperparameters and the total number of irreducible representations fixed for consistency across models. Details on the MACE-relevant hyperparameters are provided in the Appendix.

We quantify the effect of truncating the multipole expansion at $\ell_{\max} \in \{0, \ldots, 5\}$ on four solvated datasets: benzene, uracil, retinoic acid, and the transition state of the reaction of S-adenosylmethionine with cytosine (SAM). We report mean absolute errors (MAEs) for energies and forces and use five-fold cross validation throughout. The systems are visualized in Fig. S1, and further details are provided in the Appendix. Results are shown in Tables 1 and 2.

Across all systems, increasing $\ell_{\max}$ yields a consistent reduction in error up to $\ell_{\max} \approx 3$. Beyond that, improvements stagnate: higher-order multipole components rapidly decrease in magnitude, with $\ell = 5$ having values close to 10^{-7}, leading to minor improvements in predictive accuracy. A more detailed analysis of the components can be seen in the Appendix. For the energy MAEs, the most pronounced improvement is observed when going from $\ell_{\max} = 0$ to $\ell_{\max} = 2$; higher orders provide only minor additional gains, except for SAM, where $\ell_{\max} = 3$ performs best. This is consistent with the SAM system having the largest and most complex MM environment among the considered systems, thus prompting the need for higher order components. QM/MM sizes are summarized in Table 3. Force MAEs exhibit the same trend, with larger and more flexible systems (retinoic acid and SAM) showing the strongest improvements with increasing ℓ.

In addition to measuring the test error depending on the truncation level of the multipole expansion, we also timed the different models. Table 3 reports the average time per batch for each system (batch size fixed across $\ell_{\max}$). Comparing benzene and uracil, which have the same QM size but substantially different MM sizes, shows that N^{MM} has only a minor impact on training time in our implementation. This is consistent with the multipole-moment accumulation in Eq. (15) being a lightweight sum performed at the beginning of the prediction of the MACE model, requiring little computational effort. Increasing $\ell_{\max}$ also leads only to a modest increase in cost. The dominant cost is associated with the equivariant message passing over QM atoms, consistent with an overall scaling dominated by the QM representation and message passing (empirically close to $\mathcal{O}(N^{\mathrm{QM}})$ per batch for fixed local neighborhood sizes). Additional timing comparisons between the foundational MACE-MP0 model and Field-MACE are provided in the Appendix.

3.2. Benchmarks to existing approaches

We benchmark the multipole-based long-range representation against two commonly used alternatives: Ewald summation (Kosmala et al., 2023; Toukmaji & Board Jr, 1996) and scalar Coulomb-potential descriptors (Gastegger et al., 2021). For the Ewald summation method, the solvated QM/MM systems are treated with three-dimensional periodic boundary conditions, such that the classical Ewald construction applies. For each system, we compare energy and force MAEs. Both Ewald and Coulomb descriptors were implemented within the MACE architecture as long-range inputs, and we use identical MACE hyperparameters across all models to ensure a fair comparison. Results are shown in Fig. 2.

Across all systems, Ewald summation outperforms $\ell_{\max} = 0$ and achieves an accuracy comparable to $\ell_{\max} = 1$, while Field-MACE with $\ell_{\max} = 2$ consistently outperforms Ewald. The scalar Coulomb potential performs comparably to $\ell_{\max} = 0$ in larger and more complex systems, but is substantially worse than higher-$\ell_{\max}$ Field-MACE models,

Table 1. Mean Absolute Errors (MAEs) for energies (meV) of Field-MACE models evaluated on different molecules in solution. Results are shown for varying multipole expansion truncation levels ($\ell = 0$ to $\ell = 5$).

System	$\ell = 0$	$\ell = 1$	$\ell = 2$	$\ell = 3$	$\ell = 4$	$\ell = 5$
Benzene	20.8	9.2	6.0	5.0	4.9	4.9
Uracil	11.7	7.4	5.4	4.8	4.8	4.7
Retinoic Acid	62.2	54.5	36.9	35.3	35.0	34.2
SAM	75.6	71.5	66.7	50.3	51.7	50.8

Table 2. Mean Absolute Errors (MAEs) for forces (meV/Å) of Field-MACE models evaluated on different molecules in solution. Results are shown for varying multipole expansion truncation levels ($\ell = 0$ to $\ell = 5$).

System	$\ell = 0$	$\ell = 1$	$\ell = 2$	$\ell = 3$	$\ell = 4$	$\ell = 5$
Benzene	32.3	23.6	17.4	15.4	15.1	15.2
Uracil	39.2	28.7	23.6	21.7	21.7	21.5
Retinoic Acid	78.1	70.8	50.3	42.8	44.2	42.5
SAM	77.6	76.3	73.5	64.6	65.5	64.6

indicating that a purely scalar descriptor can be insufficient to resolve complex, anisotropic solvent environments. However, for smaller and relatively rigid solutes such as benzene and, to a lesser extent, uracil, the Coulomb descriptor can match or exceed the force accuracy of higher-$\ell_{\max}$ models. This can be explained by the relatively small size and restricted conformational flexibility of these molecules. Even though uracil is polar, the resulting solvent induced electrostatic environment is comparatively less complex since there is little configurational flexibility. In this situation, the scalar Coulomb potential at each QM atom already captures most of the relevant environmental information, such that including higher order multipoles provides little to no additional benefit. In contrast, the SAM system and retinoic acid are larger and more flexible, and their conformational ensembles are more strongly modulated by the solvent interactions, making higher-order angular information more beneficial. More generally, molecules with greater conformational flexibility and stronger, non-uniform solvent interactions are described insufficiently by a purely scalar Coulomb potential and, therefore, benefit more from the inclusion of higher-order multipoles. In practice, this suggests using $\ell = 0$ for small, rigid solutes such as benzene and uracil, and including multipoles up to $\ell = 2$ (or 3 for particularly complex environments) provides a robust accuracy-cost trade-off for flexible, strongly solvated systems.

3.3. Transferability of Ground State Foundational Models

In practice, Field-MACE can be trained on data generated with any of the standard strategies used for machine learning potentials, such as short QM or QM/MM molecular dynamics runs, Wigner sampling, enhanced sampling schemes, or uncertainty based approaches. For QM/MM applications, however, a key bottleneck is that such QM/MM reference data are typically much scarcer than ground state QM data without environmental effects. In this section, we therefore analyze how this burden can be reduced by initializing Field-MACE with weights of pre-trained foundational ground state models that have not been trained on data including explicit environment effects. Concretely, we leverage the foundational MACE-OFF model's pre-trained representations as initial parameters for the short-range modules of Field-MACE and then fine tune them on a comparatively small set of embedded QM/MM configurations. As shown below, this transfer-learning strategy substantially reduces the amount of system-specific QM/MM data required to reach a given accuracy. Full details on models and training can be found in the Appendix; additional transfer tests across related systems and solvents are reported in the Appendix.

Fig. 3 shows learning curves for three solvated molecular systems trained with varying amounts of fine-tuning data. The short-range blocks are initialized with the weights from the foundational model and the long-range blocks with random ones, which reduces the amount of training data required to achieve comparable accuracy to models trained from scratch, particularly for force predictions (panels d-f in Fig. 3). Across all systems, the pre-trained initialization demonstrates a clear advantage, leading to improved accuracy compared to random initialization, even though the foundational model was trained without any solvent or environmental terms.

The advantage of using the foundational representation is most pronounced in systems with few strongly electronegative atoms, as is the case in benzene, see Fig. 3a) and d), or retinoic acid, see Fig. 3 c) and f). These molecules exhibit comparatively weak and unspecific solute-solvent interac-

Table 3. Training times per batch in seconds for Field-MACE models ($\ell = 0$) to ($\ell = 5$) on various molecular systems in solution.

System	**MM Size**	**QM Size**	$\ell = 0$	$\ell = 1$	$L = 2$	$\ell = 3$	$\ell = 4$	$\ell = 5$
Benzene	250	12	0.069	0.076	0.084	0.092	0.101	0.110
Uracil	1600	12	0.068	0.075	0.082	0.088	0.098	0.108
Retinoic Acid	2600	50	0.335	0.344	0.350	0.357	0.369	0.380
SAM	3500	63	0.444	0.453	0.461	0.469	0.482	0.495

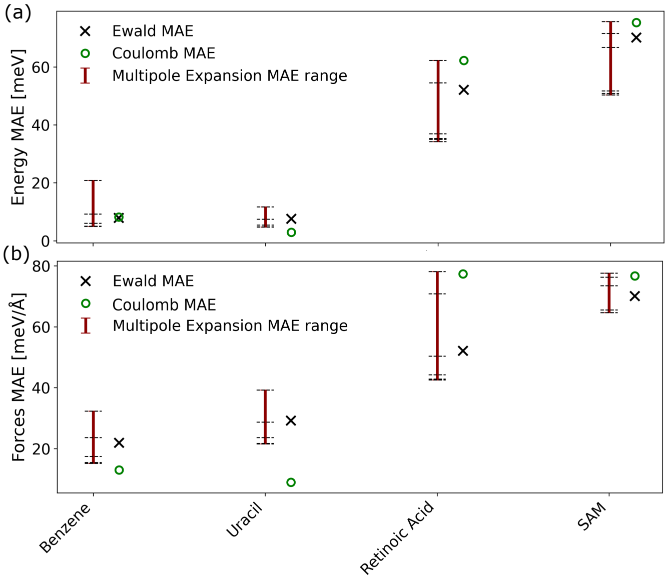

Figure 2. Comparison of energy and force prediction accuracy for solvated molecular systems. (a) Energy and (b) force MAEs for benzene, uracil, retinoic acid, and SAM in solution comparing Ewald summation, Coulomb potential, and Field-MACE with different multipole truncation levels. The dashes on the error bars represent the error for each of the cutoff values in the multipole expansion going from 0 with the top dash and 5 with the bottom dash.

tions in water. Hence, much of the short-range chemistry learned by the vacuum-trained foundational model remains relevant, and fine-tuning primarily needs to adjust the response to the environment through the added long-range terms.

In contrast, for polar systems like uracil (Fig. 3 b) and e)), which feature strong hydrogen bonding and complex electrostatic interactions, the benefits of the pre-trained initialization are less pronounced. While transfer learning still outperforms the randomly initialized model, the benefit is smaller compared to the other examples. This suggests that the foundational model's learned parameters need a larger adjustment in systems that interact more strongly with the MM environment. In such cases, additional task-specific pre-training or a larger fine-tuning set with embedded structure may be required to fully realize the potential of transfer learning.

4. Discussion

In this work, we introduced Field-MACE, a multipole-based extension of an SO(3)-equivariant molecular architecture for modeling long-range interactions in QM/MM systems. By replacing purely scalar environmental descriptors with a hierarchy of multipole features, the model captures structured electrostatic information from the MM environment while avoiding the cost of full long-range pairwise coupling. Across solvated QM/MM benchmarks, we found that multipole features consistently improve energy and force predictions relative to scalar Coulomb descriptors and Ewald-based long-range features. In particular, low-order multipoles already capture most of the empirical gains, with $\ell = 2$ or $\ell = 3$ providing a strong accuracy cost trade off. These results show that incorporating directional and higher-order environmental information can substantially improve the expressiveness of equivariant molecular ML models without sacrificing scalability.

We also demonstrated that transfer learning from vacuum trained foundational models improves data efficiency in the embedded QM/MM setting, particularly for force prediction.

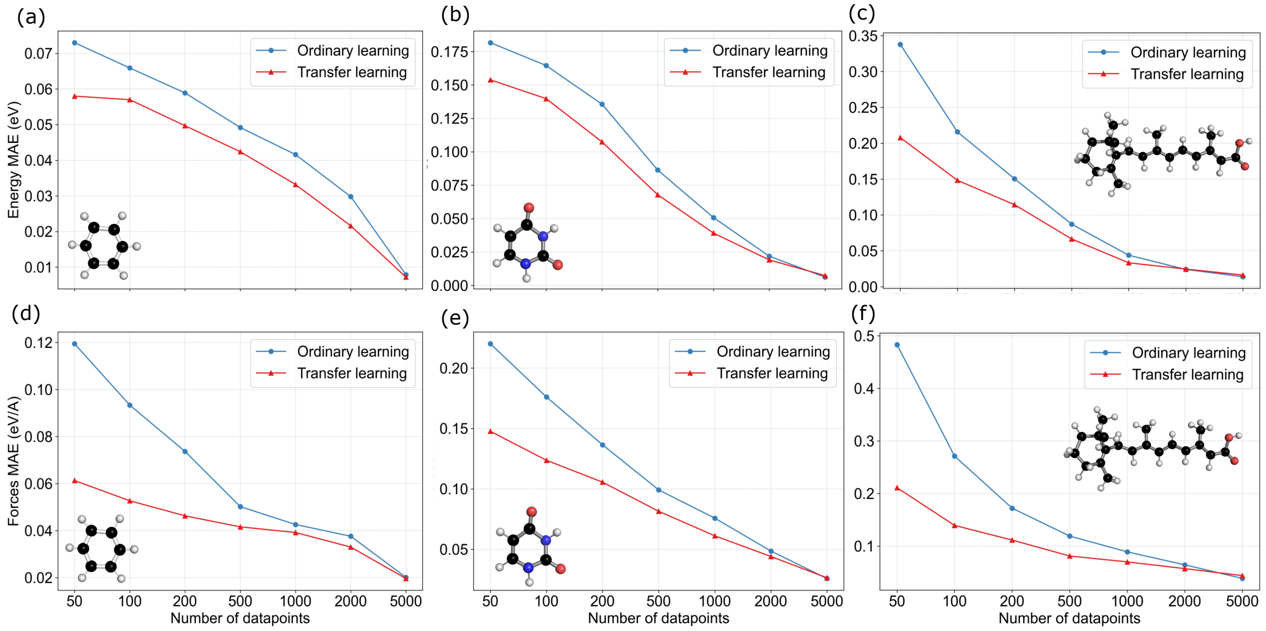

Figure 3. Comparison of transfer learning from MACE-OFF foundational model (red curves) against randomly initialized Field-MACE models (blue curves). Learning curves for energies (top row) and forces (bottom row) of molecules solvated in water: (a) and (d) benzene, (b) and (e) uracil, and (c) and (f) retinoic acid.

Even though such pre-trained models do not include explicit environmental effects, their learned short-range representations remain highly useful when combined with multipole based long-range features. This suggests that foundation-model pretraining can be an effective strategy for reducing the cost of building accurate QM/MM models in low-data regimes. At the same time, our results highlight important limitations. While the multipole expansion is effective for dominant long-range electrostatic effects, truncated expansions are less well suited to very short-range interactions near the QM/MM boundary, where the response of the QM region depends sensitively on the precise positions of nearby MM atoms. In addition, highly polar or structurally complex environments may require either higher-order terms or more localized treatments of the environment. Overall, Field-MACE provides a practical route toward accurate, scalable, and data-efficient ML/MM modeling. Our results suggest that structured long-range representations based on multipole features are a promising direction for extending molecular machine learning to more realistic embedded chemical systems.

Software and Data

The relevant code and datasets can be found here: https://figshare.com/articles/dataset/Models_data_and_code_for_publication_Incorporating_Long-Range_Interactions_via_the_Multipole_Expansion_into_Ground_and_Excited-State_Molecular_Simulations_/28497857 and https://github.com/rhyan10/FieldMACE/tree/master , any additional datasets used can be accessed on request.

Impact Statement

This work advances machine learning methods for molecular simulations, with potential applications in chemistry, materials science, and drug discovery. By improving the modeling of long-range interactions in QM/MM systems, it may enable more accurate and efficient simulations, reducing reliance on expensive computational or experimental methods.

The work does not involve human data and presents minimal direct ethical risks. However, improved molecular modeling tools could contribute to downstream applications with dual-use potential, such as the design of harmful chemical compounds. Additionally, the computational cost of training machine learning models has environmental implications.

Overall, we believe the primary impact of this work is to support scientific progress, while highlighting the importance of responsible use and consideration of computational resources.

References

Ambrosetti, A., Ferri, N., DiStasio, R. A., and Tkatchenko, A. Long-range correlation energy calculated from coupled atomic response functions. *J. Chem. Phys.*, 140(18): 18A508, 2014. doi: 10.1063/1.4874755.

Anstine, D. M. and Isayev, O. Machine learning interatomic potentials and long-range physics. *J. Phys. Chem. A*, 127 (11):2417–2431, 2023. doi: 10.1021/acs.jpca.2c08541.

Batatia, I., Kovacs, D. P., Simm, G., Ortner, C., and Csányi, G. Mace: Higher order equivariant message passing neural networks for fast and accurate force fields. *Adv. Neural Inf. Process. Syst.*, 35:11423–11436, 2022.

Batatia, I., Schaaf, L. L., Chen, H., Csányi, G., Ortner, C., and Faber, F. A. Equivariant matrix function neural networks. *arXiv preprint arXiv:2310.10434*, 2023.

Batzner, S., Musaelian, A., Sun, L., Geiger, M., Mailoa, J. P., Kornbluth, M., Molinari, N., Smidt, T. E., and Kozinsky, B. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature Communications*, 13(1), May 2022a. ISSN 2041-1723. doi: 10.1038/s41467-022-29939-5. URL `http://dx.doi.org/10.1038/s41467-022-29939-5`.

Batzner, S., Musaelian, A., Sun, L., Geiger, M., Mailoa, J. P., Kornbluth, M., Molinari, N., Smidt, T. E., and Kozinsky, B. E (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat. Commun.*, 13(1):2453, 2022b.

Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A*, 38(6):3098, 1988.

Becke, A. D. A new mixing of hartree–fock and local density-functional theories. *J. Chem. Phys.*, 98(2):1372–1377, 1993.

Bensberg, M., Eckhoff, M., Husistein, R. T., Teynor, M. S., Sora, V., Bro-Jørgensen, W., Thomasen, F. E., Krogh, A., Lindorff-Larsen, K., Solomon, G. C., et al. Hierarchical quantum embedding by machine learning for large molecular assemblies. *arXiv preprint arXiv:2503.03928*, 2025.

Berendsen, H. J., Grigera, J.-R., and Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem.*, 91(24):6269–6271, 1987.

Boereboom, J. M., Fleurat-Lessard, P., and Bulo, R. E. Explicit solvation matters: Performance of qm/mm solvation models in nucleophilic addition. *J. Chem. Theory Comput.*, 14(4):1841–1852, 2018.

Boöselt, L., Thürlemann, M., and Riniker, S. Machine learning in qm/mm molecular dynamics simulations of condensed-phase systems. *J. Chem. Theory Comput.*, 17 (5):2641–2658, 2021.

Buterez, D., Janet, J. P., Kiddle, S. J., Oglic, D., and Lió, P. Transfer learning with graph neural networks for improved molecular property prediction in the multi-fidelity setting. *Nat. Commun.*, 15(1):1517, 2024.

Cerullo, G., Polli, D., Lanzani, G., De Silvestri, S., Hashimoto, H., and Cogdell, R. J. Photosynthetic light harvesting by carotenoids: detection of an intermediate excited state. *Science*, 298(5602):2395–2398, 2002.

Chung, L. W., Sameera, W., Ramozzi, R., Page, A. J., Hatanaka, M., Petrova, G. P., Harris, T. V., Li, X., Ke, Z., Liu, F., et al. The oniom method and its applications. *Chemical reviews*, 115(12):5678–5796, 2015.

Deng, B., Zhong, P., Jun, K., Riebesell, J., Han, K., Bartel, C. J., and Ceder, G. Chgnet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nat. Mach. Intell.*, 5(9):1031–1041, 2023.

Diccianni, J. B. and Diao, T. Mechanisms of nickel-catalyzed cross-coupling reactions. *Trends in Chemistry*, 1(9):830–844, 2019.

Dunlap, B., Connolly, J., and Sabin, J. On first-row diatomic molecules and local density models. *J. Chem. Phys.*, 71 (12):4993–4999, 1979.

Edmonds, A. R. *Angular momentum in quantum mechanics*, volume 4. Princeton university press.

Esteves, C. Theoretical aspects of group equivariant neural networks. *arXiv preprint arXiv:2004.05154*, 2020.

Frank, J. T. et al. Euclidean fast attention: Machine learning global atomic representations at linear cost. *arXiv preprint*, 2024.

Frank, T., Unke, O., and Müller, K.-R. So3krates: Equivariant attention for interactions on arbitrary length-scales in molecular systems. *Adv. Neural Inf. Process. Syst.*, 35: 29400–29413, 2022.

Gao, A. and Remsing, R. C. Self-consistent determination of long-range electrostatics in neural network potentials. *Nat. Commun.*, 13(1):1572, 2022. doi: 10.1038/s41467-022-29234-x.

Gao, X., Ramezanghorbani, F., Isayev, O., Smith, J. S., and Roitberg, A. E. Torchani: a free and open source pytorch-based deep learning implementation of the ani neural network potentials. *J. Chem. Inf. Model.*, 60(7): 3408–3415, 2020.

Gastegger, M., Schütt, K. T., and Müller, K.-R. Machine learning of solvent effects on molecular spectra and reactions. *Chem. Sci.*, 12(34):11473–11483, 2021.

Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., and Dahl, G. E. Neural message passing for quantum chemistry. In *Int. Conf. Mach. Learn.*, pp. 1263–1272. PMLR, 2017.

Go, N. and Taketomi, H. Respective roles of short- and long-range interactions in protein folding. *Proc. Natl. Acad. Sci.*, 75(2):559–563, 1978. doi: 10.1073/pnas.75.2.559.

Granucci, G. and Persico, M. Critical appraisal of the fewest switches algorithm for surface hopping. *J. Chem. Phys.*, 126(13):134114, 2007. doi: 10.1063/1.2715585.

Grimme, S. Semiempirical hybrid density functional with perturbative second-order correlation. *J. Chem. Phys.*, 124(3):034108, 2006. doi: 10.1063/1.2148954.

Grimme, S., Ehrlich, S., and Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.*, 32(7):1456–1465, 2011.

Han, F.-S. Transition-metal-catalyzed suzuki–miyaura cross-coupling reactions: a remarkable advance from palladium to nickel catalysts. *Chemical Society Reviews*, 42(12): 5270–5298, 2013.

Heydari, S., Raniolo, S., Livi, L., and Limongelli, V. Transferring chemical and energetic knowledge between molecular systems with machine learning. *Commun. Chem.*, 6 (1):13, 2023.

Himanen, L., Jäger, M. O. J., Morooka, E. V., Federici Canova, F., Ranawat, Y. S., Gao, D. Z., Rinke, P., and Foster, A. S. DScribe: Library of descriptors for machine learning in materials science. *Computer Physics Communications*, 247:106949, 2020. ISSN 0010-4655. doi: 10.1016/j.cpc.2019.106949. URL `https://doi.org/10.1016/j.cpc.2019.106949`.

Jørgensen, P. B., Jacobsen, K. W., and Schmidt, M. N. Neural message passing with edge updates for predicting properties of molecules and materials. *arXiv preprint arXiv:1806.03146*, 2018.

Kang, K. et al. Accelerating the training and improving the reliability of machine-learned interatomic potentials for strongly anharmonic materials through active learning. *arXiv preprint*, 2024.

Ko, T. W., Finkler, J. A., Goedecker, S., and Behler, J. A fourth-generation high-dimensional neural network potential with accurate electrostatics including non-local charge transfer. *Nat. Commun.*, 12(1):398, 2021.

Kosmala, A., Gasteiger, J., Gao, N., and Günnemann, S. Ewald-based long-range message passing for molecular graphs. In *Proc. Int. Conf. Mach. Learn.*, pp. 17544–17563. PMLR, 2023.

Mai, S., Marquetand, P., and González, L. A general method to describe intersystem crossing dynamics in trajectory surface hopping. *Int. J. Quantum Chem.*, 115(18):1215–1231, 2015.

Mai, S., Marquetand, P., and González, L. Nonadiabatic dynamics: The sharc approach. *Wiley Int. Rev.: Comput. Mol. Sci.*, 8(6):e1370, 2018.

Mai, S., Avagliano, D., Heindl, M., Marquetand, P., Menger, M. F. S. J., Oppel, M., Plasser, F., Polonius, S., Ruckenbauer, M., Shu, Y., Truhlar, D. G., Zhang, L., Zobel, P., and González, L. SHARC3.0: Surface hopping including arbitrary couplings — program package for non-adiabatic dynamics. https://sharc-md.org/, 2023.

Malde, A. K. et al. An automated force field topology builder (atb) and repository: version 1.0. *J. Chem. Theory Comput.*, 7(12):4026–4037, 2011.

Mathew, S., Yella, A., Gao, P., Humphry-Baker, R., Curchod, B. F. E., Ashari-Astani, N., Tavernelli, I., Rothlisberger, U., Nazeeruddin, M. K., and Grätzel, M. Dye-sensitized solar cells with 13% efficiency achieved through the molecular engineering of porphyrin sensitizers. *Nat. Chem.*, 6:242–247, 2014. doi: 10.1038/nchem.1861.

Maziarka, Ł., Majchrowski, D., Danel, T., Gaiński, P., Tabor, J., Podolak, I., Morkisz, P., and Jastrzebski, S. Relative molecule self-attention transformer. *J. Cheminf.*, 16(1):3, 2024.

Mazouin, B., Schöpfer, A. A., and von Lilienfeld, O. A. Selected machine learning of homo–lumo gaps with improved data-efficiency. *Mater. Adv.*, 3(22):8306–8316, 2022.

Pultar, F., Thürlemann, M., Gordiy, I., Doloszeski, E., and Riniker, S. Neural network potential with multiresolution approach enables accurate prediction of reaction free energies in solution. *Journal of the American Chemical Society*, 147(8):6835–6856, 2025.

Rappé, A. K., Casewit, C. J., Colwell, K., Goddard III, W. A., and Skiff, W. M. Uff, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *Journal of the American chemical society*, 114(25):10024–10035, 1992.

Reichardt, C. Solvents and solvent effects: An introduction. *Org. Process Res. Dev.*, 11(1):105–113, 2007. doi: 10.1021/op600170z.

Reichardt, C. and Welton, T. *Solvents and Solvent Effects in Organic Chemistry*. John Wiley & Sons, 4 edition. doi: 10.1002/9783527632220.

Richter, M., Marquetand, P., González-Vázquez, J., Sola, I., and González, L. Sharc: ab initio molecular dynamics with surface hopping in the adiabatic representation including arbitrary couplings. *J. Chem. Theory Comput.*, 7(5):1253–1258, 2011. doi: 10.1021/ct1007394. URL `https://doi.org/10.1021/ct1007394`. PMID: 26610121.

Riniker, S., Pultar, F., Thürlemann, M., and Gordiy, I. Qm reference energies, gradients, molecular multipoles with electrostatic embedding: Dataset of qm reference properties for alanine dipeptide, nickel phosphine complexes, and pyridine-quinoline dimers. 2024.

Romero, E., Novoderezhkin, V. I., and van Grondelle, R. Quantum design of photosynthesis for bio-inspired solar-energy conversion. *Nature*, 543(7645):355–365, 2017.

Sagui, C. and Darden, T. A. Molecular dynamics simulations of biomolecules: Long-range electrostatic effects. *Annu. Rev. Biophys. Biomol. Struct.*, 28(1):155–179, 1999. doi: 10.1146/annurev.biophys.28.1.155.

Sauceda, H. E., Gálvez-González, L. E., Chmiela, S., Paz-Borbón, L. O., Müller, K.-R., and Tkatchenko, A. Bigdml—towards accurate quantum machine learning force fields for materials. *Nat. Commun.*, 13(1):3733, 2022.

Schütt, K., Unke, O., and Gastegger, M. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *Proc. Int. Conf. Mach. Learn.*, pp. 9377–9388. PMLR, 2021.

Senn, H. M. and Thiel, W. Qm/mm methods for biomolecular systems. *Angewandte Chemie International Edition*, 48(7):1198–1229, 2009.

Shu, Y., Zhang, L., Chen, X., Sun, S., Huang, Y., and Truhlar, D. G. Nonadiabatic dynamics algorithms with only potential energies and gradients: Curvature-driven coherent switching with decay of mixing and curvature-driven trajectory surface hopping. *J. Chem. Theory Comput.*, 18(3):1320–1328, February 2022a. ISSN 1549-9626. doi: 10.1021/acs.jctc.1c01080. URL `http://dx.doi.org/10.1021/acs.jctc.1c01080`.

Shu, Y., Zhang, L., Chen, X., Sun, S., Huang, Y., and Truhlar, D. G. Nonadiabatic dynamics algorithms with only potential energies and gradients: Curvature-driven coherent switching with decay of mixing and curvature-driven trajectory surface hopping. *J. Chem. Theory Comput.*, 18 (3):1320–1328, 2022b.

Smorodinskiĭ, Y. A. and Shelepin, L. A. Clebsch-gordan coefficients, viewed from different sides. *Sov. Phys. Usp.*, 15(1):1, 1972.

Song, G. and Yang, W. Nepoip/mm: Toward accurate biomolecular simulation with a machine learning/molecular mechanics model incorporating polarization effects. *J. Chem. Theory Comput.*, 2025.

Thomas, N., Smidt, T., Kearnes, S., Yang, L., Li, L., Kohlhoff, K., and Riley, P. Tensor field networks: Rotation-and translation-equivariant neural networks for 3d point clouds. *arXiv preprint arXiv:1802.08219*, 2018.

Thorne, K. S. Multipole expansions of gravitational radiation. *Rev. Mod. Phys.*, 52(2):299, 1980.

Tiefenbacher, M. X., Bachmair, B., Chen, C. G., Westermayr, J., Marquetand, P., Dietschreit, J. C., and González, L. Excited-state nonadiabatic dynamics in explicit solvent using machine learned interatomic potentials. *Digit. Discov.*, 4(6):1478–1491, 2025.

Toukmaji, A. Y. and Board Jr, J. A. Ewald summation techniques in perspective: a survey. *Computer physics communications*, 95(2-3):73–92, 1996.

Tully, J. C. Molecular dynamics with electronic transitions. *J. Chem. Phys.*, 92(2):1061–1071, 1990. doi: 10.1063/1.459170.

Tully, J. C. Nonadiabatic molecular dynamics. *Int. J. Quantum Chem.*, 40(S25):299–309, 1991. doi: https://doi.org/10.1002/qua.560400830. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/qua.560400830`.

Tzeliou, C. E., Mermigki, M. A., and Tzeli, D. Review on the qm/mm methodologies and their application to metalloproteins. *Molecules*, 27(9):2660, 2022.

Wells, B. A. and Chaffee, A. L. Ewald summation for molecular simulations. *J. Chem. Theory Comput.*, 11(8): 3684–3695, 2015.

Zhao, X., Merritt, I. C., Lei, R., Shu, Y., Jacquemin, D., Zhang, L., Xu, X., Vacher, M., and Truhlar, D. G. Nonadiabatic coupling in trajectory surface hopping: accurate time derivative couplings by the curvature-driven approximation. *J. Chem. Theory Comput.*, 19(19):6577–6588, 2023.

A. Ground State Dynamics: Nickel Complexes

To illustrate some representative applications, FieldMACE is applied to electrostatic embedding for ML/MM simulations. In such a scheme, the total energy is written as

$$E_{\text{tot}}(R_{\text{QM}}, R_{\text{MM}}) = E_{\text{MM}}(R_{\text{MM}}) + E_{\text{QM}}(R_{\text{QM}}) + E_{\text{QM/MM}}(R_{\text{QM}}, R_{\text{MM}}), \tag{20}$$

where E_{MM} describes the MM-MM interactions and is evaluated with a classical force field, E_{QM} is the energy of the QM region in isolation, and $E_{\text{QM/MM}}$ is the coupling term that accounts for the interaction between the QM region and the MM charges q_{MM}. In this work, the neural network models the QM region and the coulomb interaction between the embedded QM region and the MM region $E_{\text{QM}} + E^{C}_{\text{QM/MM}}$. The other interactions between the QM and MM regions are modeled with a classical MM force field. The forces on the QM and MM atoms induced by the coulomb potential are obtained by differentiating this learned energy with respect to R_{QM} and R_{MM}. Further details of how the model computes the gradients of the MM region with respect to the interaction with the QM region are provided in the methods section.

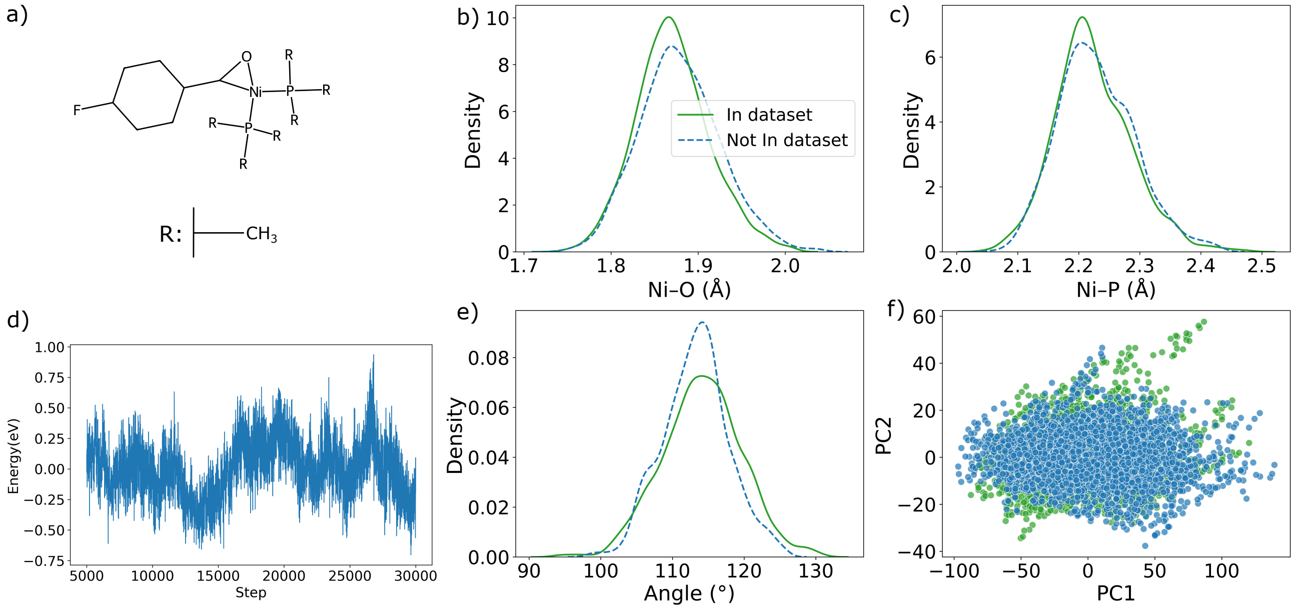

Figure 4. Nickel phosphine complex with methyl substituents in explicit benzene solvent. (a) Structure of the nickel complex. (b) Distribution of Ni-O bond lengths and (c) distribution of Ni-P bond lengths over the course of the simulation for the models trained with and without the methyl substituent. (d) Time evolution of the predicted energy during the molecular dynamics simulation. (e) Distribution of P-Ni-P bond angles for the models trained with and without the methyl substituent. (f) Projection of the sampled configurations onto the first two principal components.

As an initial ground-state example, we investigate nickel phosphine complexes, which are widely used as homogeneous catalysts in cross-coupling and related organometallic transformations (Han, 2013; Diccianni & Diao, 2019). Each complex consists of two phosphine ligands and a third neutral donor ligand, as illustrated in Figures 4a and 5a. Both systems are explicitly solvated in benzene. We consider two variants differing in the substituents on phosphorus: methyl and isopropyl.

For each system, the ML/MM dynamics were run at 300 K for 15 ps. To train either model we start from the MACE-MP-0 representation and further train the models using the relevant nickel complexes, all nickel complexes used to train the model can seen in ref (Pultar et al., 2025). Full simulation details and dataset descriptions are provided in Section S6c of the SI. Scatter plots of the trained models show accurate predictions for both QM/MM energies and forces Figure S6a. The total energy traces after 2.5 ps of equilibration are shown in Figures 4d and 5d. In both cases, the total energy fluctuates around a stable mean without noticeable drift, and the magnitude of fluctuations is similar for both systems. This indicates that the ML/MM dynamics with electrostatic embedding produces numerically stable trajectories for these relatively difficult transition metal complexes in an explicit benzene environment.

To evaluate generalization, we assessed performance on these complexes by comparing dynamics obtained from a model trained with the target system and a model trained without the target system but similar structures. The target systems can be

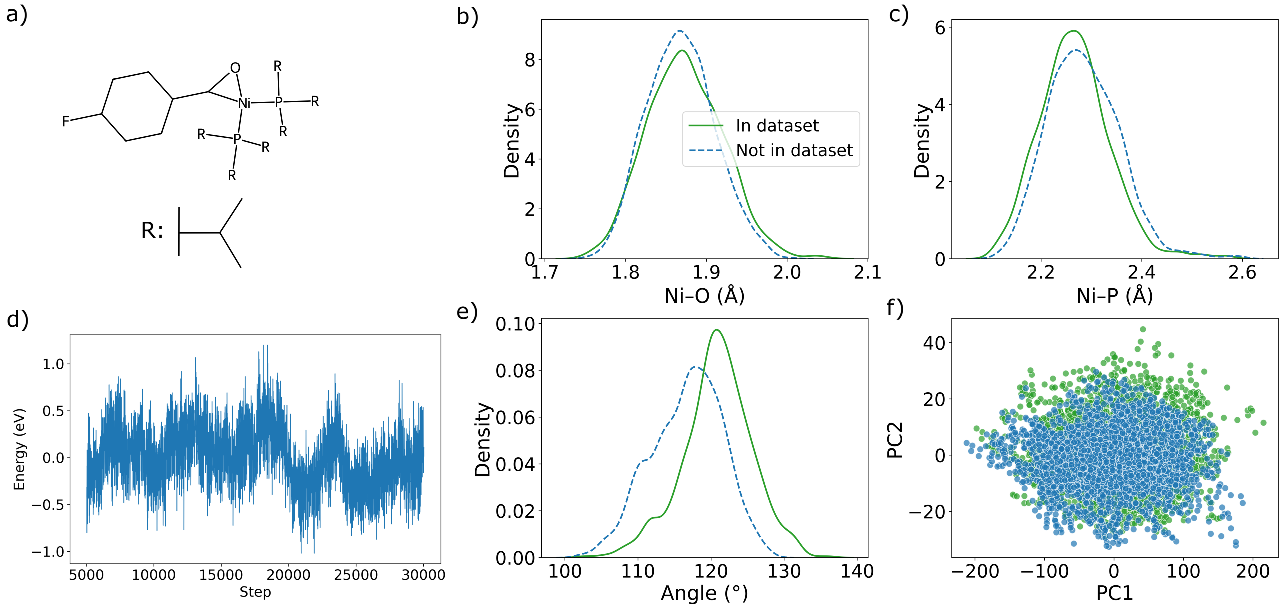

Figure 5. Nickel phosphine complex with isopropyl substituents in explicit benzene solvent. (a) Structure of the nickel complex. (b) Distribution of Ni-O bond lengths and (c) Distribution of Ni-P bond lengths over the course of the simulation for the models trained with and without the isopropyl substituent. (d) Time evolution of the predicted energy during the molecular dynamics simulation. (e) Distribution of P-Ni-P bond angles for the models trained with and without the isopropyl substituent. (f) Projection of the sampled configurations onto the first two principal components.

seen in Figure 4a and 5a. Illustrations of all similar structures can be seen in ref (Pultar et al., 2025). We monitored several key internal coordinates and compared their distributions between the two cases. First, we examined the Ni-P bond distance for the phosphine ligands directly coordinated to the nickel center. For both methyl and isopropyl substituents (Figures 4a and 5a), the Ni-P distributions from the model without the relevant system closely match those from the model that include, see figures 4c and 5c. Additionally, both models display a shorter Ni-P bond for the methyl-substituted complex, consistent with the smaller steric size of the methyl group allowing the phosphines to approach the nickel center more closely and form a slightly stronger bond. Next, we analyzed the Ni-O distance to the neutral donor ligand, see figures 4b and 5b. For both substituents, this distance remains nearly constant throughout the simulations and agrees closely between models trained with and without the relevant system, again demonstrating that the model generalizes effectively to unseen systems. Finally, we examined the P-Ni-P bond angle, see figures 4e and 5e. For the methyl-substituted complex, both models yield nearly identical angular distributions. In contrast, for the isopropyl system, the model, which was not trained on iospropyl, shows a slight shift toward larger bond angles. This is likely due to the training data containing complexes with bulkier substituents and larger bond angles, slightly biasing the model toward predicting larger angles.

Because internal coordinates cannot fully capture the full configuration space, we also performed principal component analysis (PCA) on SOAP (Himanen et al., 2020) descriptors computed for structures sampled from the dynamics. For both methyl and isopropyl-substituted complexes, the resulting PCA projections overlap closely, as shown in figures 4f and 5f confirming that the model generates dynamics highly similar to those obtained with system-specific training. Overall, this nickel complex case study demonstrates that the electrostatically embedded ML/MM approach yields stable ground-state dynamics for transition metal systems and generalizes well to new, unseen systems. Additionally, to illustrate the limitations of our model, we simulated the same system with a cyclohexyl substituent, as shown in the Appendix. In this case, the difference between the model trained on the cyclohexyl substituted system and the one without substitution was larger, likely due to the larger size and increased complexity of the interactions in the system. The detailed training errors associated with the nickel site are provided in Appendix.

A.1. Furan Excited State Dynamics

Excited states play a central role in many photochemical and photophysical processes, such as biological light harvesting (Cerullo et al., 2002; Romero et al., 2017) and converting solar energy (Mazouin et al., 2022; Mathew et al., 2014).

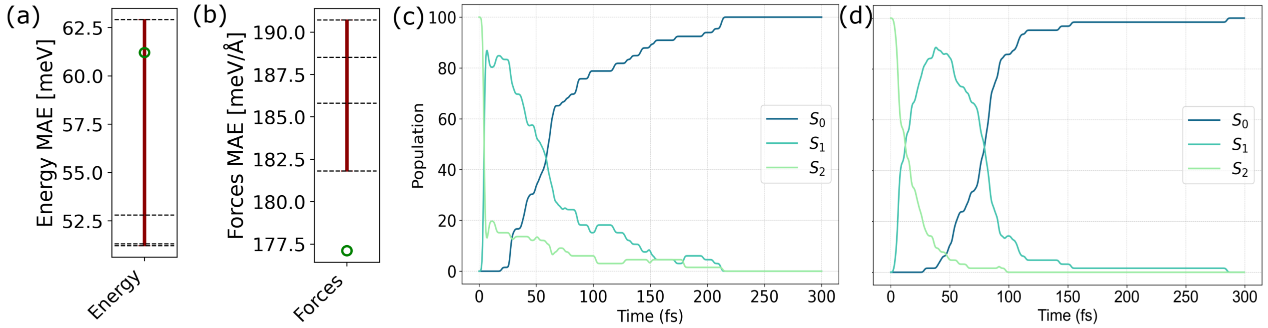

Figure 6. Model accuracy distributions and photodynamics simulation results. Distribution of (a) energy and (b) force mean absolute errors (MAEs) for the Field-MACE models, with dotted lines representing errors for each ℓ cutoff value starting from 0 and going down to 3. The green marker indicates the error associated with the Coulomb potential. (c) Population curves resulting from photodynamics simulations using the Field-MACE model and (d) quantum chemistry, for the ring-opening reaction of furan following excitation to the second excited singlet state (S_2).

Understanding the potential energy surfaces of these electronic states is essential for accurately modeling how molecules behave under light exposure or during electronic transitions. As a test case for excited states, we simulate the photodynamics of furan solvated in water. The key feature of this excited state reaction is the ring-opening reaction involving the C-O bond in the five membered ring.

Panels a) and b) of Fig. 6 summarize energy and force errors, respectively. For energies, higher multipole truncation levels achieve the lowest errors, consistent with the increased angular resolution of the environment. For forces, the scalar Coulomb descriptor yields the lowest MAE among the tested long-range variants; however, although force errors remain uniformly large across all approaches, indicating either that the dataset may be insufficiently sampled or that the underlying potential energy surface contains intricate topological features such as conical intersections (points where potential energy surfaces approach each other and form cusps), which are inherently difficult to approximate accurately.

To assess whether the learned surfaces are nevertheless adequate for dynamics, we performed nonadiabatic molecular dynamics (NAMD) with Field-MACE using trajectory surface hopping (TSH) (Tully, 1990; 1991). NAMD simulations describe the behavior of a molecule after being excited by light and are computationally much more demanding than ground state simulations, especially when an explicit environment is included, since energies and forces are required for multiple electronic states along each trajectory. In addition, transitions between different excited-state potentials need to be accounted for, which can be done either via second-order derivatives (Zhao et al., 2023; Shu et al., 2022b) or nonadiabatic coupling vectors (Mai et al., 2018). TSH recovers the quantum nature of the excited wave packet by averaging over many trajectories that have different initial conditions (*i.e.* configurations and momenta). In Tully's Fewest Switches TSH (Tully, 1990; Granucci & Persico, 2007), used here, each trajectory changes its active state via so-called hops stochastically. For the photodynamics simulations, we integrated our Field-MACE approach into the SHARC (Surface Hopping including Arbitrary Couplings) program suite (Richter et al., 2011; Mai et al., 2023; 2015; 2018). Details on the molecular dynamics, data, and model training are specified in Section S4 in the SI.

Panels c) and d) of Fig. 6 show the time evolution of the average population of the electronic states obtained from simulations with Field-MACE (panel c) and the quantum chemical reference method (panel d). The two sets of simulations were initiated from the identical set of initial conditions. At the beginning of the simulations, the system is excited to the second excited singlet state, S_2. Both sets of simulations exhibit rapid non-radiative decay toward lower states, with fast transfer from S_2 to S_1 followed by relaxation to the ground state S_0. Field-MACE reproduces both the rapid initial decay of S_2 and the subsequent shift from S_1 to S_0 in close agreement with the quantum-chemical reference: The S_1 state is populated until around 100 fs, at which point the majority of the trajectories fall back into the ground state. Notably, Field-MACE yielded stable trajectories across a range of training-set sizes (see next section), whereas existing approaches such as FieldSchNarc (Tiefenbacher et al., 2025) exhibit much larger sensitivity to hyperparameters and difficulties in obtaining stable dynamics for this system.

A.2. Transferability to excited states

To further test whether Field-MACE can also benefit from transfer learning in excited-state simulations, we revisit the furan-in-water system introduced above. Full details on the model and training can be found in Section S4.3 in the SI. Rather than considering the raw MAE values, we evaluate the learned models by running NAMD for the same set of initial conditions used previously and compare the resulting population curves as the amount of training data is reduced. Models trained from scratch with randomly initialized parameters are compared to models whose short-range block parameters are taken from a MACE-OFF model. We consider training set size of 600, 150, and 30 frames, compared to the 1,900 data points used in the previous section on furan. Since excited-state data is computationally much more costly to obtain than ground state data, this is a realistic use case. The results are shown in Fig. 7.

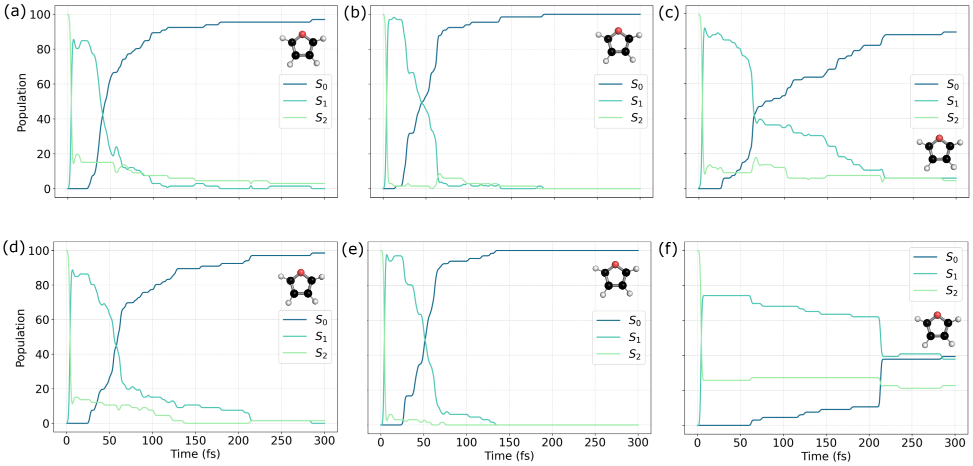

Figure 7. Population curves of furan starting in the second excited singlet state for transfer learning models using (a) 600, (b) 150, and (c) 30 data points and non-transfer learning models using (d) 600, (e) 150, and (f) 30 data points.

For the largest dataset with 600 data points, Fig. 7 first column, both transfer a) and non-transfer models d) result in similar population curves that align well with the reference populations (Fig. 6 c). Deviations in the S_2, S_1, and S_0 populations remain minimal throughout the simulation. When the dataset size is reduced to 150 points (middle column of Fig. 7), the predictions are reasonable for both transfer (panel b)) and non-transfer (panel e)) models.

In contrast, for the smallest dataset of 30 data points (right-most column in Fig. 7), a stark difference emerges between the two initialization strategies. The transfer model (panel c)), leveraging the foundational representation, maintains a reasonable albeit worse approximation of the population curves, but captures the general trends in decay and redistribution among electronic states. However, the model trained from scratch (panel f)) fails to capture the transition dynamics or the populations of S_2, S_1, and S_0. These results highlight that transfer learning becomes critical in low-data regimes, where the foundational representation provides a strong inductive bias that enables the model to generalize despite very limited excited-state training data.

Overall, the results demonstrate that transfer learning significantly enhances the performance of Field-MACE, particularly when QM/MM training data is scarce. With sufficient data (600 or 150 data points), both transfer and non-transfer models can perform reasonably well, though transfer learning provides an edge in accuracy. In low-data scenarios (30 data points), transfer learning is essential for recovering qualitatively correct nonadiabatic population dynamics.

B. Model Inference Timings and Data Generation

The tables below (Tab 4, 5, and 6) compare inference timings of the foundational MACE-MP-0 model (medium size) and the Field-MACE model for nickel complexes solvated in benzene, using systems with 75, 100, and 125 solvent molecules. The Field-MACE model and MACE-MP-0 models have similar parameterizations. These correspond to systems of approximately 1000, 1300, and 1600 atoms in total. All timings were measured on a single NVIDIA H100 GPU and averaged over 50 forward passes.

For the foundational MACE-MP-0 model, the entire system (including both the nickel complex and all solvent molecules) is treated explicitly. As a result, the computational cost scales roughly linearly with the total number of atoms. This scaling quickly becomes prohibitive for large solvated systems. In practice, systems exceeding roughly 100 solvent molecules (about 1300 atoms total) frequently encounter out-of-memory errors when evaluated on a single GPU. For even larger environments, stable inference would require the parallelization of model partitioning.

In contrast, Field-MACE is applied only to the QM region, while the surrounding solvent is incorporated implicitly through the multipole expansion of the electrostatic field. During the construction of this field, the MM point charges associated with the solvent molecules are aggregated into multipole moments at an early stage. Hence, the number of MM atoms has only a minor influence on the overall computational cost. For a fixed QM region, the inference time of Field-MACE remains effectively constant as the size of the MM environment increases, in stark contrast to the linear scaling observed for the foundational model.

Table 4. Comparison of inference times (seconds) for one Field-MACE and MP-0 medium foundational model prediction for nickel complexes with 75 solvent molecules. Timings were averaged over 50 predictions on an H100 GPU.

System	**Foundational Medium MP-0 Model**	**Field-MACE**
Methane	0.146	0.0316
Isobutane	0.655	0.0324
Hexane	0.285	0.0318

Table 5. Comparison of inference times (seconds) for one Field-MACE and MP-0 medium foundational model prediction for nickel complexes with 100 solvent molecules. Timings were averaged over 50 predictions on an H100 GPU.

System	**Foundational Medium MP-0 Model**	**Field-MACE**
Methane	0.405	0.0319
Isobutane	0.970	0.0329
Hexane	0.658	0.0324

Table 6. Comparison of inference times (seconds) for one Field-MACE and MP-0 medium foundational model prediction for nickel complexes with 125 solvent molecules. Timings were averaged over 50 predictions on an H100 GPU.

System	**Foundational Medium MP-0 Model**	**Field-MACE**
Methane	0.826	0.0331
Isobutane	-	0.0322
Hexane	-	0.0321

It is important to note that the total training time for both the foundational and Field-MACE models should include the cost of generating the underlying QM/MM training data. The data generation typically dominates the total computational expense, as it requires performing high-level quantum-chemical calculations for representative configurations of the system. However, the amount of training data required varies substantially depending on the chemical complexity of the system, the size of the QM region, the chosen level of theory, and the availability of transferable pretrained models. In particular, the use of transfer learning or model fine-tuning can drastically reduce the number of expensive QM calculations needed, often by one or more orders of magnitude, while maintaining accuracy comparable to that of a model trained completely from scratch.

C. Transferability between Similar Systems

In Figure 8 we examine how transfer learning affects the data efficiency of Field-MACE for a small series of related quinoline dimer systems solvated in dichloromethane. Full details of the underlying QM/MM dataset are reported in Ref. (Riniker et al., 2024). The molecule in panel (e) is used as the initial system that Field-MACE model is first trained on. This pretrained model is then fine-tuned on two substituted target systems shown in panels (f) and (g). For each target, we randomly subsample different fractions of the available QM/MM configurations and compare the resulting mean absolute errors (MAEs) in energies and forces against models trained from scratch on exactly the same data. Panels (a) and (b) report the learning curves for the system in panel (f), while panels (c) and (d) show the corresponding results for the system shown in panel (g).

In both cases, pretraining on similar quinoline dimer systems in Figure 8 panel (e) leads to a clear reduction in error in the low-data regime. The transfer-learned models achieve near-full-dataset accuracy using only a fraction of the target-system configurations, whereas models trained from scratch require substantially more data to reach comparable performance. As one might expect, the advantage of transfer learning diminishes as the fraction of target-system data increases and the models become increasingly dominated by direct supervision on the target system. This demonstrates that knowledge learned on one embedded system can be reused to accelerate training on closely related environments.

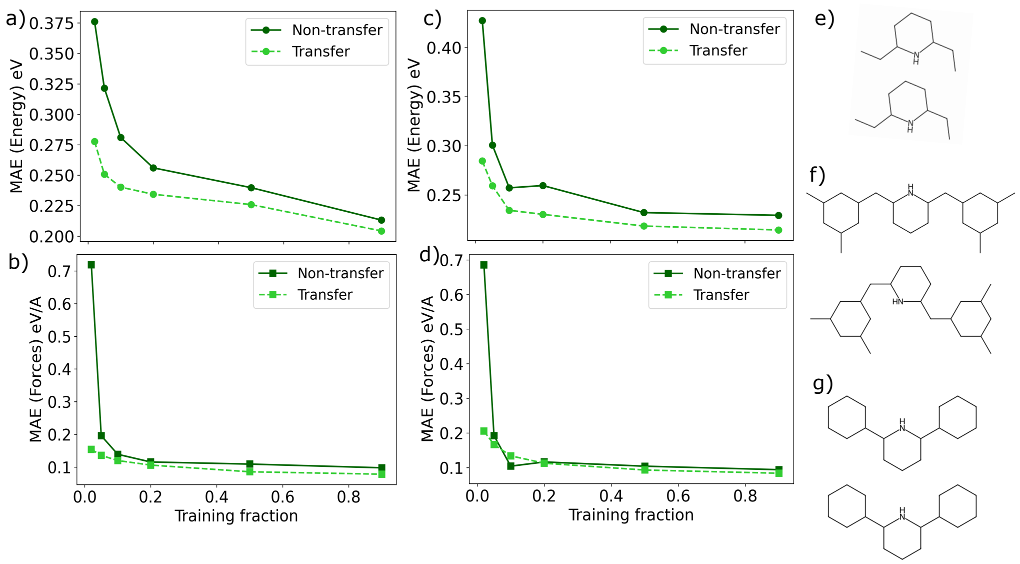

Figure 8. Transfer-learning performance for a series of quinoline dimers systems solvated in dichloromethane. Errors in energies and forces are shown as a function of the fraction of target-system data used for training. In all cases, the transfer-learning models are initialized from a Field-MACE model pretrained on the system shown in panel (e). (a) Energy MAEs for the target system in panel (f). (b) Force MAEs for the target system in panel (f). (c) Energy MAEs for the target system in panel (g). (d) Force MAEs for the target system in panel (g).

D. Transferability between Biological and Solvent environments

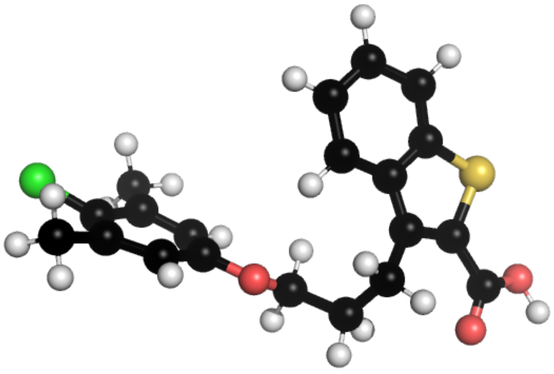

Figure 9. Structure of the small-molecule inhibitor 19G bound to the protein MCL1.

In this section, we test how well Field-MACE can transfer between two different environments for the same ligand molecule, namely bulk water and the binding pocket of a protein. As a representative example, we consider the small-molecule inhibitor 19G bound to the protein MCL1, illustrated in Fig. 9. We used two separate QM/MM datasets, which were obtained from ref. (Bensberg et al., 2025), one describing the ligand solvated in water and the ligand in the protein pocket.

As a baseline, independent Field-MACE models were first trained from scratch on each dataset. The scatter plots in Fig. 10b and c show that in both environments the final models reproduce the reference QM/MM energies with small errors, confirming that Field-MACE can accurately learn ligand energetics in both solvent and protein environments when trained directly on the corresponding data.

To assess cross-environment transferability, we then performed transfer-learning experiments in both directions. In Figs. 10d and e, a model pre-trained on the ligand-in-water dataset is fine-tuned on the protein-bound ligand data and compared to a model trained from scratch using only protein configurations, while Figs. 10f and g show the reverse scenario, where a model pre-trained on the protein dataset is fine-tuned on the solvated ligand data. In both directions, the learning curves for energies and forces remain very similar, with at most a modest improvement at the smallest training fractions. This suggests that, for this particular system, the protein environment induces interaction patterns that are sufficiently different from those in bulk solvent that pre-training provides only limited additional benefit. We expect stronger gains from pretraining when starting from a foundational model representation as seen in the main text.

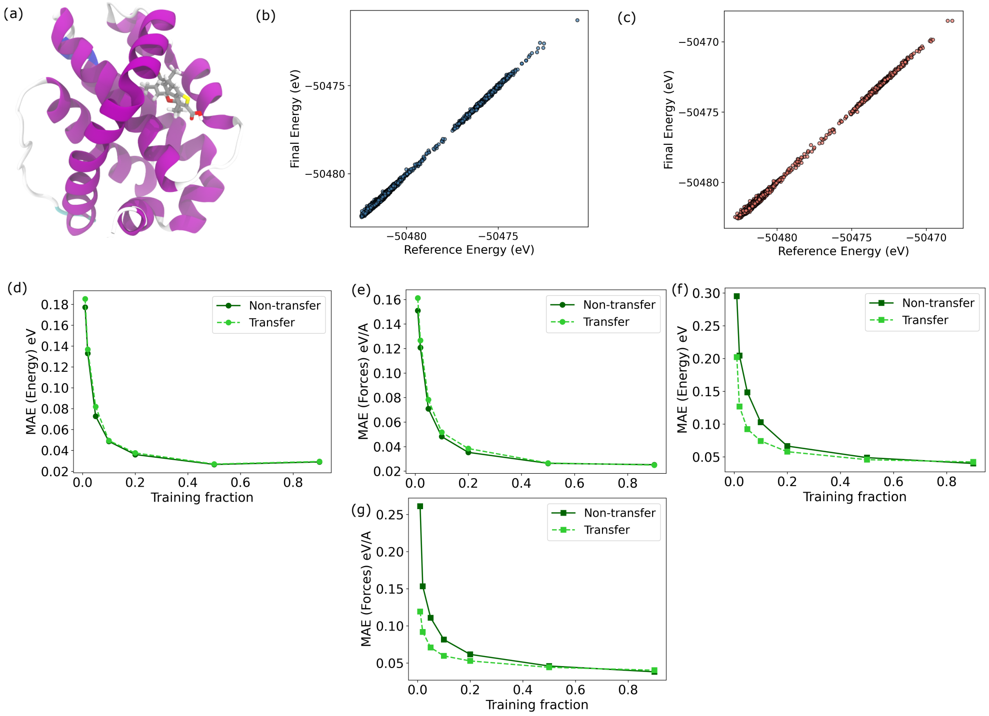

Figure 10. Transfer learning between solvent and biological environments for the 19G ligand. (a) Illustration of the MCL1-19G protein complex with the protein represented by its secondary structure elements and the small ligand in atomistic detail, with hydrogen, carbon, oxygen, and sulfur atoms in white, gray, red, and yellow, respectively. Parity plots of the predicted energies of the model for (b) the protein complex and (c) the ligand in water. (d,e) Learning curves for energy and force MAE, respectively, when training on the protein dataset either from scratch or by fine-tuning a model pre-trained on the solvent data. (f,g) Corresponding learning curves when training on the solvent dataset, comparing models trained from scratch to those fine-tuned from a model pre-trained on the protein data.

E. Ground State Simulation Details

E.1. Nickel Model Validation

Figures 11 and 12 show scatter plots comparing reference and predicted energies, QM forces, and MM Coulomb forces for the nickel complexes.

Both the models trained on the full dataset and the model without the target system exhibit good agreement with the reference data. In particular, energy and QM force errors remain small even when one substituent class is excluded from training, indicating robust generalization across closely related nickel complexes. The MM Coulomb force errors are several orders of magnitude smaller than the QM force errors, reflecting the relatively smooth dependence of the electrostatic interaction on MM atom positions.

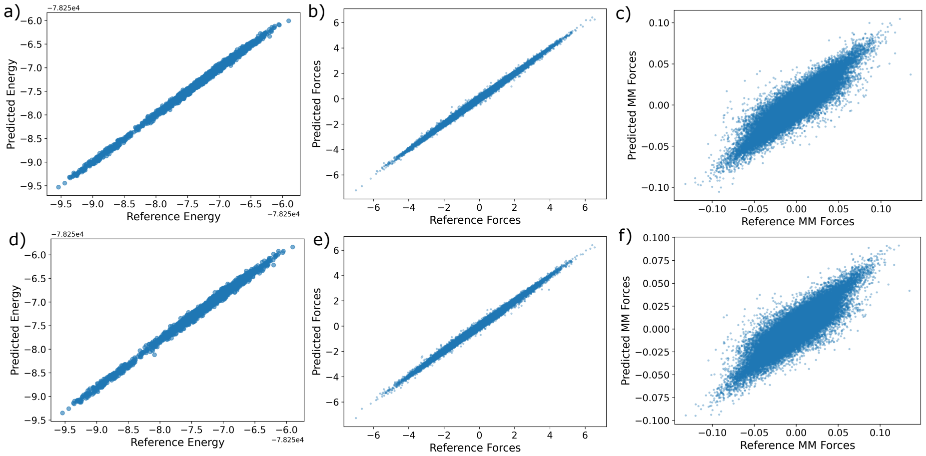

Figure 11. Scatter plots for the nickel complexes with methyl substituents. The upper row corresponds to scatter plots trained on all systems and the lower row corresponds to models trained without the methyl substituent. (a,b) Reference versus predicted energies. (c,d) Reference versus predicted QM forces. (e,f) Reference versus predicted Coulomb forces acting on MM atoms.

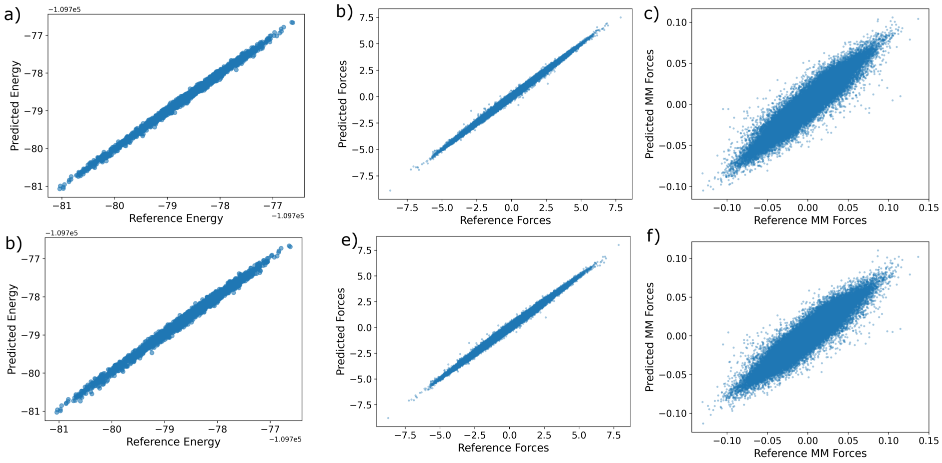

Figure 12. Scatter plots for the nickel complexes with isopropyl substituents. The upper row corresponds to scatter plots trained on all systems and the lower row corresponds to models trained without the isopropyl substituent. (a,b) Reference versus predicted energies. (c,d) Reference versus predicted QM forces. (e,f) Reference versus predicted Coulomb forces acting on MM atoms.

Table 7. Mean absolute errors (MAEs) for energies and forces on the validation set. Energies are given in eV and forces in eV Å^{-1}.

Quantity	**Full**	**w/o Methyl**	**w/o Isopropyl**
Energy	0.05	0.063	0.059
QM forces	0.037	0.040	0.040
MM forces	0.00032	0.00037	0.00038

E.2. Nickel Dynamics Cyclohexyl Substituent

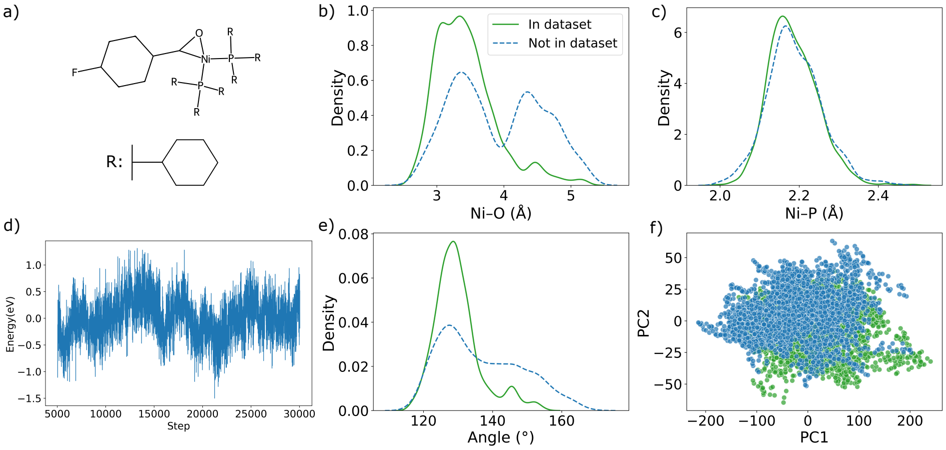

Figure 13. Nickel phosphine complex with cyclohexyl substituent in explicit benzene solvent. (a) Structure of the nickel complex with the hexane substituent. (b) Distribution of the of Ni-O bond lengths and (c) distribution of the of Ni-P bond lengths over the course of the simulation for the models trained with and without the hexane substituent. (d) Time evolution of the predicted energy during the molecular dynamics simulation. (e) Distribution of the of P-Ni-P bond angles for the models trained with and without the hexane substituent. (f) Projection of the sampled configurations onto the first two principal components.

In addition to the methyl- and isopropyl-substituted systems discussed in the main text, we also simulated the cyclohexyl-substituted system for 15 ps. During the trajectory, we monitored the same internal coordinates as before: the Ni–O bond length, Ni–P bond lengths, and P–Ni–P bond angles. The distributions of the Ni–P bond distances are very similar between the two models. In contrast, the Ni-O bond length is substantially larger, leading to a dissociation, for the model that was not trained on this system. This likely reflects an underestimated dissociation barrier that is not well represented by the other nickel complexes in the dataset. This, in turn, has an effect on the P-Ni-P bond angle, which spans a much wider range than in simulations produced by the model that included the system during training.

Overall, this example suggests that while the Field-MACE generalizes well to simpler systems, it may struggle to capture behavior in larger systems than those present in the training set, particularly in cases where dissociation barriers are small and highly sensitive to modest prediction errors or rare events.

E.3. Molecular Datasets

The configurations, annotated with energies and forces for the four molecules solvated in water in the ground state were taken from ref (Boöselt et al., 2021). Simulation details are briefly summarized here: All simulations were performed using a QM/MM approach with the DFT functionals BP86 (Becke, 1988) for retinoic acid and the SAM system, ωB97X-D3 (Becke, 1993; Grimme et al., 2011) for uracil, and B2-PLYP (Grimme, 2006) for benzene; the def2-TZVP basis set was used for all calculations. The resolution of identity (RI) approximation (Dunlap et al., 1979), Grimme's dispersion correction with Becke–Johnson damping (Grimme et al., 2011), and electrostatic embedding were used to account for long-range interactions. The MM region consists of water molecules modeled with the SPC/E water model (Berendsen et al., 1987). The QM/MM Hamiltonian incorporated MM point charges within a cutoff radius of 0.6 nm for benzene and 1.4 nm for uracil, retinoic acid and the SAM system.

Each solute was solvated in a water box, with the number of MM partial charges included approximately 1600 for uracil, 2600 for retinoic acid and 3500 for the SAM system, while benzene had a smaller number of MM partial charges of 250 due to the smaller box size. The initial structures were generated using the ATB server (Malde et al., 2011) and relaxed at 0 K

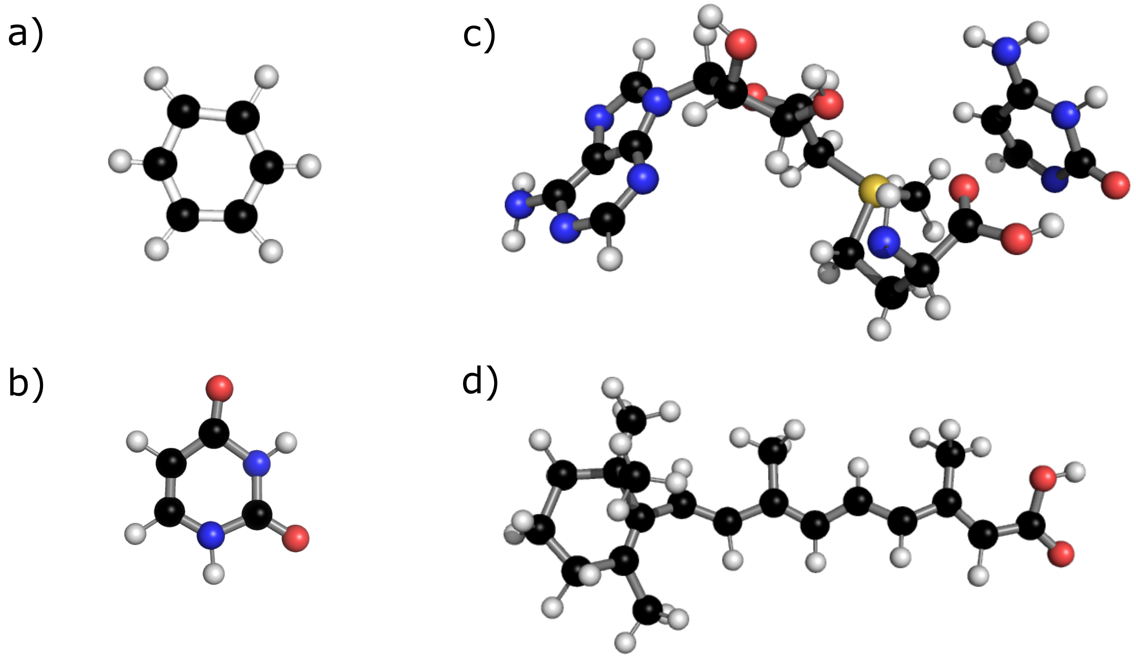

Figure 14. The following shows images of the molecules used in the paper, a) benzene, b) uracil c) SAM system, d) retinoic acid

with a gradient descent algorithm until the predicted energy change fell below 0.1 kJ mol^{-1}.

F. Multipole Component Analysis

In order to better rationalize the choice of ℓ, we plotted the distribution of coefficient magnitudes for each value as a series of box plots with a logarithmic scale. Since the coefficients decay steeply as ℓ decreases, every one unit drop in ℓ typically corresponds to roughly an order of magnitude decrease. For the SAM system, uracil, and retinoic acid, the coefficients at $\ell = 5$ already fall below 10^{-7}. Numbers in this range brush up against the roughly 7 significant digit resolution of single-precision (float32) arithmetic, so rounding noise may become non negligible. However, at such small values the influence of higher ℓ values will be negligible for the prediction, as seen when looking at $\ell = 4$ and $\ell = 5$ compared to lower order ℓ values.

F.1. Training and Model Details

In all trainings, a 5 Ångstrom cutoff was applied to the ML region to construct the local graph. The training parameters shown in the Field-MACE section below were also used for the transfer learning, dynamics and the ablation study models shown, with the exception that the channel size was fixed to 128 for the transfer learning/dynamics models and 32 for the ablation studies. Additionally the multipole parameter ℓ value was fixed to 3 for the transfer learning study and 1 for the dynamics models. For the transfer learning and dynamics models (Heydari et al., 2023; Buterez et al., 2024) the short-range MACE blocks were initialized using the medium size foundational MACE-OFF model parameters for the transfer learning. Additionally, the other blocks such as the long-range blocks and the readouts were randomly initialized.

All trainings were performed on a single NVIDIA H100 GPU. For all Field-MACE models, we used two sets of short and long-range layers. We trained models with the initial learning rate of 0.01 with a decay factor of 0.5 occurring after 10 consecutive epochs of no change in validation error. All models used for benchmarking were trained for 1000 epochs. A weighting of 100:100 was applied to the energies and forces during the training. Additionally, all MACE models were trained with spherical harmonics up to and including degree 1 and 8 Bessel basis functions.

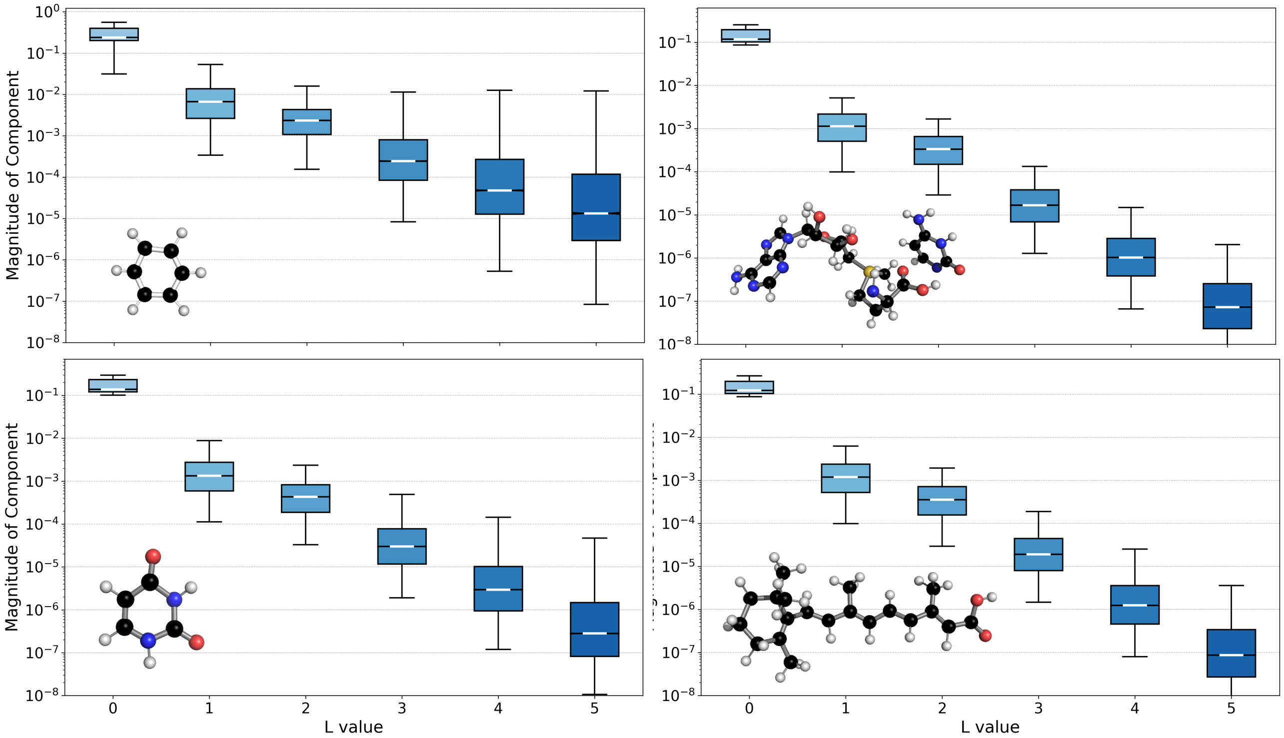

Figure 15. Box plot visualization showing the distribution of the multipole moment coefficients (ℓ values) for benzene, retinoic acid, uracil, and the complete system.

All models were trained with the loss function:

$$L = c_E \sum_{i=1}^{n} \big(E_{REF}^{(i)} - \hat{E}^{(i)}\big)^2 + c_F \sum_{i=1}^{n} \big(\mathbf{F}_{REF}^{(i)} - \hat{\mathbf{F}}^{(i)}\big)^2 \tag{21}$$

The coefficients c_E and c_F were both set to 100. The optimization was performed using the default optimizer in the MACE library with an initial learning rate of 0.01 as mentioned previously.

F.2. Point charge gradient computation

As introduced in the main text, the QM/MM reference calculations used to generate the training data represent the MM environment by a set of fixed point charges $(q_j^{\mathrm{MM}}, \mathbf{r}_j^{\mathrm{MM}})$, which enter the electronic Hamiltonian as an external electrostatic potential. Therefore, in an electrostatic embedding framework, the interaction of the QM and MM subsystems must also be considered. In particular, because the coupling of the two subsystems gives rise not only to forces on the QM atoms but also to forces acting on the MM atoms. The forces on the MM atoms are obtained as the negative gradient of the total QM/MM energy with respect to their positions $\mathbf{r}_j^{\mathrm{MM}}$,

$$\mathbf{F}_j^{\mathrm{MM}} = -\frac{\partial E_{\mathrm{QM/MM}}}{\partial \mathbf{r}_j^{\mathrm{MM}}}. \tag{22}$$

In Field-MACE, during training, the model is trained to reproduce the QM/MM energy, QM atomic forces, and, if available, MM atomic forces. Internally, the forces on the MM atoms are computed through the chain rule, first differentiating the energy with respect to the multipole coefficients, and then the multipole coefficients with respect to the MM positions. Formally, this can be written as

$$\mathbf{F}_j^{\mathrm{MM}} = -\frac{\partial E_{\mathrm{ML}}}{\partial \mathbf{r}_j^{\mathrm{MM}}} = -\sum_{\ell,m,i} \frac{\partial E_{\mathrm{ML}}}{\partial \Phi_{\ell m}^{i}} \frac{\partial \Phi_{\ell m}^{i}}{\partial \mathbf{r}_j^{\mathrm{MM}}}. \tag{23}$$

In practice, all derivatives are evaluated automatically using PyTorch's automatic differentiation engine. Both the MACE architecture and the multipole feature construction are implemented within a single computational graph, allowing gradients to be obtained by standard backpropagation without any additional analytical derivations, ensuring that the forces on QM and MM atoms are consistent by construction. It is worth noting that the MM forces can still be computed even in the absence of explicit MM force labels during training, albeit with reduced accuracy compared to models trained with full QM/MM force information. Nevertheless, the availability of gradients with respect to MM coordinates enables stable ML/MM molecular dynamics simulations within an electrostatic embedding framework.

F.3. Nickel Simulation Details

All ML/MM molecular dynamics simulations were performed on a nickel phosphine complex with a neutral donor ligand embedded in an explicit benzene environment. The reference data and initial structures were taken from Ref. (Riniker et al., 2024), with the starting configuration selected to be the first structure of the corresponding dataset.

As an initial preparation step, the isolated nickel complex was relaxed using Field-MACE until the maximum force on any atom was below 0.1 eV Å^{-1}. To construct a physically reasonable solvent environment, benzene molecules exhibiting unphysical overlaps with the nickel complex were removed prior to simulation, full prior construction of the initial conditions can be found in ref (Pultar et al., 2025). More concretely, any benzene molecule whose closest atom lay closer than 3.5 Å to any atom of the QM region was discarded. The structures obtained this way served as starting points for the subsequent ML/MM simulations. The spatial extent of all atoms was determined, and uniform vacuum padding of 2 Å was added in each Cartesian direction. A cubic simulation box was then defined with side length equal to the largest system extent plus twice this padding. The structure was recentered within the box, and periodic boundary conditions were applied in all three dimensions.

Prior to production dynamics, the QM region was relaxed once more at the ML/MM level in the presence of the frozen MM environment, using the FieldMACE model trained on the relevant complexes, the same model as in the production MD run. Geometry optimization was performed using a quasi-Newton algorithm until the maximum force acting on any QM atom was again below 0.1 eV Å^{-1}. This optimized configuration was used as the initial structure for the molecular dynamics trajectories. The MM region consisted exclusively of rigid benzene molecules such that intramolecular interactions could be omitted.Intermolecular MM-MM interactions were described by the sum of Lennard Jones and Coulomb potential acting only on pairs of atoms belonging to different benzene molecules. The MM energy is given by

$$E_{\text{MM}} = \sum_{i<j}^{\text{MM}} \left[4\varepsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right) + k_e \frac{q_i q_j}{r_{ij}} \right], \tag{24}$$

where $k_e = 14.3996$ eV Å/e^2 and r_{ij} denotes the interatomic distance. Lennard Jones parameters for carbon and hydrogen were chosen to be compatible with a simple OPLS-AA-type benzene parameterization, hydrogen and carbon with an epsilon parameters 0, 0.0030 and sigma parameters 0, 3.55 respectively. Parameters on the hydrogens were omitted. Partial atomic charges were assigned such that each benzene molecule remained charge neutral, charges of -0.1299 for the carbons and 0.1299 for the hydrogen, consistent with the dataset for QM training.

The QM region was described using a machine-learned neural network potential evaluated on GPU. In the electrostatic-embedding scheme, the QM energy depends explicitly on the positions and charges of the surrounding MM atoms, yielding an embedded QM energy of the form

$$E_{\text{ML}}(R_{\text{QM}}; R_{\text{MM}}, q_{\text{MM}}) = E_{\text{QM}}(R_{\text{QM}}) + E^{C}_{\text{QM/MM}}(R_{\text{QM}}, R_{\text{MM}}, q_{\text{MM}}), \tag{25}$$

along with forces on both QM atoms and MM atoms arising from the electrostatic interaction. Here $E^{C}_{\text{QM/MM}}(R_{\text{QM}}, R_{\text{MM}}, q_{\text{MM}})$ is the interaction energy resulting from the coulomb interaction. The van der Waals interactions between the QM and MM regions were modeled using Leonard Jones parameters taken from the Universal force field (UFF) (Rappé et al., 1992).

The total potential energy of the system was therefore approximated as

$$\begin{aligned} E_{\text{tot}}(R_{\text{QM}}, R_{\text{MM}}) = {} & E_{\text{MM}}(R_{\text{MM}}) + E_{\text{ML}}(R_{\text{QM}}; R_{\text{MM}}, q_{\text{MM}}) \\ & + E^{\text{LJ}}_{\text{QM–MM}}(R_{\text{QM}}, R_{\text{MM}}). \end{aligned} \tag{26}$$

where the final term denotes the explicit Lennard-Jones interaction between QM and MM atoms. However, this classical force field scheme can be switched for any other classical force field, this example was taken for simplicity.

All molecular dynamics simulations were carried out using a time step of 0.5 fs. Initial velocities were drawn from the Maxwell–Boltzmann distribution at a temperature of 300 K. A Langevin thermostat was applied to the QM degrees of freedom to maintain the target temperature while minimally perturbing the dynamics. For each nickel complex (methyl, isobutane, and cyclohexyl), trajectories of 15 ps picoseconds were generated.

F.4. Surface Hopping Molecular Dynamics

Trajectory surface hopping (TSH) (Tully, 1990; 1991; Richter et al., 2011; Mai et al., 2015) is a mixed quantum-classical approach to recover the behavior of the nuclear wave packet after electronic excitation. In this nonadiabatic molecular dynamics approach, the nuclei positions evolve classically on a single potential energy surface (PES), while stochastic hops between electronic states occur based on nonadiabatic coupling probabilities. The quantum nature of the system is then recovered as an average over many different initial conditions. In Tully's fewest switches surface hopping (FSSH) (Tully, 1990; 1991), electronic state populations are propagated via the time-dependent Schrödinger equation,

$$\frac{d}{dt}c_k(t) = -\sum_{\ell}\left[\frac{i}{\hbar}H_{k\ell} + \mathbf{d}_{k\ell}\cdot\mathbf{v}\right]c_\ell(t), \tag{27}$$

where $c_k(t)$ denotes the time-dependent amplitude of electronic state k, $H_{k\ell}$ is the electronic Hamiltonian matrix element of states k and ℓ, $\mathbf{d}_{k\ell} = \langle\phi_k|\nabla_R|\phi_\ell\rangle$ is the nonadiabatic coupling vector between states k and ℓ, and $\mathbf{v}$ is the nuclear velocity vector. The probability of a hop from state k to state ℓ in a time step Δt is computed according to

$$P_{k\to\ell} = \max\left(0, \frac{2\,\mathrm{Re}\left(c_n^* c_\ell\left[\frac{i}{\hbar}H_{k\ell} + \mathbf{d}_{k\ell}\cdot\mathbf{v}\right]\right)}{|c_k|^2}\Delta t\right). \tag{28}$$

At every time step, a random variable ξ is drawn from the interval $[0,1]$, and the system transitions from state k to state ℓ if ξ is less than $P_{k\to\ell}$ (Tully, 1990). Since calculating nonadiabatic couplings $\mathbf{d}_{k\ell}$ directly is resource-intensive, alternative methods like curvature-driven surface hopping (Shu et al., 2022a) have emerged, which approximate the coupling between states via second-order time derivatives of the PESs. The curvature-driven surface hopping was utilized in all surface hopping simulations in this study. These techniques enable efficient propagation of surface hopping trajectories while accurately incorporating nonadiabatic effects in simulations of excited-state molecular dynamics.

F.5. Excited State Dataset Details

The excited-state data set contains calculations of furan solvated in water (Zenodo archive `10.5281/zenodo.14536036`) and was produced for an ML/MM study using FieldSchNet (Tiefenbacher et al., 2025). Computational details are briefly summarized here: The furan molecule is solvated in a cubix box (side length 15 Å) of TIP3P water (1365 molecules). Initial structures for surface hopping dynamics were obtained from classical MD simulations. These frames were split into two sets for subsequent excited-state QM/MM simulations. One set contains the initial conditions for those trajectories that constitute the training and validation sets for model, the other initial conditions were used to compare dynamics of the trained model to unseen QM/MM simulations. We followed the same scheme to train and evaluate our models. For more details on the initial conditions generation see (Tiefenbacher et al., 2025).

F.6. Excited State Simulation and Training Details

All nonadiabatic simulations were conducted using an interface between Field-MACE and the SHARC engine (Mai et al., 2023; Richter et al., 2011). For the transfer learning experiments, the short-range blocks were initialized using the medium-size foundational MACE-OFF model parameters. The other blocks, such as the long-range blocks and the readouts, were randomly initialized. Hyperparameters listed below were kept consistent for both the full data set and the smaller subsets used in transfer learning. These smaller datasets were constructed by subsampling frames randomly from the full training set. A 5 Å cutoff radius was applied along with 8 equally spaced Bessel basis functions. Only two interaction layers were used, with 128 channels and spherical harmonics up to degree 3. The initial vectors for the MM atoms were constructed from spherical harmonics up to degree 3, combined with an embedding vector of up to size 16. The readout functions were modified to produce multiple outputs rather than a single node. In this case, 3 electronic state energies were predicted.

The loss function was defined as mean squared error over all states for a property (energy and forces)

$$\begin{aligned} L =&\frac{c_E}{n_L}\sum_{j=1}^{n_L}\sum_{i=1}^{n}\left(E_{REF}^{(i,j)} - \hat{E}^{(i,j)}\right)^2 \\ &+ \frac{c_F}{n_L}\sum_{j=1}^{n_L}\sum_{i=1}^{n}\left(\mathbf{F}_{REF}^{(i,j)} - \hat{\mathbf{F}}^{(i,j)}\right)^2 \end{aligned} \tag{29}$$

where c_E and c_F were both set to 100. The parameter n_L represents the number of energy states, here 3. For optimization we used the default optimizer in the MACE library with a learning rate of 0.001. Training was performed on a single NVIDIA H100 GPU. Full details of the parameters used for the SHARC simulations can be found in (Tiefenbacher et al., 2025).

F.7. Ewald Sum Implementation

To use the Ewald sum (Wells & Chaffee, 2015) it is necessary to have periodic system thus to evaluate the electrostatic potential from a set of point charges q_j at positions $\mathbf{R}_j$, we embed them in a cubic periodic cell whose side length

$$L = 2\,\max_j \|\mathbf{R}_j\| \tag{30}$$

ensures every point charge lies within the periodic box. Next the Coulomb interaction is split into three parts, the real term, the momentum or reciprocal term and the self interaction term. Each term is also given a cutoff value, these are shown below for the real and reciprocal term,

$$r_{\text{cut}} = \tfrac{L}{2} \quad \text{and} \quad k_{\text{cut}} \approx \frac{2\,\alpha\,L}{\pi}. \tag{31}$$

Here α corresponds to the weighting of the real space to reciprocal part. The idea here being the real part captures the close range interactions and the reciprocal term captures low frequency momentum wave vectors which normally correspond to charge distributions at a long distance. In real space, the short-range potential at point $\mathbf{R_i}$ is

$$\Phi^{(\text{real})}(\mathbf{R_i}) = \sum_j q_j\,\frac{\operatorname{erfc}(\alpha\,r_{ij})}{r_{ij}}\,H\big(r_{\text{cut}} - r_{ij}\big), \tag{32}$$

where r_{ij} is the minimum-image distance to $\mathbf{R}_j$, i.e $r_{ij} = \min_{\mathbf{n}\in\mathbb{Z}^3}\big\|\mathbf{R}_i - \mathbf{R}_j + \mathbf{n}L\big\|$, where n is the index of the periodic image, H masks beyond r_{cut}.

In reciprocal space, one sums over discrete wavevectors $\mathbf{k} = 2\pi(n_x, n_y, n_z)/L$ with $|n| \leq k_{\text{cut}}$. Defining a weight

$$W(\mathbf{k}) = \frac{4\pi}{V}\,\frac{e^{-\|\mathbf{k}\|^2/(4\alpha^2)}}{\|\mathbf{k}\|^2}, \tag{33}$$

where V is the volume of the periodic box and charge structure factors

$$C(\mathbf{k}) = \sum_j q_j\cos\big(\mathbf{k}\cdot\mathbf{R}_j\big), \qquad S(\mathbf{k}) = \sum_j q_j\sin\big(\mathbf{k}\cdot\mathbf{R}_j\big), \tag{34}$$

the long-range potential is

$$\Phi^{(\text{recip})}(\mathbf{R_i}) = \sum_{\mathbf{k}\neq 0} W(\mathbf{k})\Big[\cos(\mathbf{k}\cdot\mathbf{R_i})\,C(\mathbf{k}) + \sin(\mathbf{k}\cdot\mathbf{R_i})\,S(\mathbf{k})\Big]. \tag{35}$$

A simple self-term

$$\Phi^{(\text{self})} = -\frac{\alpha}{\sqrt{\pi}}\sum_j q_j \tag{36}$$

removes each charge's interaction with its own screening cloud.

The total potential is then

$$\Phi(\mathbf{R_i}) = \Phi^{(\text{real})}(\mathbf{R_i}) + \Phi^{(\text{recip})}(\mathbf{R_i}) + \Phi^{(\text{self})}. \tag{37}$$

The total potential at each is then embedded with a radial basis function in a similar way as previously done with the multipole expansion except this time there is only a scalar value and the message becomes,

$$\mathbf{m}_{m,i}^{\ell,(t+1)} = \mathbf{m}_{m,i}^{\ell,(t)} + \mathbf{h}_{\ell,m}^{i,\text{QM}} Rbf(\Phi(\mathbf{R_i})) \tag{38}$$

where Rbf is a radial basis function mapping the scalar value to the correct channel size. This tensor of scalar values is then multiplied with the node features tensor along the channel dimension.

F.8. Coulomb Potential Implementation

As a baseline for long-range interactions, we compute the Coulomb potential of a set of point charges q_j at positions $\mathbf{R}_j$. At any QM point $\mathbf{R}_i$, the coulomb potential induced by a cloud of point charges is simply the sum of pairwise terms as shown:

$$\Phi(\mathbf{R}_i) = \sum_j \frac{q_j}{\left\|\mathbf{R}_i - \mathbf{R}_j\right\|} \tag{39}$$

No cutoff or Fourier space decomposition, every charge contributes directly, and the resulting scalar values are collected for each QM point. Similar to the ewald sum previously, we embed this scalar value via a radial basis function and construct the same message as shown in equation 38.