Incorporating Long-Range Interactions via the Multipole Expansion into Ground and Excited-State Molecular Simulations

Rhyan Barrett

Leipzig University, Wilhelm Ostwald Institute for Physical and Theoretical Chemistry, Linnéstraße 2, 04103 Leipzig, Germany

Johannes C. B. Dietschreit

Institute of Theoretical Chemistry, Faculty of Chemistry, University of Vienna, Währinger Straße 17, 1090 Vienna, Austria

Julia Westermayr*

Leipzig University, Wilhelm Ostwald Institute for Physical and Theoretical Chemistry,

Linnéstraße 2, 04103 Leipzig, Germany and

Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI), Dresden/Leipzig, Germany

(Dated: March 3, 2025)

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 introduce FieldMACE, an extension of the message-passing atomic cluster expansion (MACE) architecture that integrates the multipole expansion to model long-range interactions more efficiently. By incorporating the multipole expansion, Field-MACE effectively captures environmental and long-range effects in both ground and excited states. Benchmark evaluations demonstrate its superior performance in predictions and computational efficiency compared to previous architectures, as well as its ability to accurately simulate nonadiabatic excited-state dynamics. Furthermore, transfer learning from foundational models enhances data efficiency, making FieldMACE a scalable, robust, and transferable framework for large-scale molecular simulations.

I. INTRODUCTION

Over the past few years, graph neural networks (GNNs) have shown great promise in modeling molecules and materials due to the natural graph-like structure of chemical systems [1–4]. However, capturing long-range interactions, particularly electrostatics, remains a key challenge [5]. These interactions are crucial when considering solvent effects [6, 7] or for phenomena such as protein

^{*} julia.westermayr@uni-leipzig.de

folding [8, 9]. However, their slow decay with distance makes them difficult to represent accurately [10, 11].

Many frequently applied GNN models focus on interactions among immediate neighbors [12, 13]. As a result, purely local descriptors often miss subtle, but important long-range forces, leading to inaccuracies when simulating systems like proteins or hybrid organic-inorganic interfaces [14, 15] that are often dominated by these effects. Various strategies have been proposed to address this limitation [16, 17], among them, message-passing methods, which exchange information among neighboring nodes, but their limited depth restricts the effective receptive field, making it difficult to account for distant atoms. Other approaches use additional physics-based terms [18, 19] or incorporate data from large, extended systems [16, 20] such as some of the systems in the OC20 [21] dataset, which describes molecules on surfaces, although these can be prohibitively expensive to generate.

A more efficient alternative takes inspiration from QM/MM (quantum mechanics/molecular mechanics) [22, 23] methodologies. In these methods, the region undergoing critical chemical transformations is modeled quantum mechanically, while the surrounding environment is treated classically. An illustration of this approach is shown in the top right of Fig. 1. This reduces the computational cost for large systems without sacrificing essential chemical accuracy.

Traditional machine learning (ML) models that include long-range effects often assign node features to every atom in the system and perform periodic updates thereof, even if many atoms are only characterized by a simple charge [24–26]. This leads to excessive computational overhead and poor scaling. To address this issue, some approaches incorporate physics-based features from the classically treated region into the quantum region. For example, FieldSchNet [27] includes the electric field generated by the partial charges of the MM atoms as an extra feature for the ML model that describes the QM region. This significantly reduces the computational time for training compared to methods like SO3krates [24], since only a subset of atoms requires full periodically updated node features, while the rest can be captured through external interactions. While this strategy efficiently represents some long-range effects without sacrificing computational speed, it relies on a single scalar value (the electric field), which may not fully capture the complex nature of the interactions between MM and QM region.

One way to expand existing ML methods within the QM/MM methodology is via the introduction of higher-order features through the multipole expansion [28, 29]. The multipole expansion offers a systematic way to decompose electrostatic interactions into hierarchical terms, which naturally fits with equivariant neural networks, especially those based on spherical harmonics. In this

work, we incorporate these features into the message-passing atomic cluster expansion (MACE) [30] architecture, resulting in FieldMACE, and demonstrate how this approach can be applied to both ground- and excited-state simulations. Additionally, transfer learning from the foundational MACE-OFF [31] model originally trained on ground-state data without the inclusion of an environment is shown to significantly reduce the amount of data needed [32, 33] for accurate predictions. Most notably, we show that this refinement can replicate reference population curves obtained from nonadiabatic excited-state dynamics [34, 35]. These simulations are, when conducted with quantum chemistry, computationally intensive, often requiring months of computation[36], and still pose challenges to ML models [37–39]. In this work, as little as 30 excited-state data points are required when using transfer learning for reproducing excited-state nonadiabatic molecular dynamics.

II. PRELIMINARIES AND RELATED WORK

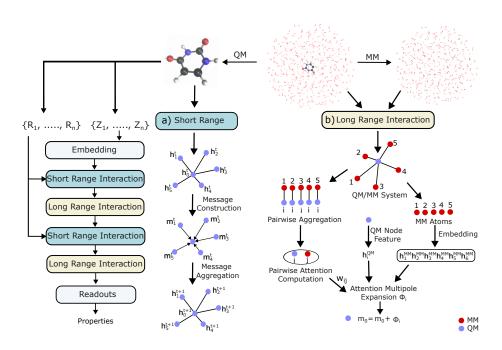


FIG. 1. Architecture of the message-passing neural network integrated with the multipole expansion a) The short-range blocks operate through a message-passing scheme aggregating information in the atoms local neighborhood, whereas the long-range blocks collect information on molecular mechanics (MM) nodes through the multipole expansion. b) The long-range block uses pairwise attention weightings that are computed between the quantum mechanics (QM) region and MM atoms. These are combined with MM node features and the QM node features in the multipole expansion described in Section III.

A. 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 graphs, where atoms correspond to nodes \mathcal{V}_i , and all nodes in the neighborhood of central atom i, $\mathcal{N}(i) = \{\mathcal{V}_j \mid ||\mathbf{r}_{ij}||_2 = ||\mathbf{r}_i - \mathbf{r}_j||_2 < r_{\text{cut}}\}$ are connected by edges, where r_{cut} is a predefined cutoff radius. The categorical atomic numbers are transformed into a learnable invariant embedding vector $X_i \in \mathbb{R}^k$. During the message-passing process [40, 41] each atom i is associated with a set of features $\mathbf{h}_i \in \mathbb{R}^k$, and each edge (i,j) is assigned the vector between the two atoms \mathbf{r}_{ij} . The features are initialized with the atomic embedding $\mathbf{h}_i^0 = X_i$ and then updated through an iterative process:

$$\mathbf{m}_{ij}^{(t)} = \phi_{\text{msg}}\left(\mathbf{h}_i^{(t)}, \mathbf{h}_j^{(t)}, \mathbf{r}_{ij}\right) \tag{1}$$

$$\mathbf{m}_{i}^{(t)} = \sum_{j \in \mathcal{N}(i)} \mathbf{m}_{ij}^{(t)} \tag{2}$$

$$\mathbf{h}_{i}^{(t+1)} = \phi_{\text{update}} \left(\mathbf{h}_{i}^{(t)}, \mathbf{m}_{i}^{(t)} \right)$$
(3)

This process involves three main steps: i) creation of the messages $\mathbf{m}_{ij}^{(t)}$ that are computed for each edge using a learnable function ϕ_{msg} and the vector \mathbf{r}_{ij} . ii) These messages are aggregated over the neighborhood of each node using a permutation-invariant operation, such as summation. iii) Finally, node features \mathbf{h}_i are updated using another learnable function ϕ_{update} . Messages passed between atoms in the neighborhood $\mathcal{N}(i)$ can represent scalar or higher-order features depending on the approach chosen. An illustration of the message-passing process can be seen in Fig. 1a (short-range block).

B. Equivariance

Vectorial properties of a molecular system are uniquely defined by the orientation of it in space, i.e. if the entire system rotates, associated properties have to transform accordingly, which is known as equivariance [42]. Formally the set of features \mathbf{h}_i is called equivariant under a group G if

$$D(R)\mathbf{h}_i(\mathbf{r}_1,...,\mathbf{r}_n) = \mathbf{h}_i(R \circ (\mathbf{r}_1,...,\mathbf{r}_n)) , \qquad (4)$$

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), these equivariant features can be broken down into smaller building blocks [43], the spherical harmonic functions, $Y_{\ell m}(\theta, \phi)$ acting on the sphere, where ℓ represents the degree and m corresponds to the order, $m \in \{-\ell, ..., \ell\}$. Under a rotation $R \in SO(3)$, the spherical harmonics transform as follows:

$$Y_{\ell m'}(\theta', \phi') = \sum_{m} D_{m'm}^{\ell}(R) Y_{\ell m}(\theta, \phi), \tag{5}$$

with $D^{\ell}_{m'm}(R)$ being the Wigner-D matrices. In most cases, the equivariant feature vectors are expressed as a series of vectors where each is associated to a certain spherical harmonic, known as channels. The equivariant features are indexed as $\mathbf{h}^{\ell}_{i,m,k}$ where k is the channel index. The message-passing framework can be rewritten in terms of these spherical harmonics. In this case only two body features will be considered, describing only interactions between neighbors around atom i.

$$\mathbf{m}_{M,i,j,k}^{L,(t)} = R_{\ell_1 m_1}(r_{ij}) Y_{\ell_1 m_1}(\theta,\phi) \mathbf{h}_{j,m_2,k}^{\ell_2,(t)}$$
(6)

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

$$\tag{7}$$

$$\mathbf{h}_{i,m,k}^{\ell,(t+1)} = w_{mkk}^{l} \mathbf{m}_{m,i,k}^{\ell,(t)} + \mathbf{h}_{i,m,k}^{\ell,(t)}$$

$$L = (\ell_1, \ell_2), \quad M = (m_1, m_2)$$
(8)

The angles θ and ϕ correspond to the directionality of the unit vector of $\mathbf{r_{ij}}$ and r_{ij} the distance between atoms i and j. The values $w_{mkk'}^l$ correspond to learnable, linear transformations across the channel of each spherical harmonic, an additional non-linear function can be placed along with the linear transformation. In order to preserve the correct equivariance for each order of spherical harmonics we must include the standard Clebsch Gordan coefficients $C_{L,M}^{lm}$ [44]. The equations above form the basic block to the short range interactions in the following section. These blocks can also be exchanged with MACE which is discussed in the Appendix A.

C. Multipole Expansion

The multipole expansion is a technique that 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. These moments are constructed in terms of spherical harmonics allowing a direct integration into equivariant neural networks that already use spherical harmonics such as MACE.

A scalar potential $\Phi(\mathbf{r})$ arising from 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

$$\Phi(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0} \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}'.$$
 (9)

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

$$\frac{1}{|\mathbf{r} - \mathbf{r}'|} = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \frac{4\pi}{2\ell + 1} \left(\frac{r}{r'}\right)^{\ell} Y_{\ell m}(\theta, \phi) Y_{\ell m}^*(\theta', \phi') \tag{10}$$

Substituting this expansion into Coulomb's law yields:

$$\Phi(\mathbf{r}) = \frac{1}{\varepsilon_0} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \frac{1}{2\ell+1} \frac{Y_{\ell m}(\theta, \phi)}{r^{\ell+1}} Q_{\ell m}, \tag{11}$$

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

$$Q_{\ell m} = \int r'^{\ell} \rho(\mathbf{r}') Y_{\ell m}^*(\theta', \phi') \, d\mathbf{r}'. \tag{12}$$

Each term in this expansion corresponds to a specific multipole order, with higher-order terms decaying more rapidly with distance. These multipole moments will be used as feature vectors in the following section.

III. INTEGRATION OF QM/MM FRAMEWORK WITH EQUIVARIANT NEURAL NETWORKS

Our method splits a chemical system into a QM and an MM region. The descriptors for the QM region are derived in Sections 2.1 and 2.2. The MM atoms are described by their positions and partial charges that are assigned by a classical force field. While this is sufficient for the construction of the multipole expansion, the direct summation leads to significant information loss. To address this, we introduce weightings towards each of the MM atoms. Concretely, we compute pairwise attention coefficients [45] between the QM node features and the MM atom features to establish appropriate weightings.

A. Attention Weightings

Initial feature vectors for the MM atoms comprise spherical harmonics values, $Y_{\ell m}^{\rm MM}$, computed from the position values. These values are expanded to produce MM node features,

$$\mathbf{h}_{j,m,k'}^{\ell,\mathrm{MM}} = w_{mk'}^{\ell} Y_{j,\ell m}^{\mathrm{MM}} \tag{13}$$

where $w^l_{mk'}$ is a mapping $\mathbb{R} \to \mathbb{R}^{k'}$. To reduce computational costs, k' can be tuned as a hyperparameter and remains k'< k. A similar contraction operation is performed for the QM node features:

$$\mathbf{h}_{i,m,k'}^{\ell,\mathrm{QM},(t)} = w_{mk'k}^{\ell} \mathbf{h}_{i,m,k}^{\ell,\mathrm{QM},(t)},\tag{14}$$

where $w_{mk'k}^l$ is a mapping $\mathbb{R}^k \to \mathbb{R}^{k'}$. Given these feature vectors, pairwise attention is taken between QM atoms and each of the MM atoms. Note that no attention is computed between MM atoms. This significantly reduces the computational cost for large systems, full element-wise comparison scales with $O(n^2)$ whereas element-wise comparison between only MM and QM scales with O(n) with the QM region being assumed to be constant. The element-wise comparisons are performed using an L_1 norm

$$\kappa_{ijk'}^{(t)} = \xi_{mk'k'}^{\ell} (|\mathbf{h}_{i,m,k'}^{\ell,QM,(t)} - \mathbf{h}_{j,m,k'}^{\ell,MM}|)$$
(15)

Here ξ consists of a linear transformation followed by a softmax normalization. These weighting values can be integrated into the multipole expansion to obtain the attention weighted expansion. An illustration of computing the attention weights can be seen beneath the long-range interaction block in Figure 1.

B. Attention Multipole Expansion

The charge distribution surrounding the QM region can be described by a set of point charges obtained from the MM atoms, q_i^{MM} ;

$$\rho(\mathbf{r}') = \begin{cases} q_j^{\text{MM}}, & \text{if } \mathbf{r}' = \mathbf{r}_j^{\text{MM}}, \\ 0, & \text{otherwise.} \end{cases}$$
(16)

This simplifies the multipole moments, eq. (12), into a sum over these discrete charges.

$$Q_{m,k'}^{\ell} = \sum_{j} (r_j^{\text{MM}})^l q_j^{\text{MM}} \mathbf{h}_{j,m,k'}^{\ell,\text{MM}} .$$
 (17)

It is important to note that the system is shifted so that the QM region is centered on the origin to maintain translational invariance. The multipole expansion can now be written including the weighting values for atom i in the QM region,

$$\Phi_{i,k'}^{\ell m,(t)}(\mathbf{r}_i^{\mathrm{QM}}) = \frac{1}{\varepsilon_0} \sum_{i} \kappa_{ijk'}^{(t)} \frac{Q_{m,j,k'}^{\ell}}{2\ell + 1} \frac{\mathbf{h}_{i,m,k'}^{\ell,\mathrm{QM},(t)}}{|\mathbf{r}_i^{\mathrm{QM}} - \mathbf{r}_i^{\mathrm{MM}}|^{\ell + 1}}$$
(18)

$$Q_{m,j,k'}^{\ell} = (r_j^{\text{MM}})^l q_j^{\text{MM}} \mathbf{h}_{j,m,k'}^{\ell,\text{MM}}$$

$$\tag{19}$$

The term $\Phi_{i,k'}^{\ell m,(t)}(\mathbf{r}_i^{\mathrm{QM}})$ incorporates the necessary long-range information from the MM atoms for node feature $\mathbf{h}_i^{\mathrm{QM},(t)}$ at atom i.

C. Message Construction

The message is updated again after the short range message update eq. 7 by incorporating the multipole expansion term after expanding it to the correct dimensionality:

$$\mathbf{m}_{m,i,k}^{\ell,(t)} = \mathbf{m}_{m,i,k}^{\ell,(t)} + w_{mkk'}^{\ell,(t)} \Phi_{i,k'}^{\ell m,(t)} (\mathbf{r}_i^{QM})$$
(20)

The node features are constructed from these messages in the same way as seen in eq. (8). These node features can be expanded to include many body interactions as well as other effects in the QM region. An illustration of the long range interaction implementation can be seen in the right half of Figure 1. Our long-range multipole message is integrated into the MACE framework.

IV. EXPERIMENTS

TABLE I. Mean Absolute Error (MAE) for Energy and Forces of FieldSchNet and FieldMACE models evaluated on the ground states of benzene, uracil, and retinoic acid systems. Results are shown for varying representation sizes (8, 16, 32, 64, 128) to assess model performance.

Model	System	Energy (eV)					Force (eV/Å)				
	by Stelli	Representation Size					Representation Size				
		8	16	32	64	128	8	16	32	64	128
FieldSchNet	Benzene	0.065	0.058	0.049	0.014	0.005	0.054	0.053	0.047	0.034	0.017
	Uracil	0.182	0.151	0.079	0.016	0.015	0.128	0.118	0.112	0.054	0.048
	Retinoic acid	0.145	0.053	0.029	0.028	0.015	0.190	0.141	0.099	0.064	0.031
FieldMACE	Benzene	0.041	0.033	0.018	0.006	0.003	0.039	0.038	0.031	0.018	0.010
	Uracil	0.060	0.035	0.012	0.006	0.005	0.082	0.066	0.040	0.026	0.020
	retinoic acid	0.061	0.036	0.019	0.013	0.012	0.117	0.092	0.065	0.043	0.028

In this section, we evaluate the performance of our approach, called FieldMACE, which constitutes the inclusion of long-range message-passing into the MACE architecture. To this end, we test our model on the datasets by [46] comprising three molecular systems, namely benzene, uracil, and retinoic acid modeled under explicit environment. Each data set comprises molecules solvated

in water, full details of the dataset can be found in the Appendix B1 or respective publication. We showcase the performance of our method using different representation sizes and compare it to another state-of-the-art model (FieldSchNet) that can be used for ML/MM, where ML is used to approximate the QM part of the system. In addition, we compute the vibrational power spectrum with FieldMACE for benzene solvated in water and compare the spectrum to a quantum chemical reference spectrum. In addition, we gauge FieldMACE's capacity to handle electronically excited states by computing population curves resulting from nonadiabatic photodynamics simulations. Finally, we explore the potential for reusing parameters from foundational ground-state models trained without an external environment, particularly the MACE-OFF model, which allows us to obtain lower errors with less amount of data compared to directly trained models.

A. Ground State Simulations

To evaluate the performance of FieldMACE in predicting energies and forces in the presence of an environment, we benchmark it against FieldSchNet [39], a recently developed model that can be used for QM/MM simulations. Full architecture and training details can be found in the Appendix C. To the best of our knowledge, there are currently few ML approaches capable of effectively describing molecular systems [39, 47–57] in the presence of implicit or explicit environments. Many other models that are used for ML/MM or similar simulations often do not take the influence of an environment into account, by relying for instance on so-called mechanical embedding approaches (neglecting any electrostatic interactions), or by using specific embedding schemes, like the Buffered Region Neural Network approach (BuRNN), [58] where the whole system is split into three regions rather than two and the ML model learns the difference between regions rather than the QM region. Thus, to allow for a fair comparison, the benchmark focuses exclusively on the FieldSchNet and FieldMACE models. The results, presented in Table I summarize the mean absolute errors (MAEs) for energy and force predictions. As can be seen, FieldMACE consistently outperforms FieldSchNet, particularly at smaller representation sizes, demonstrating its capability to capture both short- and long-range interactions while being more compact. For energy predictions, both models show reduced errors with increasing representation size. Across all systems, FieldMACE consistently exhibits lower errors for both energy and forces.

In addition to evaluating test scores, we simulate the photodynamics of solvated benzene and compare the associated power spectra of the ML/MM and QM/MM reference dynamics in Figure 2a, computational details can be found in Appendix B.2. The trained FieldMACE method produces a

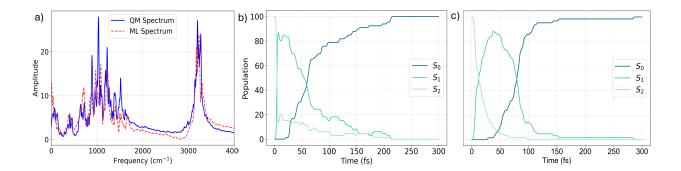


FIG. 2. a) Power spectrum produced from the dynamics of benzene in a water box. Population curves resulting from photodynamics simulations using b) FieldMACE and c) quantum chemistry, showing the ring-opening reaction of furan after being excited to the second excited state (S_2) . Transitions to S_1 and S_0 are present in both approaches.

power spectrum in agreement with the QM spectrum, particularly, it reproduces the characteristic C–H stretching modes located near 3000 cm⁻¹ as well as the C=C ring stretching and C-H bending modes contained in the distributions centered around 1000 cm⁻¹. In general, the peaks of the ML spectrum are located at the same frequencies as in the QM reference, only their relative intensity varies slightly.

B. Excited State Dynamics

Going beyond ground state simulations, FieldMACE is used to simulate nonadiabatic dynamics using trajectory surface hopping (TSH) [59, 60]. These 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, as is the case here. 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 [59, 61], used here, each trajectory changes its active state via so-called hops stochastically. For the photodynamics simulations, we integrated our FieldMACE approach into the SHARC (Surface Hopping including Arbitrary Couplings) program suite [34, 62–64].

We investigate the molecule furan solvated in water to assess Field-MACE for excited-state dynamics. Details on the model and training are specified in Appendix D. Upon excitation, the molecule can undergo a ring-opening reaction. We compare our model to recent simulations of this system using FieldSchNarc [39].

Figure 2b and c show the population dynamics obtained using FieldMACE (panel b) and the

quantum chemical reference method (panel c). Population curves represent a distribution over the active state showing how the system evolves over time and hops between different electronic states. At the beginning of the simulations, the system is excited to the second excited singlet state, S_2 . The nonadiabatic dynamics then show non-radiative decay back to the electronic ground state, S_0 . FieldMACE accurately captures the rapid decay from the S_2 and the subsequent redistribution of populations to the S_1 and the ground state, S_0 . The predicted population transfer closely matches quantum chemical reference populations. As can be seen, in both cases the S_2 population transfers quickly to the S_1 state, which is populated until around 100 fs, at which point the majority of the trajectories fall back into the ground state.

V. TRANSFERABILITY

A. Ground State

In this section, we analyze the possibility of reducing the amount of data necessary when starting with weights from pre-trained foundational ground-state models that have not been trained including an environment or excited-state effects. Therefore, we leverage the foundational MACE model's pre-trained representations as initial parameters for FieldMACE's short-range modules to improve performance and data efficiency. Full details on the models and training can be found in Appendix B.

As can be seen in Figure 3, which shows learning curves using different amounts of data for fine-tuning for three molecular systems in solution, initializing the short-range blocks of FieldMACE with the foundational model weights reduces the amount of training data required to achieve comparable accuracy to models trained from scratch, particularly for force predictions (second line, panels d-f). Across all systems, the pre-trained initialization demonstrates a clear advantage, leading to improved accuracy compared to random initialization, even though the foundational models do not incorporate any solvent effects.

The benefits of using the foundational representation are most pronounced in systems with few strongly electronegative atoms, as is the case in benzene (first column, panels a and d) or retinoic acid (last column, panels c and f). These molecules are characterized by weaker intermolecular interactions and limited hydrogen bonding with the surrounding water molecules. The resulting potential energy surfaces for these systems will thus be more similar to the potential energy surfaces in a vacuum compared to the uracil system.

In contrast, for polar systems like uracil (middle column, panels b and e), which feature strong hydrogen bonding and complex electrostatic interactions, the benefits of the pre-trained initialization are less pronounced. While transfer-learned FieldMACE still outperforms the randomly initialized model, the benefit is smaller compared to the other examples. This suggests that the foundational model's learned parameters have to be changed more in systems that interact more strongly with the MM environment. Here, additional refinement on another dataset or task-specific training would improve performance for highly polar systems.

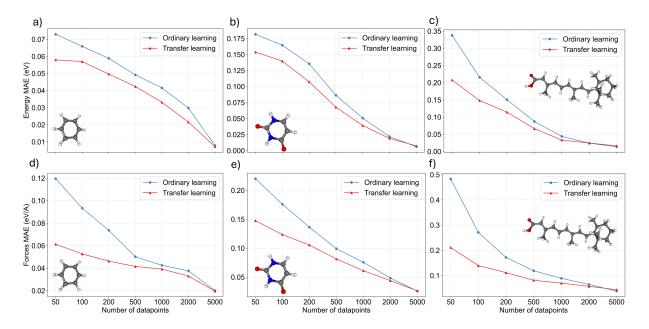


FIG. 3. Comparison of transfer learning from MACE-OFF foundational model (red curves) against randomly initialized FieldMACE models (blue curves). Evaluation of transfer and ordinary learning models with respect to energies (top row) and forces (bottom row) of molecules solvated in water: a) and d) benzene, b) and e) uracil and c) and f) retinoic acid.

B. Excited States

To further test the ability of FieldMACE to profit from transfer learning by heavily reducing the amount of training data needed, we revisit the furan in water system. Full details on the model and training can be found in Appendix D. Rather than considering the mere test statistics (meaning error for energies and forces on the test set), we generate the population curves for the same set of initial conditions used in the previous section with FieldMACE models trained with a decreasing number of data. We compare FieldMACE models trained from scratch with randomly

initialized parameters and those where the short-range block parameters are taken from a MACE-OFF model. We use only 600, 150, and 30 data points to train these models. Since excited-state data are computationally much more costly to obtain, this is a realistic use case. The results are shown in Figure 4.

For the largest dataset with 600 data points (first column), both transfer (panel a) and non-transfer models (panel d) result in similar population curves that align well with the reference results (Figure 2 c). Deviations in the S_2 , S_1 , and S_0 populations remain minimal throughout the simulation compared to the reference population curves. When the dataset size is reduced to 150 points (middle column), 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 (last column), 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 and less stable simulations, capturing the general trends in decay and redistribution among electronic states. However, the model trained from scratch (panel f) fails to accurately capture the transition dynamics or the populations of S_2 , S_1 , and S_0 . These results highlight the critical importance of transfer learning in low-data regimes, where the foundational representation enables FieldMACE to generalize effectively despite very limited training data.

Overall, the results demonstrate that transfer learning significantly enhances FieldMACE's performance, particularly when 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 maintaining reasonable predictions albeit less stable simulations, but the non-transfer model fails to capture the dynamics adequately. These findings emphasize the use of pre-trained foundational representations for obtaining accurate models.

VI. DISCUSSION

The attention weighted multipole expansion presented in this paper highlights an effective way of combining spherical harmomics-based equivariant neural networks with external charge distributions. In particular, results presented in this study highlight the significant advancements achieved by combining the multipole expansion with the MACE neural network architecture, FieldMACE, particularly its transferability and efficiency in modeling both ground-state and excited-state molec-

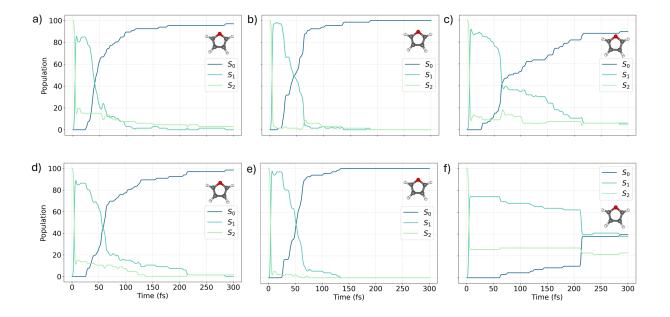


FIG. 4. Population curves of furan starting in the third excited singlet state for transfer learning models using a) 600, b) 150, and c) 30 data points and non-transfer learning models using d) 600, e) 150, and f) 30 data points.

ular processes. By leveraging the foundational MACE model representations, FieldMACE demonstrates its ability to extrapolate to new molecular systems while requiring less training data, especially for force predictions. This capability underscores the potential of pre-trained machine learning models in quantum chemistry.

One of the most notable findings is the effectiveness of FieldMACE in low-data regimes. For excited-state dynamics, the transfer learning approach from ground-state foundational models enabled FieldMACE to produce reasonable population dynamics even with as little as 30 data points, whereas models initialized with random weights failed to capture the underlying processes. This is particularly impactful given the computational expense associated with generating training data for excited-state simulations. The ability to maintain predictive accuracy with reduced datasets makes FieldMACE a practical choice for studying complex systems where data availability is limited.

When comparing FieldMACE with other molecular architectures that incorporate long-range interactions such as SO3Krates, the balance between accuracy and computational expense must be considered. In these architectures pairwise attention weightings are computed for all pairs in the system which leads to quadratic scaling, $O(n^2)$. In some cases considering interactions between all atoms in the system may be necessary but in many cases it is not. For instance, in the presented uracil system, which contains more than 3000 atoms, the only area of high significance is the uracil

molecule. Training a single epoch while considering all pairwise interactions and assigning large node features can lead to prohibitively large computational costs to train these models. FieldMACE significantly reduces this computational burden by only considering interactions between the QM and MM atoms, thus allowing linear scaling when increasing the size of the external environment. This leads to a dramatic reduction in computational overhead, and thus allows us to train on extended systems.

In summary, FieldMACE represents a significant step forward in transferable machine learning models for molecular simulations in the presence of external charge distributions such as QM/MM simulations. Its ability to train on very large systems, combined with its accuracy in low-data regimes, makes it a powerful tool for QM/MM simulations in both ground and excited states. Future work will focus on enhancing its performance for highly polar systems, expanding its application to broader chemical spaces, and applying FieldMACE to systems such as proteins where long-range interactions are particularly dominant.

ACKNOWLEDGEMENTS

This work is funded in parts by the Deutsche Forschungsgemeinschaft (DFG) – Project-ID 443871192 - GRK 2721: "Hydrogen Isotopes ^{1,2,3}H". The authors acknowledge the ZIH TU Dresden, the URZ Leipzig University, and Paderborn Center for Parallel Computing (PC2) for providing the computational resources to conduct this study.

AUTHOR CONTRIBUTIONS

RB (Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Validation, Writing - original draft). JCBD (Methodology, Writing - review). JW (Methodology, Conceptualization, Supervision, Writing - review).

MATERIALS AND CORRESPONDENCE

Correspondence to Julia Westermayr.

COMPETING INTERESTS

The authors declare no competing interests.

CODE AND DATA AVAILABILITY

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.

 H. E. Sauceda, L. E. Gálvez-González, S. Chmiela, L. O. Paz-Borbón, K.-R. Müller and A. Tkatchenko, Nat. Commun., 2022, 13, 3733.

- [3] X. Gao, F. Ramezanghorbani, O. Isayev, J. S. Smith and A. E. Roitberg, J. Chem. Inf. Model., 2020, 60, 3408–3415.
- [4] T. W. Ko, J. A. Finkler, S. Goedecker and J. Behler, Nat. Commun., 2021, 12, 398.
- [5] D. M. Anstine and O. Isayev, J. Phys. Chem. A, 2023, 127, 2417–2431.
- [6] C. Reichardt, Org. Process Res. Dev., 2007, 11, 105–113.
- [7] C. Reichardt and T. Welton, Solvents and Solvent Effects in Organic Chemistry, John Wiley & Sons, 4th edn., 2011.
- [8] N. Go and H. Taketomi, Proc. Natl. Acad. Sci., 1978, 75, 559–563.
- [9] C. Sagui and T. A. Darden, Annu. Rev. Biophys. Biomol. Struct., 1999, 28, 155-179.
- [10] A. Ambrosetti, N. Ferri, R. A. DiStasio and A. Tkatchenko, J. Chem. Phys., 2014, 140, 18A508.
- [11] K. Kang et al., arXiv preprint, 2024.
- [12] S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt and B. Kozinsky, Nat. Commun., 2022, 13, 2453.
- [13] K. Schütt, O. Unke and M. Gastegger, Proc. Int. Conf. Mach. Learn., 2021, pp. 9377–9388.
- [14] M. M. Gromiha and S. Selvaraj, *Biophys. Chem.*, 1999, 77, 49–68.
- [15] J. Westermayr, S. Chaudhuri, A. Jeindl, O. T. Hofmann and R. J. Maurer, Digit. Discov., 2022, 1, 463–475.
- [16] A. Kosmala, J. Gasteiger, N. Gao and S. Günnemann, Proc. Int. Conf. Mach. Learn., 2023, pp. 17544–17563.
- [17] P. Loche, K. K. Huguenin-Dumittan, M. Honarmand, Q. Xu, E. Rumiantsev, W. B. How, M. F. Langer and M. Ceriotti, arXiv preprint arXiv:2412.03281, 2024.
- [18] D. M. Anstine and O. Isayev, J. Phys. Chem. A, 2023, 127, 2417–2431.
- [19] A. Gao and R. C. Remsing, Nat. Commun., 2022, 13, 1572.

^[2] B. Deng, P. Zhong, K. Jun, J. Riebesell, K. Han, C. J. Bartel and G. Ceder, Nat. Mach. Intell., 2023, 5, 1031–1041.

- [20] J. Behler and G. Csányi, Eur. Phys. J. B, 2021, 94, 1–11.
- [21] L. Chanussot, A. Das, S. Goyal, T. Lavril, M. Shuaibi, M. Riviere, K. Tran, J. Heras-Domingo, C. Ho, W. Hu et al., ACS Catal., 2021, 11, 6059–6072.
- [22] C. E. Tzeliou, M. A. Mermigki and D. Tzeli, *Molecules*, 2022, 27, 2660.
- [23] J. M. Boereboom, P. Fleurat-Lessard and R. E. Bulo, J. Chem. Theory Comput., 2018, 14, 1841–1852.
- [24] T. Frank, O. Unke and K.-R. Müller, Adv. Neural Inf. Process. Syst., 2022, 35, 29400–29413.
- [25] I. Batatia, L. L. Schaaf, H. Chen, G. Csányi, C. Ortner and F. A. Faber, arXiv preprint arXiv:2310.10434, 2023.
- [26] J. T. Frank et al., arXiv preprint, 2024.
- [27] M. Gastegger, K. T. Schütt and K.-R. Müller, Chem. Sci., 2021, 12, 11473-11483.
- [28] A. R. Edmonds, Angular momentum in quantum mechanics, Princeton university press, 1996, vol. 4.
- [29] K. S. Thorne, Rev. Mod. Phys., 1980, **52**, 299.
- [30] I. Batatia, D. P. Kovacs, G. Simm, C. Ortner and G. Csányi, Adv. Neural Inf. Process. Syst., 2022, 35, 11423–11436.
- [31] D. P. Kovács, J. H. Moore, N. J. Browning, I. Batatia, J. T. Horton, V. Kapil, W. C. Witt, I.-B. Magdău, D. J. Cole and G. Csányi, arXiv preprint arXiv:2312.15211, 2023.
- [32] C. Cai, S. Wang, Y. Xu, W. Zhang, K. Tang, Q. Ouyang, L. Lai and J. Pei, J. Med. Chem., 2020, 63, 8683–8694.
- [33] D. Buterez, J. P. Janet, S. J. Kiddle, D. Oglic and P. Lió, Nat. Commun., 2024, 15, 1517.
- [34] S. Mai, P. Marquetand and L. González, Wiley Interdiscip. Rev. Comput. Mol. Sci., 2018, 8, e1370.
- [35] T. R. Nelson, A. J. White, J. A. Bjorgaard, A. E. Sifain, Y. Zhang, B. T. Nebgen, S. Fernandez-Alberti, D. Mozyrsky, A. E. Roitberg and S. Tretiak, Chem. Rev., 2020, 120, 2215–2287.
- [36] J. Westermayr, S. Ghosh, T. Mori and L. González, Nat. Chem., 2022, 14, 914–919.
- [37] J. Westermayr and P. Marquetand, Chem. Rev., 2021, 121, 9873–9926.
- [38] S. Mausenberger, C. Müller, A. Tkatchenko, P. Marquetand, L. González and J. Westermayr, Chem. Sci., 2024, 15, 15880–15890.
- [39] M. X. Tiefenbacher, B. Bachmair, C. G. Chen, J. Westermayr, P. Marquetand, J. C. Dietschreit and L. González, Excited-state nonadiabatic dynamics in explicit solvent using machine learned interatomic potentials, 2025.
- [40] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals and G. E. Dahl, Int. Conf. Mach. Learn., 2017, pp. 1263–1272.
- [41] P. B. Jørgensen, K. W. Jacobsen and M. N. Schmidt, arXiv preprint arXiv:1806.03146, 2018.
- [42] C. Esteves, arXiv preprint arXiv:2004.05154, 2020.
- [43] N. Thomas, T. Smidt, S. Kearnes, L. Yang, L. Li, K. Kohlhoff and P. Riley, arXiv preprint arXiv:1802.08219, 2018.
- [44] Y. A. Smorodinskii and L. A. Shelepin, Sov. Phys. Usp., 1972, 15, 1.
- [45] A. Vaswani, Adv. Neural Inf. Process. Syst., 2017.

- [46] L. Boöselt, M. Thürlemann and S. Riniker, J. Chem. Theory Comput., 2021, 17, 2641–2658.
- [47] A. Hofstetter, L. Böselt and S. Riniker, Phys. Chem. Chem. Phys., 2022, 24, 22497–22512.
- [48] G. Song and W. Yang, arXiv preprint, 2025.
- [49] K. Yao, J. E. Herr, D. W. Toth, R. Mckintyre and J. Parkhill, Chem. Sci., 2018, 9, 2261–2269.
- [50] B. Lier, P. Poliak, P. Marquetand, J. Westermayr and C. Oostenbrink, J. Phys. Chem. Lett., 2022, 13, 3812–3818.
- [51] A. Hofstetter, L. Böselt and S. Riniker, Phys. Chem. Chem. Phys., 2022, 24, 22497–22512.
- [52] K. Zinovjev, J. Chem. Theory Comput., 2023, 19, 1888–1897.
- [53] M. Thürlemann and S. Riniker, Chem. Sci., 2023, 14, 12661–12675.
- [54] K. Zinovjev, L. Hedges, R. Montagud Andreu, C. Woods, I. Tuñón and M. W. van der Kamp, J. Chem. Theory Comput., 2024.
- [55] J. A. Semelak, I. Pickering, K. K. Huddleston, J. Olmos, J. S. Grassano, C. Clemente, S. I. Drusin, M. Marti, M. C. G. González, A. E. Roitberg et al., 2024.
- [56] J. S. Grassano, I. Pickering, A. E. Roitberg, M. C. González Lebrero, D. A. Estrin and J. A. Semelak, J. Chem. Inf. Model., 2024, 64, 4047–4058.
- [57] Y.-K. Lei, K. Yagi and Y. Sugita, J. Chem. Phys., 2024, 160, 214109.
- [58] B. Lier, P. Poliak, P. Marquetand, J. Westermayr and C. Oostenbrink, J. Phys. Chem. Lett., 2022, 13, 3812–3818.
- [59] J. C. Tully, J. Chem. Phys., 1990, **92**, 1061–1071.
- [60] J. C. Tully, Int. J. Quantum Chem., 1991, 40, 299–309.
- [61] G. Granucci and M. Persico, J. Chem. Phys., 2007, 126, 134114.
- [62] M. Richter, P. Marquetand, J. González-Vázquez, I. Sola and L. González, J. Chem. Theory Comput., 2011, 7, 1253–1258.
- [63] S. Mai, D. Avagliano, M. Heindl, P. Marquetand, M. F. S. J. Menger, M. Oppel, F. Plasser, S. Polonius, M. Ruckenbauer, Y. Shu, D. G. Truhlar, L. Zhang, P. Zobel and L. González, SHARC3.0: Surface Hopping Including Arbitrary Couplings Program Package for Non-Adiabatic Dynamics, https://sharcmd.org/, 2023.
- [64] S. Mai, P. Marquetand and L. González, Int. J. Quantum Chem., 2015, 115, 1215–1231.
- [65] R. Drautz, Phys. Rev. B, 2019, 99, 014104.
- [66] I. Batatia, S. Batzner, D. P. Kovács, A. Musaelian, G. N. Simm, R. Drautz, C. Ortner, B. Kozinsky and G. Csányi, Nat. Mach. Intell., 2025, 1–12.
- [67] A. D. Becke, Phys. Rev. A, 1988, 38, 3098.
- [68] A. D. Becke, J. Chem. Phys., 1993, 98, 1372–1377.
- [69] S. Grimme, S. Ehrlich and L. Goerigk, J. Comput. Chem., 2011, 32, 1456–1465.
- [70] S. Grimme, J. Chem. Phys., 2006, 124, 034108.
- [71] B. Dunlap, J. Connolly and J. Sabin, J. Chem. Phys., 1979, 71, 4993–4999.
- [72] H. J. Berendsen, J.-R. Grigera and T. P. Straatsma, J. Phys. Chem., 1987, 91, 6269-6271.

- [73] A. K. Malde et al., J. Chem. Theory Comput., 2011, 7, 4026–4037.
- [74] D. J. Evans and B. L. Holian, J. Chem. Phys., 1985, 83, 4069–4074.
- [75] P. Mark and L. Nilsson, J. Phys. Chem. A, 2001, 105, 9954–9960.
- [76] A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus et al., J. Phys. Condens. Matter, 2017, 29, 273002.
- [77] S. Heydari, S. Raniolo, L. Livi and V. Limongelli, Commun. Chem., 2023, 6, 13.
- [78] Y. Shu, L. Zhang, X. Chen, S. Sun, Y. Huang and D. G. Truhlar, J. Chem. Theory Comput., 2022, 18, 1320–1328.

Appendix A: MACE Architecture

To extend message-passing neural networks (MPNNs) with many-body interactions, the atomic cluster expansion (ACE) [65] formulation can be utilized. This expansion incorporates higher-order terms, which are essential for accurately modeling complex molecular interactions. The message $\mathbf{m}_i^{(t)}$ at iteration t is expressed as a sum over many-body contributions:

$$\mathbf{m}_{i}^{(t)} = \sum_{j \in \mathcal{N}_{1}(i)} \phi_{1}\left(\mathbf{h}_{i}^{(t)}, \mathbf{h}_{j}^{(t)}, \mathbf{r}_{ij}\right)$$

$$+ \sum_{j_{1} \in \mathcal{N}_{2}(i)} \sum_{j_{2} \in \mathcal{N}_{2}(j_{1})} \phi_{2}\left(\mathbf{h}_{i}^{(t)}, \mathbf{h}_{j_{1}}^{(t)}, \mathbf{h}_{j_{2}}^{(t)}, \mathbf{r}_{ij_{1}}, \mathbf{r}_{j_{1}j_{2}}\right)$$

$$\vdots$$

$$+ \sum_{j_{1}, \dots, j_{\nu} \in \mathcal{N}_{\nu}(i)} \phi_{\nu}\left(\mathbf{h}_{i}^{(t)}, \dots, \mathbf{h}_{j_{\nu}}^{(t)}, \mathbf{r}_{ij_{1}}, \dots, \mathbf{r}_{j_{\nu-1}j_{\nu}}\right),$$

where $N_{\nu}(i)$ denotes the set of ν -th order neighbors of node i, and ϕ_{ν} represents the learnable interaction function incorporating many-body effects. This operation is computationally expensive but can be significantly reduced by constructing higher-order terms using tensor products of two body features then symmetrizing. The full derivations of the MACE architecture can be seen in the following sources [30, 66].

Appendix B: Ground State Simulation Details

1. Molecular Datasets

The configurations, energies, and forces for the three molecules solvated in water were taken from Boöselt et al. [46]. Simulation details are briefly summarized here: All simulations were performed

using a QM/MM approach with the DFT functionals BP86 [67] for retonic acid, ω B97X-D3 [68, 69] for uracil, and B2-PLYP [70] for benzene; the def2-TZVP basis set was used for all calculations. The resolution of identity (RI) approximation [71], Grimme's dispersion correction with Becke–Johnson damping [69], 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 [72]. The QM/MM Hamiltonian incorporated MM point charges within a cutoff radius of 0.6 nm for benzene and 1.4 nm for uracil and retinoic acid.

Each solute was solvated in a periodic water box, with sizes of 1.6 nm, 3.3 nm and 5.0 nm respectively for benzene, uracil and retonic acid. The number of MM partial charges included was approximately 1500–2000 for uracil and 2500 for retinoic acid, while benzene had a smaller number of MM partial charges of 200 due to the smaller box size. The initial structures were generated using the ATB server [73] and relaxed at 0 K with a gradient descent algorithm until the predicted energy change fell below 0.1 kJ mol⁻¹.

For the molecular dynamics simulations that were performed, a time step of 0.5 fs was used, with a Nose–Hoover thermostat [74] (0.1 ps coupling constant) and a weak-coupling barostat (0.5 ps coupling constant). Long-range electrostatics beyond 1.4 nm were included using a reaction-field method. The system was simulated for 20,000 steps at 400 K and 1 bar, with the first 10,000 steps discarded for equilibration and the remaining 10,000 were saved.

2. Simulation Details for benzene in water

The following parameters were used to generate the power spectrum in Fig. 2a. In the QM/MM simulations, an electrostatic embedding was used. The benzene atoms were treated quantum mechanically, while the surrounding water molecules were treated classically with the TIP3P [75] model. The simulation was done with the atomic simulation environment (ASE) [76]. The simulation cell was defined as a box of 16.0 Å per side with periodic boundary conditions applied in all directions. The QM region, was evaluated at a B3LYP [67]/def-TZVP level. For the ML/MM simulations of the same system, we used the FieldMACE model trained on the benzene dataset. The model used 128 channels with spherical harmonics up to degree 3. The vectors assigned to the MM atoms used spherical harmonics up to degree 3 and with 16 channels.

The initial geometry was optimized using the BFGS [76] minimizer until the maximum residual force was reduced to 0.1 eV/A, this was performed in the atomic simulation environment in ASE [76]. Initial velocities were drawn from a Maxwell-Boltzmann distribution at 300 K. We performed

molecular dynamics with a Langevin integrator using a 1 fs timestep, a friction coefficient of 10 ps⁻¹, and a target temperature of 300 K.The system was propagated for a total of 15,000 steps.

In order to facilitate a fair comparison between the ML and QM methods, the benzene data set from [46] was recomputed at the same level of theory, B3LYP-D3(BJ)/def-TZVP for the QM region with the TIP3P model for the MM region. The ML model was trained on the recomputed frames from this data set (see details for the training in the following Sections). The ML/MM-MD producing the ML-based power spectrum was performed with identical settings as the QM/MM-MD.

Appendix C: Training and Model Details

In all systems, a 5 Ångstrom cutoff was applied to the ML region to construct the local graph. The training parameters shown in the FieldMACE section below were also used for the transfer learning models, with the exception that the channel size was fixed to 128 for the transfer learning [33, 77] and 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.

1. FieldMACE

All trainings were performed on a single NVIDIA H100 GPU. For the three datasets (see Section B.1.) we used all available configurations, which correspond to 10000 frames for benzene and uracil, and 9719 structures for retinoic acid. In all cases, we performed an 80:10:10 train:validation:test split. For all FieldMACE models, we used two sets of short- and long-range layers. We trained models with varying channel sizes of 8, 16, 32, 64, 128.Additionally, all models were trained with spherical harmonics up to degree 3 and 8 Bessel basis functions. The initial vectors for the MM atoms consisted of spherical harmonics up to degree 3, combined with an embedding vector of size 16 (not varied between models).

All models were trained with the loss function:

$$L = c_E \sum_{i=1}^{n} (E_{REF}^{(i)} - \hat{E}^{(i)})^2 + c_F \sum_{i=1}^{n} (\mathbf{F}_{REF}^{(i)} - \hat{\mathbf{F}}^{(i)})^2$$
 (C1)

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 0.001.

2. FieldSchNet

All models used 6 interaction layers with 8-128 channels. A total of 25 Gaussian basis functions were used to featurize each edge. The learning rate was set to 10^{-4} , with a decay factor of 0.8 triggered after 25 training epochs without improvement. Both energy and force terms were assigned a weight of 1.0 in the loss function, in order to get equal emphasis during optimization. Training took place on a single A30 GPU. The training splits used were identical to FieldMACE.

Appendix D: Excited State Simulations

1. Surface Hopping Molecular Dynamics

Trajectory surface hopping (TSH) [59, 60, 62, 64] is a nonadiabatic molecular dynamics approach where for each time step nuclear motion evolves classically on a single potential energy surface (PES), while stochastic hops between electronic states occur based on nonadiabatic coupling probabilities. In Tully's fewest switches surface hopping (FSSH) [59, 60], electronic state populations are propagated via the time-dependent Schrödinger equation,

$$\frac{d}{dt}c_k(t) = -\sum_{\ell} \left[\frac{i}{\hbar} H_{k\ell} + \mathbf{d}_{k\ell} \cdot \mathbf{v} \right] c_{\ell}(t), \tag{D1}$$

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

$$P_{k\to\ell} = \max\left(0, \frac{2\operatorname{Re}\left(c_n^*c_\ell\left[\frac{i}{\hbar}H_{k\ell} + \mathbf{d}_{k\ell} \cdot \mathbf{v}\right]\right)}{|c_k|^2}\Delta t\right). \tag{D2}$$

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 $P_{k\to\ell}$ [59]. Since calculating explicit nonadiabatic couplings $\mathbf{d}_{k\ell}$ is resource-intensive, alternative methods like curvature-driven surface hopping [78] have emerged. 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.

2. 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 [39]. 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 trajectories used for model training and validation, the other was 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 [39].

3. Simulation and Training Details

All nonadiabatic simulations were conducted using an interface between FieldMACE and the SHARC engine [62, 63]. 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 taking the relevant amount of data randomly from the full dataset. A 5 Ågstrom 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 states were used.

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

$$L = \frac{c_E}{n_L} \sum_{j=1}^{n_L} \sum_{i=1}^n \left(E_{REF}^{(i,j)} - \hat{E}^{(i,j)} \right)^2 + \frac{c_F}{n_L} \sum_{j=1}^{n_L} \sum_{i=1}^n \left(\mathbf{F}_{REF}^{(i,j)} - \hat{\mathbf{F}}^{(i,j)} \right)^2$$
(D3)

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 [39].