LEARNING THE ENERGY RELAXATION MANIFOLD FROM UNRELAXED STRUCTURES WITH RELAXNET

Anonymous authors
Paper under double-blind review

ABSTRACT

In efforts to bypass computationally-expensive density functional theory (DFT) calculations for energy minimization and structure relaxation, rapid progress in the development of machine learning force fields (MLFF) and more robust models that adhere to quantum chemistry/physical paradigms and constraints have been realized. However, most research to date involves energy predictions in a static frame only (i.e., given a specific atomic configuration, predict the energy of the current or final instance), which neglects intermediary physical insight-providing contexts. In this work, we developed RelaxNet, a dynamics-aware, equivariant deep learning model that leverages neural ordinary differential equations (neural ODEs) and message passing neural networks (MPNNs) for predicting the energy relaxation landscape between the initial unrelaxed structure and final relaxed structure for the first time. From just the initial structure, which is often the configuration that is fed into DFT simulations, we can accurately recover the energy/forces for the entire trajectory at a competitive prediction accuracy. We further provide extensive insights on the use of implicit vs. explicit latent embedding evolution schemes to offer perspectives on optimal methods for future works integrating expensive graph-based neural networks and neural ODEs.

1 INTRODUCTION

Density functional theory (DFT) simulations are long-standing prerequisites for most conventional atomistic simulation methods (e.g., Grand Canonical Monte Carlo, molecular dynamics) that require force field inputs. DFT can effectively optimize an unrelaxed structure to a more stable configuration by minimizing the potential energy landscape as guided by the forces acting on each atom in the structure. However, DFT is computationally-expensive and often exhibit slow convergence, especially for large/complex systems. There is also a growing need for adaptable and scalable force fields that can easily extend to different molecular structure groups (e.g., metal-based compounds vs. inorganic molecules). To bypass these time-consuming energy-based methods and address these point, machine-learned force fields (MLFF), which can be computed using surrogate models, have been proposed as a viable method of obtaining efficiently-computed atomic interaction parameters with DFT-level accuracy (Botu et al.; Chmiela et al. (a;c); Unke et al.). These models generally emphasize representation learning, specifically developing adequate architectures and features that can capture the underlying latent physics and quantum mechanics principles that are intrinsic to these molecular systems. While there are many advances in this direction, there is currently no work devoted to expanding these concepts towards dynamics-level frameworks. The ability to model the DFT relaxation trajectory from a structure's initial unrelaxed state to its final relaxed state, for example, can be valuable. Most of the existing research focuses on static, single-frame energy and force field predictions, which limits the scope of understanding how a structure evolves towards equilibrium. Traditional one-to-one structure–property mappings can be useful for obtaining the energy at a very specific given state (but provides little information on the equilibrium) or for predicting final states given the initial states (but lacks information on the intermediate states).

To model these dynamics, a framework that can inherently learn the derivatives and vector fields of the system, like neural ordinary differential equations (ODEs), would be ideal. Neural ODEs, however, can be time-consuming and memory-intensive, especially when large networks, like graph-based models, are used as the ODE backbone, since this method requires numerical integration evaluation at each solver step. The intermediate states are usually stored as well, which could

similarly blow up memory usage. Thus, we need to either (1) develop an efficient, yet sufficiently expressive, neural representation that can be updated in each ODE step, or (2) reformulate how the latent dynamics are updated within the model. In this work, we explore both options in depth, which can help inform appropriate model selection in the future. We bridge these primary knowledge gaps by introducing RelaxNet, a fully-equivariant, neural ODE-based model that can learn energies and force fields at the intermediate and final states from only the initial state. We also note that RelaxNet can be used to aid existing DFT efforts by predicting near-equilibrium states, which can serve as a starting point for DFT simulations, thus reducing the expected convergence time (compared to starting from the initial state). As the first trajectory-learning work in the MLFF field, this method can contribute greatly to advancing equivariant ODE-based models in addition to their applications in scientific domains.

To summarize, this work makes the following key contributions:

1. We propose a new physics-informed modeling framework to predict the relaxation trajectory, for the first time, based on energy and dynamics-conserving principles.
2. We develop RelaxNet, a fully-equivariant neural ODE model that is capable of learning smooth surrogate dynamics exhibited in DFT relaxation. Provided only the initial unrelaxed atomic configuration, the model can predict the energies and forces for n frames of the trajectory at the competitive accuracies demonstrated by other state-of-the-art models.
3. We analyzed the effects of using implicit vs. explicit latent embedding evolution in neural ODEs, which can inform future relaxation trajectory prediction works that marry equivariant graph-based and ODE-based neural networks.

2 Related Work

Equivariant graph neural networks. Since a crystal structure can be defined as an undirected graph, a graph neural network (GNN) can incorporate spatial information in the learning of local atomic properties and their interactions. Thus, GNNs are effective message passing-based frameworks that can allow for high-quality molecular structure representations. A graph, $G = (V, E)$, can be deconstructed into nodes, $v_i \in V$, and edges, $e_{ij} \in E$. Additional node- (h - e.g., electronegativity, atomic number) or edge-based (r - e.g., bond distance) features can be imposed. A general GNN workflow can be outlined in Eqs. (1a) and (1b), where m is the message used to update the nodal representation, σ_a is an aggregative operation, ϕ_m is the message passing function, and ϕ_u is a generic update function.

$$\mathbf{m}_{ij} = \sigma_a\big(\phi_m(\mathbf{h}_i^{(l)}, \mathbf{h}_j^{(l)}, r_{ij})\big), \forall j \in \mathcal{N}(i) \tag{1a}$$

$$\mathbf{h}_i^{(l+1)} = \phi_u\big(\mathbf{h}_i^{(l)}, \mathbf{m}_{ij}\big) \tag{1b}$$

Depending on the feature representation methods and message passing layers, different graph transformations can yield different predictions. This is nonideal for anisotropic predictions (e.g., forces), thus motivating the need for equivariant models. Specifically, the energy and forces of the molecular structure should be conserved and consistent after applying translational (Eq. (2a)) and rotational (Eq. (2b), where S is a rotational group function) transformations.

$$E(\mathbf{r}) = E(\mathbf{r} + \Delta\mathbf{x}) \tag{2a}$$

$$E(\mathbf{r}) = S(G(\mathbf{r})) \tag{2b}$$

Several methods of establishing equivariance in graph-based networks have been proposed. NequIP (Batzner et al.), for example, uses E(3)-equivariant convolutions to describe geometric tensor interactions. These convolution filters are based on learnable radial functions and spherical harmonics (to achieve rotational invariance). Other works have extended equivariance to an E(n)-space, as proposed by Satorras et al. (EGNN) and Mao et al. (ENINet), which ensures invariance with respect to translational, rotational, and reflective operations. In the former work, spherical harmonics were not used; instead, equivariant graph convolution layers were established with only distance-based inputs. Likewise, in the latter study, equivariant vectorial representations were introduced using directional message passing *via* many-body tensor representation (MBTR).

Energy and force field prediction. Naturally, these equivariant models can be extended to periodicity-dependent applications, like molecular modeling and DFT. In one such case, symmetric

gradient domain machine learning (sGDML) (Chmiela et al. (b)) model has been used to reproduce global potential energy surfaces (PES) for molecules with a few dozen atoms. Other works like DeepEF (Wu et al.) and DeePMD (Wang et al.) can predict energy/forces *via* an atomic self-attentive model coupled with a geometric optimizer (in the former case) and deep, rotationally-invariant neural network (in the latter case). Research efforts have also explored the use of genetic/evolutionary algorithms (Bin Faheem et al.) and KNN/random forest regressors (Nakata & Bai) for obtaining optimized force field parameters. In recent years, many studies have developed more robust ML frameworks for modeling quantum interactions and predicting energy (e.g., internal energy, atomization energy, in addition to other thermodynamic properties (e.g., enthalpy, Gibbs free energy, using the QM9 database, which features small, organic molecules. SchNet (Schütt et al.), for example, introduced continuous-filter convolutional neural networks, which more-accurately models local correlations, that allows for non-discretized PES reconstruction. In 2020, DimeNet (Gasteiger et al. (b)) (and later, DimeNet++ (Gasteiger et al. (a))) established directional message passing, which embeds messages passed rather than the atoms themselves. Later in 2023, Allegro (Musaelian et al.) presented a locally-equivariant deep neural network that doesn't use atom-centered message passing. Chen & Ong also developed M3GNET, which is a universal interatomic potential that incorporates three-body effects, unlike earlier models (e.g., CGCNN) that only consider pairwise interactions. Finally, there is ALIGNN (Choudhary & DeCost), a GNN-based model that performs message passing on the interatomic bond graph and line graph corresponding to bond angles.

Molecular structure relaxation. Earlier works on structure relaxation have also leveraged machine learning methods to obtain or accelerate the search for stable crystal configurations, although the discipline is currently not as well-explored as general force field prediction. In one such study, an iterative active learning approach can be used alongside machine-learning interatomic potentials (e.g., moment tensor potentials (MTPs)) to construct interatomic interaction models, thereby facilitating rapid crystal structure prediction (Podryabinkin et al.). Another study combined finetuned-MLFF and machine-learning surrogate models for learning reconstruction motifs and optimizing initial sampling structures to achieve energy-minimized structures, which can assist in point defect predictions (Mosquera-Lois et al.). Similarly, the introduction of DeepRelax (Yang et al.), a non-iterative equivariant graph neural network (EGNN)-based generative model that can predict relaxation quantities, has enabled low-energy crystal determination.

3 Background & Preliminaries

Energy minimization and force field calculations. For a given potential energy surface (PES), which can be defined by the total potential energy of a system (that is a function of the atomic coordinates, $\mathbf{r} = \{r_1, r_2, ..., r_N\} \in \mathbb{R}^{3N}$), we can minimize this energy, $E = E(\mathbf{r})$, to achieve the most stable molecular configuration. To enforce energy conservation and directly couple energy and force, we define the force acting on each atom in the x, y, z-directions by taking the negative gradient of energy with respect to position, as shown in Eq. (3). By jointly considering energy, force, and atomic coordinates, we can ensure physical consistency and allow the model to learn dynamic updates that mirror common iterative techniques, like Velocity Verlet and quasi-Newton algorithms.

$$\vec{F}(\mathbf{r}) = -\nabla_r E(r) \quad \Rightarrow \quad \vec{F}_i(\mathbf{r}) = -\partial E(\mathbf{r})/\partial r_i \text{ for } i = 1, ..., 3N \tag{3}$$

Modeling smooth surrogate relaxation dynamics. Neural ODEs can be used to capture the continuous, physics-based relaxation dynamics *via* hidden state evolution, as generally defined in Eqs. (4a) and (4b). The evolution of the latent state $\mathbf{s}(t)$ at pseudotime t is determined by an ODE, in which the governing dynamics are parameterized by a neural network f. A visualization of the relaxation trajectory is shown in Figure 1.

$$\frac{d\mathbf{s}(t)}{dt} = f(\mathbf{s}(t), t; \theta_p) \tag{4a}$$

$$\mathbf{s}(T) = \mathbf{s}(0) + \int_0^T f(\mathbf{s}(t), t; \theta_p) dt \tag{4b}$$

For physical systems, however, the ODE system (shown below), in its simplest form, can be defined as a function of the position $(\mathbf{r}(t))$ and velocity $(\mathbf{u}(t))$, like in Newton's second law. In this case, the force (F) is derived from energy, which is a learnable neural network.

$$\frac{d}{dt}\begin{bmatrix}\mathbf{r}(t)\\ \mathbf{u}(t)\end{bmatrix} = \begin{bmatrix}\mathbf{u}(t)\\ \frac{1}{m}F(\mathbf{r}(t))\end{bmatrix}$$

It should also be noted that DFT time, in this context, is not real physical time, rather a pseudotime (similar to those in score-based generative and normalizing flow models). Since we are observing the problem from an optimization standpoint, quasi-Newton methods with artificial timesteps should suffice as long as the natural order of the relaxation trajectory is preserved. The learnable network should intrinsically learn the surrogate progression rule (i.e., surrogate dynamic) to equilibrium, since force is still evaluated at each step.

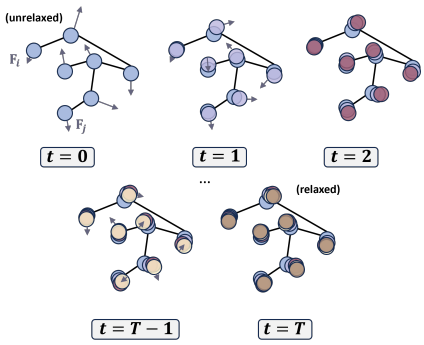

Figure 1: An overview visualization of the relaxation trajectory from the initial unrelaxed state to final relaxed state, as governed by forces (and structure-level energy minimization).

4 METHODS

4.1 DATA

The dataset was derived from JARVIS DFT (3D) database, which contains 33,473 unique crystal structures. However, after data processing, only 28,342 samples was used for training. From the `vasprun.xml` file of each structure, we extracted the atomic energy and per-atom forces (at different timesteps of the energy minimization process), their corresponding structures (x, y, z-coordinates), and the atomic numbers. We can also extract the total charges from the `OUTCAR` file. Finally, we can retrieve the adjacency/connectivity matrix of each structure *via* crystallographic (`.cif`) files. The structural information, atomic number, and coordination number will define the input features, and the energy/forces will constitute the outputs. Unlike the QM9 database, which has organic structures exhibiting only 5 unique elements and made up of less than 30 atoms, the JARVIS DFT database features structures with 86 distinct elements, including metals, and a maximum of 96 atoms. The data distribution for both datasets can be found in Figure A.1.

4.2 ENFORCING EQUIVARIANCE

Spherical harmonics. To establish equivariance and maintain translational/rotational invariance, explicit cartesian inputs should be avoided. Instead, we can map the cartesian space onto a spherical surface, which can help establish equivariance while also maintaining the flexibility of point inputs. Spherical harmonic expansion, for example, can be used to enforce rotational invariance, since it is constituted by orthonormal integration over all angles. Spherical harmonics, $Y_m^{(l)}(\theta, \varphi)$, form basis functions for angular-dependent models (and follows Wigner D-matrices under rotation) by encoding the angular position of the node, and by extension, the global rotational symmetries.

The spherical harmonics are given by Eq. (5), where $N_m^{(l)}$ is the normalization constant that enforces orthonormality, $\exp(im\varphi)$ is the azimuthal component, and $P_m^{(l)}(\cos\theta)$ are the Legendre polynomials. Here, θ is the polar angle/colatitude, defined as $\theta = \arccos(z/r)$, $\phi = \arctan(y/x)$, l is the orbital angular momentum quantum number, and m is the magnetic quantum number ($-l \leq m \leq l$). The quantum numbers ultimately dictate the complexity of the spherical harmonic embeddings.

$$Y_m^{(l)}(\theta, \varphi) = \sum_m N_m^{(l)} \exp(im\varphi) P_m^{(l)}(\cos\theta) \tag{5}$$

Linking spherical Bessel functions and spherical harmonics expansion. To incorporate local, relative spatial information (e.g., distance), we can expand the spherical harmonics to include a radial coefficient, $c_m^{(l)}(r)$. By using relative displacement, translational invariance can also be preserved. This term can be expressed using (1) gaussian radial basis functions (RBFs), defined as $c_m^{(l)}(r) = \exp(-\beta(r - r_0)^2)$, (2) polynomial basis functions (which can also be used as an envelope function for radial functions to increase the general expressivity of the representation), and (3) spherical Bessel functions, which can be used to capture wave-like properties as commonly seen in quantum mechanical systems, like in the Schrödinger equation (defined in Eq. (6a) for Bessel

function of the first kind). RBFs alone are not as robust for molecular systems, since they can only approximate, rather than fully capture, the oscillatory and decaying effects inherent to atomistic simulations. Additionally, RBFs do not abide by orthogonality, unlike spherical Bessel functions. The full spherical harmonics expansion (Eq. (6b)) considers both angular and radial information.

$$c_m^{(l)}(r) = \sqrt{\pi/2r} J_{l+1/2}(r) \tag{6a}$$

$$f(r,\theta,\varphi) = \sum_{l=0}^{l_{max}} \sum_{m=-1}^{l} c_m^{(l)}(r) Y_m^{(l)}(\theta,\varphi) \tag{6b}$$

4.3 RelaxNet Architecture

The primary inputs to RelaxNet are the invariant node features described in Section 4.1 and the transformed radial features described in Section 4.2. The overall network can be found in Figure 2.

Interaction layer. The atomic interactions can be encoded with the interaction layer (DTNN module), which is comprised of a GNN for capturing pairwise interactions and a MPNN (denoted as TripletMPNN) for higher-order, three-body interactions. The angles are computed for each three-atom system and are cosine-transformed and passed through a zonal spherical harmonics function (which neglects dihedral effects). These angular embeddings are concatenated with the bond distances and spherical harmonics-expanded radial features, and will serve as inputs to the two MLP blocks, MLP_s and MLP_v, which will yield scalar and vector messages, respectively. These scalar messages are responsible for updating the existing node features. The outputs of the MLP_v block will be multiplied with an orthogonal basis set, B, which is constructed with the Gram-Schmidt algorithm, and are then aggregated on a per-node level. The outputs from the mixed vector embedding layers, MLP_m, can be multiplied with the aggregated embeddings and added to the gated vector embeddings (calculated from MLP_g) to get the updated vector features. Similarly, the positions/coordinates can be updated by multiplying the outputs from the MLP_c block with the direction unit vectors and subsequently adding them to the original position. The embeddings from each DTNN layer in the encoder block will be updated and concatenated to make up the final embedding.

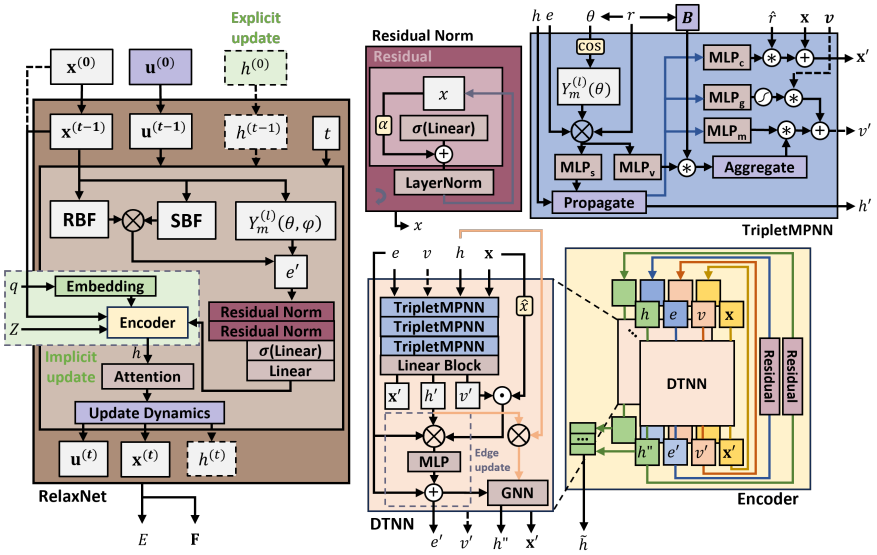

Figure 2: The **RelaxNet** architecture leverages a neural ODE wrapper to evolve the surrogate dynamics involved in DFT relaxation *via* either **implicit or explicit latent embedding update** steps. The latent embeddings, h, are computed from the **encoder block**, which is made up of **DTNN layers** with **residual** connections. The DTNN layer is comprised of **TripletMPNN layers** and a pairwise GNN. The dotted boxes/arrows indicate variables/blocks that can be excluded (vector messages, v) or interchanged (h updates) in the overall model.

Neural ODE. We initialize the neural ODE with the states $\mathbf{x}^{(0)}$ and $\mathbf{u}^{(0)}$, where $\mathbf{x}^{(0)}$ is the initial unrelaxed position and $\mathbf{u}^{(0)}$ is a learned (initial velocity) embedding from the spherical harmonics-expanded edge features for the initial structure. The encoder is used to produce interaction-based

embeddings based on the position of the atoms at each pseudotime step, and are passed to an attention block. Finally, we pass these scalar embeddings to a MLP block to get the per-structure energy, after which the per-atom forces can be computed by taking the gradient of the energy with respect to the position of each atom. We can then update the surrogate dynamics states by setting $d\mathbf{x}/dt = \mathbf{u}$ and $d\mathbf{u}/dt = \hat{F}/m$, where m is the atomic mass. For the ODE solver, we used a `rk4` fixed step solver with a relative tolerance of `rtol`=1e-3 and an absolute tolerance of `atol`=1e-3.

Implicit vs. explicit latent embedding evolution. We experimented with two different methods of latent embedding updates. When the latent embeddings evolve implicitly (i.e., the interaction embedding is not an initialized state for the neural ODE), the latent variable, h, is recomputed for each evolved positional state, $\mathbf{x}^{(t)}$. On the contrary, the h for explicit latent embedding evolution is established as an initial state (denoted as $h^{(0)}$), and is calculated using the same encoding block (but at the initial unrelaxed position) and updated *via* a MLP for each pseudotime step.

4.4 Training

Loss function. The model is directly supervised on several types of losses, including the per-structure (1) energy (E) and (2) energy landscape's monotonicity, and the per-atom (3) forces (F), (4) displacement ($\mathbf{x}$), and (5) directions. Here, t is the frame number in a trajectory with a total of T frames, s is the sample number in a batch with S structures, and N is the number of atoms in the system. Loss terms 1, 3, and 4 simply ensure that the predicted values (e.g., $\hat{y}$) are close to their ground-truth values (e.g., y), while loss term 2 enforces a smooth energy function and a monotonically-decreasing energy throughout the relaxation trajectory. Finally, loss term 5, modeled by cosine similarity, imposes a stricter alignment between the true and predicted directions. The loss function (Eq. (7)) is defined as the following, where λ represents the weight for each loss term.

$$\begin{aligned}\mathscr{L} = &\frac{\lambda_1}{T\cdot S}\sum_{t=1}^{T}\sum_{s=1}^{S}\left|E_{t,s}-\hat{E}_{t,s}\right| + \frac{\lambda_2}{T(S-1)}\sum_{t=1}^{T}\sum_{s=1}^{S-1}\mathrm{ReLU}\big(\hat{E}_{t,s+1}-\hat{E}_{t,s}\big)\\ &+\frac{\lambda_3}{T\cdot S\cdot 3N}\sum_{t=1}^{T}\sum_{s=1}^{S}\sum_{i=1}^{3N}\big(F_{t,s,i}-\hat{F}_{t,s,i}\big)^2 + \frac{\lambda_4}{T\cdot S\cdot 3N}\sum_{t=1}^{T}\sum_{s=1}^{S}\sum_{i=1}^{3N}\big(\mathbf{x}_{t,s,i}-\hat{\mathbf{x}}_{t,s,i}\big)^2\\ &+\lambda_5\frac{\sum_{i\in\mathcal{M}}(1-\cos\theta_i)\|\Delta\mathbf{x}_i\|}{\sum_{i\in\mathcal{M}}\|\Delta\mathbf{x}_i\|+\delta},\quad \text{where } \mathbf{x}=\mathbf{x}_{t_f}-\mathbf{x}_{t_0} \text{ and } \cos\theta_i=\frac{\mathbf{x_i}}{\|\Delta\mathbf{x}_i\|}\cdot\frac{\hat{\mathbf{x}}_\mathbf{i}}{\|\Delta\hat{\mathbf{x}}_\mathbf{i}\|}\end{aligned} \tag{7}$$

Hyperparameters. During training, we use an AdamW optimizer with an initial learning rate of 5e-4 (`ReduceLROnPlateau` scheduler with a factor of 0.8 and an 8-epoch patience) and a weight decay of 5e-5. The model is trained over 1000 epochs with a training patience of 20 epochs to prevent overfitting. The train, validation, and test splits are 90/5/5. We used two NVIDIA Volta V100 GPUs (1 GPU = 32 GB), each accommodating a batch size of 64 crystals.

5 Experiments

We performed several experiments to quantify the performance of the model. First, we establish a baseline model with just the encoder block (no neural ODE), examining the effects of the number of DTNN blocks on energy prediction quality. Next, we explore the effects of using implicit vs. explicit latent embedding evolution by comparing the energy/force mean absolute error (MAE) and average per-epoch training time for various cases by modulating the number of frames and structures. Finally, we benchmarked RelaxNet with other state-of-the-art energy prediction models and further investigated the effects of energy/displacement on the MAE_E.

5.1 Baseline Model: Final Frame Prediction

We first consider a baseline model that is constructed from only an encoder block (no neural ODE envelope). For various training set sizes and different numbers of DTNN blocks (e.g., 2, 4), we evaluate the MAE_E at the final frame and the average time it takes the model to complete an epoch, as tabulated in Table 1. The results indicate that using 4 DTNN blocks yields higher MAE_E and per-epoch training duration, which could be attributed to overfitting.

Thus, for future experiments, we constrain our model to only 2 DTNN blocks. The lower memory cost and faster runtime associated with using less DTNN blocks would also be advantageous for regulating the neural ODE's computational efficiency. We similarly observe that in using more unique training samples, the model can achieve better energy prediction performance.

Table 1: Global energy MAEs for the baseline model (no neural ODE) predictions at the final frame only using 2 DTNN blocks/4 DTNN blocks for different number of samples.

S	MAE_E [eV] (↓)	Avg. time per epoch [min] (↓)
1,000	6.362/16.947	1.24/2.10
2,000	5.824/13.103	2.32/3.74
5,000	3.436/8.606	5.39/9.38

5.2 Ablation Study: Trajectory Prediction

In the second experiment, we examine RelaxNet's performance by comparing the use of an implicit vs. explicit latent embedding evolution scheme. To also better understand the effects of trajectory length (i.e., the number of considered frames) and the number of unique crystal structures on prediction quality, we performed an auxiliary ablation study. For demonstration, we used only 2 DTNN blocks for all cases in this section. Additionally, to ensure fair comparisons for each case, we collected all samples that have at least n frames in the trajectory. We then sample n equidistant frames from the full trajectory (including the initial and final frames).

Table 2: An ablation study comparing the implicit vs. explicit latent embedding evolution cases for different numbers of trajectory frames and unique crystal structures. The shaded rows indicate cases where the same structures are used. The global MAEs for energy/force predictions and the average training time per epoch are tabulated. The percent change from the baseline (2 DTNN blocks) final frame MAE_E and per-epoch runtime are reported. The percent change from the implicit all-frame MAE_E and per-epoch runtime are also noted.

	n	S	S_{tot}	Global MAE_E (final/all) [eV] (↓)	MAE_F (final/all) [eV/Å] (↓)	Avg. time per epoch [min] (↓)
Implicit latent embedding evolution	3	5,000	15,000	3.670/3.671	0.0005/0.0068	133.30
	5	5,000	25,000	5.053/5.042 (+47.1%)	0.0001/0.0051	293.63 (+5347.7%)
	10	5,000	50,000	6.457/6.472	0.0003/0.0044	311.25
	5	2,000	10,000	8.727/8.718 (+49.8%)	0.0008/0.0081	61.88 (+2567.2%)
	10	1,000	10,000	5.892/5.918 (-7.4%)	0.0005/0.0120	31.16 (+2412.9%)
Explicit latent embedding evolution	3	5,000	15,000	2.983/3.031 (-17.4%)	0.0008/0.0031	17.95 (-86.5%)
	5	5,000	25,000	2.764/2.781 (-44.8%)	0.0004/0.0025	27.50 (-90.6%)
	10	5,000	50,000	2.235/2.257 (-65.1%)	0.0003/0.0023	49.84 (-84.0%)
	5	2,000	10,000	4.582/4.652 (-46.6%)	0.0007/0.0043	11.34 (-81.7%)
	10	1,000	10,000	4.503/4.174 (-29.4%)	0.0015/0.0108	11.54 (-63.0%)
	5	24,311	121,555	2.919/2.996	0.0000/0.0005	136.80

Comparing implicit vs. explicit latent evolution methods. In Table 2, we disentangle the effects of (1) the number of frames by using the same 5,000 structures (shaded rows) and (2) the number of total samples in the training set. For the explicit cases, the global MAE_E and per-atom MAE_F decreases as the number of frames increase for the same 5,000 unique structures, which can be attributed to the intermediate frames providing more context to the model in reconstructing a smoother relaxation trajectory. We also note that after training with the same total number of samples, S_{tot}, the final and aggregated MAE_E is comparable for both cases, while the MAE_F is higher for the case with less unique samples but more frames, which highlights the importance of diversity and size of the training set. On the contrary, the implicit model results reveal that MAE_E increases as more frames are used. We also want to note that despite the long training duration of each epoch, the model converges relatively quickly (20-200 epochs), unlike other models that require 100s or 1000s of epochs for full convergence.

From this ablation study, we demonstrate that using explicit latent embedding updates is sufficient for obtaining full trajectory energy predictions with lower MAE_E (17.4-65.1% better) and quicker per-epoch training time (63.0-90.6% faster) than the implicit scheme. This method would also be more suitable for large training sets.

Baseline vs. trajectory-level prediction performance. For the same number of unique structures, the MAE_E at the final frame for the non-neural ODE (baseline) cases are 19.6-29.2% higher (compared to the explicit latent embedding evolution case), highlighting the advantage of using a neural ODE-based model with an equivariant encoder layer as the foundation. Compared to the baseline MAE_E, however, the implicit model achieves better energy predictions for the 1,000 sample case.

Per-frame MAEs. We can further break down the global energy and force errors into per-frame and cumulative MAEs, as depicted in Figure 3, which can provide clearer insights into the individual contributions of each state along the trajectory. In particular, we performed this analysis for the implicit latent embedding case with 10 frames and 1,000 samples. From the MAE vs. frame index plots, it is evident that the per-frame energy and force MAEs decrease exponentially, with the lowest MAEs observed at the final relaxed frame, which is generally expected since the energy/forces decrease as the structure approaches a more stable configuration. Similarly, we note that the cumulative probability curve shifts to the left for each relaxation step, indicating that there are lower errors with more relaxed frames. The overall energy predictions also show strong agreements, as indicated by the parity plot (PCC=0.979).

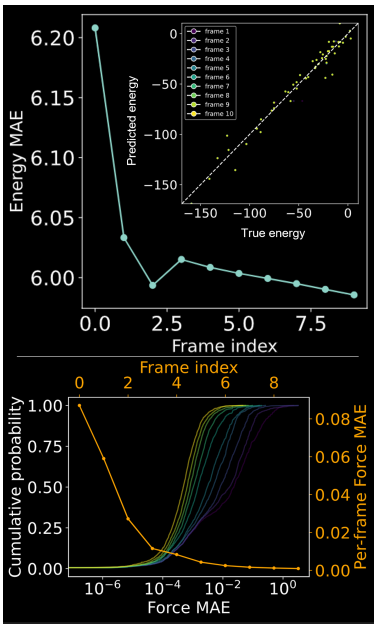

Figure 3: RelaxNet's energy and force prediction performance for the implicit latent embedding evolution/10 frame/1,000 structure case. The (top) per-frame energy MAEs are plotted with an inset parity plot comparing the ground-truth and predicted energies. Similarly, the (bottom) per-frame force MAEs and cumulative density function of these MAEs curves are plotted.

5.3 Benchmarking

We then compared the RelaxNet model with explicit latent embedding evolution to other state-of-the-art energy prediction models (detailed in the JARVIS leaderboard from Choudhary et al.), like ALIGNN (Choudhary & DeCost), MatFormer (Cui et al.), PotNet (Lin et al.), and KGCNN (Reiser et al.), as shown in Table 3.

Table 3: Benchmarking RelaxNet with other state-of-the-art energy prediction models.

Model	S	MAE_E [meV/atom] (↓)
ALIGNN (all)	55,713	33.1
MatFormer (all)	55,713	32.2
potnet (all)	55,713	29.3
kgcnn_coNGN (all)	55,713	29.1
kgcnn_coGN (all)	55,713	27.1
RelaxNet (final/all)	24,311	17.060/24.646

Specifically, we compared the MAE_E (obtained from k-fold cross-validation on the test set) in our work for the 24,311 unique structure case. Currently, the kgcnn_coGN can achieve the lowest MAE_E at 27.1 meV/atom, followed by kgcnn_coNGN at 29.1 meV/atom. We can also broadly deduce the models' performance with other datasets. For example, M3GNET (Chen & Ong), which is trained on the trajectory of 62,783 unique crystals (total of 187,687 samples) from the Materials Project (MP) database, yielded a MAE_E of 35 meV/atom. In the extensive benchmarking study conducted by Choudhary & DeCost that compared the MAE_E for different models across the MP and JARVIS databases, the authors noted that the models attained higher performance on the MP database, likely due to MP's larger dataset and lower target energies (50% lower) for MP structures. RelaxNet can achieve a MAE_E of 17.060 (final frame) and 24.646 (all frame) meV/atom.

We want to emphasize, however, that this benchmarking study has a few qualifications. First, RelaxNet was trained with fewer samples overall, since the neural ODE requires exactly n frames, which can lead to a truncated database if $n > n_{tot}$ (n_{tot} = total number of frames for the structure) or complete sample omission if $n < n_{tot}$. As evident from Table 2, the MAE_E (especially on a per-atom level) generally decreases with more training samples. Second, the RelaxNet predicts the entire trajectory energies with a single initial frame (so intermediate energy predictions are also reliant on predicted intermediate states), while the other works predict the energies with absolute states. With

these considerations and the benchmarking results in mind, we note that while this study does not provide a strict one-to-one comparison, we can extrapolate that RelaxNet can achieve a relatively competitive prediction accuracy.

5.4 Effects of energy and displacement on energy prediction

We can also visualize the general effect of energy and average per-frame displacement ($\mathbf{x}_t - \mathbf{x}_0$) on MAE_E, as outlined in Figure 4. First, we observed that most true displacements are concentrated between 10^{-3} and 10^{-2} $\mathring{A}$. Second, the MAE_E decreases as the movement increases and energy magnitude decreases. This is consistent with earlier results, such that RelaxNet generated better final frame predictions, which could be attributed to the flatter energy landscape near equilibrium (e.g., smaller gradients), hence improving final energy predictions. Moreover, for higher energy magnitudes (regardless of signs), the resulting MAE_E is higher, as expected.

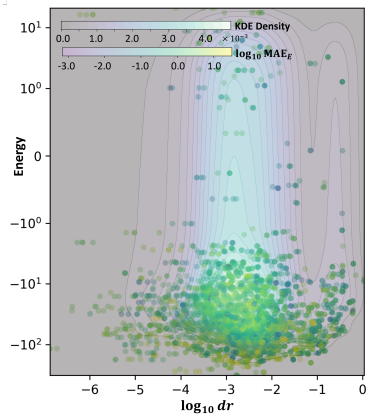

Figure 4: The MAE_E (and overlayed KDE density) plotted against the energy and average per-frame displacement ($\mathbf{x}_t - \mathbf{x}_0$).

6 Conclusion

In this work, we developed RelaxNet, a surrogate dynamics-aware neural ODE-based model with a fully-equivariant backbone. We demonstrated the ability to evolve this descriptive, graph-based neural network ODE scaffold without memory explosion and within a reasonable training time. Moreover, unlike earlier works that predict energy/forces at static frames, our current model, for the first time, can predict the per-structure energy and per-atom forces for the entire DFT relaxation trajectory, provided only the initial unrelaxed frame (i.e., positions). In addition, we thoroughly explored the effects of using implicit vs. explicit latent embedding evolution schemes *via* ablation studies. Afterwards, we benchmarked RelaxNet with other state-of-the-art force field prediction models, which revealed the exceptional predictive capabilities of our model. To summarize, this work can expedite computationally-expensive DFT studies by (1) informing researchers on the expected final energy/forces and (2) providing users a more equilibrated (i.e., more stable crystal configuration) starting point for their DFT simulations.

7 Limitations & Future Works

To our knowledge, we are the first to predict the DFT relaxation trajectory from the initial unrelaxed state. While we were able to achieve promising results in regards to energy and force predictions, we recognize that the model could benefit from higher computational efficiency (i.e., faster training). There are also opportunities for extending this work to include stress prediction for the full trajectory, since the stress can be calculated by taking the gradient of energy with respect to the strain or lattice geometry. To achieve this, we can adapt our existing model to jointly train energy, force, displacement, and stress. An additional state to account for the lattice dynamics in the neural ODE may also be required. Another research direction could potentially include shifting the problem from a predictive to a generative framework. While neural ODEs can naturally digest and learn dynamics, they are also limited by their requirement for a specified trajectory length. Transformers that are constrained with physics-based equations, therefore, can be a powerful alternative, since they can generate each frame autoregressively, irrespective of trajectory length. This method can also be more suitable for the JARVIS dataset and offer more flexibility, because the trajectory lengths for each structure can vary widely.

By acknowledging these limitations, we highlight the large opportunity space and potential for future exploration in relaxation trajectory research. For example, while this work is currently constrained to JARVIS materials, it can also be extended to the biology discipline (e.g., protein folding, which is essentially structure relaxation) to accelerate drug discovery efforts. Similarly, we can apply these principles to other traditional atomistic simulation methods, which can significantly decrease computational requirements and runtimes.

ACKNOWLEDGMENTS

Hidden for anonymity.

REFERENCES

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

Abdullah Bin Faheem, Jong-Yun Kim, Sang-Eun Bae, and Kyung-Koo Lee. Efficient parameterization of intermolecular force fields for molecular dynamics simulations via genetic algorithms. 337:116579. ISSN 0167-7322. doi: 10.1016/j.molliq.2021.116579. URL https://www.sciencedirect.com/science/article/pii/S0167732221013039.

V. Botu, R. Batra, J. Chapman, and R. Ramprasad. Machine Learning Force Fields: Construction, Validation, and Outlook. 121(1):511–522. ISSN 1932-7447. doi: 10.1021/acs.jpcc.6b10908. URL https://doi.org/10.1021/acs.jpcc.6b10908.

Chi Chen and Ping Ong. A universal graph deep learning interatomic potential for the periodic table. 2(11):718–728. ISSN 2662-8457. doi: 10.1038/s43588-022-00349-3. URL https://www.nature.com/articles/s43588-022-00349-3.

Stefan Chmiela, Huziel E. Sauceda, Klaus-Robert Müller, and Alexandre Tkatchenko. Towards exact molecular dynamics simulations with machine-learned force fields. 9(1):3887, a. ISSN 2041-1723. doi: 10.1038/s41467-018-06169-2. URL https://www.nature.com/articles/s41467-018-06169-2.

Stefan Chmiela, Huziel E. Sauceda, Igor Poltavsky, Klaus-Robert Müller, and Alexandre Tkatchenko. sGDML: Constructing accurate and data efficient molecular force fields using machine learning. 240:38–45, b. ISSN 0010-4655. doi: 10.1016/j.cpc.2019.02.007. URL https://www.sciencedirect.com/science/article/pii/S0010465519300591.

Stefan Chmiela, Valentin Vassilev-Galindo, Oliver T. Unke, Adil Kabylda, Huziel E. Sauceda, Alexandre Tkatchenko, and Klaus-Robert Müller. Accurate global machine learning force fields for molecules with hundreds of atoms. 9(2):eadf0873, c. doi: 10.1126/sciadv.adf0873. URL https://www.science.org/doi/full/10.1126/sciadv.adf0873.

Kamal Choudhary and Brian DeCost. Atomistic Line Graph Neural Network for improved materials property predictions. 7(1):185. ISSN 2057-3960. doi: 10.1038/s41524-021-00650-1. URL https://www.nature.com/articles/s41524-021-00650-1.

Kamal Choudhary, Daniel Wines, Kangming Li, Kevin F. Garrity, Vishu Gupta, Aldo H. Romero, Jaron T. Krogel, Kayahan Saritas, Addis Fuhr, Panchapakesan Ganesh, Paul R. C. Kent, Keqiang Yan, Yuchao Lin, Shuiwang Ji, Ben Blaiszik, Patrick Reiser, Pascal Friederich, Ankit Agrawal, Pratyush Tiwary, Eric Beyerle, Peter Minch, Trevor David Rhone, Ichiro Takeuchi, Robert B. Wexler, Arun Mannodi-Kanakkithodi, Elif Ertekin, Avanish Mishra, Nithin Mathew, Mitchell Wood, Andrew Dale Rohskopf, Jason Hattrick-Simpers, Shih-Han Wang, Luke E. K. Achenie, Hongliang Xin, Maureen Williams, Adam J. Biacchi, and Francesca Tavazza. JARVIS-Leaderboard: A large scale benchmark of materials design methods. 10(1):93. ISSN 2057-3960. doi: 10.1038/s41524-024-01259-w. URL https://www.nature.com/articles/s41524-024-01259-w.

Jiyu Cui, Fang Wu, Wen Zhang, Lifeng Yang, Jianbo Hu, Yin Fang, Peng Ye, Qiang Zhang, Xian Suo, Yiming Mo, Xili Cui, Huajun Chen, and Huabin Xing. Direct prediction of gas adsorption via spatial atom interaction learning. 14(1):7043. ISSN 2041-1723. doi: 10.1038/s41467-023-42863-6. URL https://www.nature.com/articles/s41467-023-42863-6.

Johannes Gasteiger, Shankari Giri, Johannes T. Margraf, and Stephan Günnemann. Fast and Uncertainty-Aware Directional Message Passing for Non-Equilibrium Molecules, a. URL http://arxiv.org/abs/2011.14115.

Johannes Gasteiger, Janek Groß, and Stephan Günnemann. Directional Message Passing for Molecular Graphs, b. URL http://arxiv.org/abs/2003.03123.

Yuchao Lin, Keqiang Yan, Youzhi Luo, Yi Liu, Xiaoning Qian, and Shuiwang Ji. Efficient Approximations of Complete Interatomic Potentials for Crystal Property Prediction. URL http://arxiv.org/abs/2306.10045.

Jiashun Mao, Jianmin Wang, Amir Zeb, Kwang-Hwi Cho, Haiyan Jin, Jongwan Kim, Onju Lee, Yunyun Wang, and Kyoung Tai No. Transformer-Based Molecular Generative Model for Antiviral Drug Design. 64(7):2733–2745. ISSN 1549-9596. doi: 10.1021/acs.jcim.3c00536. URL https://doi.org/10.1021/acs.jcim.3c00536.

Irea Mosquera-Lois, Seán R. Kavanagh, Alex M. Ganose, and Aron Walsh. Machine-learning structural reconstructions for accelerated point defect calculations. 10(1):1–9. ISSN 2057-3960. doi: 10.1038/s41524-024-01303-9. URL https://www.nature.com/articles/s41524-024-01303-9.

Albert Musaelian, Simon Batzner, Anders Johansson, Lixin Sun, Cameron J. Owen, Mordechai Kornbluth, and Boris Kozinsky. Learning local equivariant representations for large-scale atomistic dynamics. 14(1):579. ISSN 2041-1723. doi: 10.1038/s41467-023-36329-y. URL https://www.nature.com/articles/s41467-023-36329-y.

Hiroya Nakata and Shandan Bai. Development of a new parameter optimization scheme for a reactive force field based on a machine learning approach. 40(23):2000–2012. ISSN 1096-987X. doi: 10.1002/jcc.25841. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.25841.

Evgeny V. Podryabinkin, Evgeny V. Tikhonov, Alexander V. Shapeev, and Artem R. Oganov. Accelerating crystal structure prediction by machine-learning interatomic potentials with active learning. 99(6):064114. doi: 10.1103/PhysRevB.99.064114. URL https://link.aps.org/doi/10.1103/PhysRevB.99.064114.

Patrick Reiser, André Eberhard, and Pascal Friederich. Graph neural networks in TensorFlow-Keras with RaggedTensor representation (kgcnn). 9:100095. ISSN 2665-9638. doi: 10.1016/j.simpa.2021.100095. URL https://www.sciencedirect.com/science/article/pii/S266596382100035X.

Victor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E(n) Equivariant Graph Neural Networks. URL http://arxiv.org/abs/2102.09844.

Kristof T. Schütt, Pieter-Jan Kindermans, Huziel E. Sauceda, Stefan Chmiela, Alexandre Tkatchenko, and Klaus-Robert Müller. SchNet: A continuous-filter convolutional neural network for modeling quantum interactions. URL http://arxiv.org/abs/1706.08566.

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

Han Wang, Linfeng Zhang, Jiequn Han, and Weinan E. DeePMD-kit: A deep learning package for many-body potential energy representation and molecular dynamics. 228:178–184. ISSN 0010-4655. doi: 10.1016/j.cpc.2018.03.016. URL https://www.sciencedirect.com/science/article/pii/S0010465518300882.

Guoxin Wu, Qilei Liu, Jian Du, Qingwei Meng, and Lei Zhang. A deep learning-based energy and force prediction framework for high-throughput quantum chemistry calculations. In Flavio Manenti and Gintaras V. Reklaitis (eds.), *Computer Aided Chemical Engineering*, volume 53 of *34 European Symposium on Computer Aided Process Engineering / 15 International Symposium on Process Systems Engineering*, pp. 715–720. Elsevier. doi: 10.1016/

B978-0-443-28824-1.50120-4. URL https://www.sciencedirect.com/science/article/pii/B9780443288241501204.

Ziduo Yang, Yi-Ming Zhao, Xian Wang, Xiaoqing Liu, Xiuying Zhang, Yifan Li, Qiujie Lv, Calvin Yu-Chian Chen, and Lei Shen. Scalable crystal structure relaxation using an iteration-free deep generative model with uncertainty quantification. 15(1):8148. ISSN 2041-1723. doi: 10.1038/s41467-024-52378-3. URL https://www.nature.com/articles/s41467-024-52378-3.

A Appendix

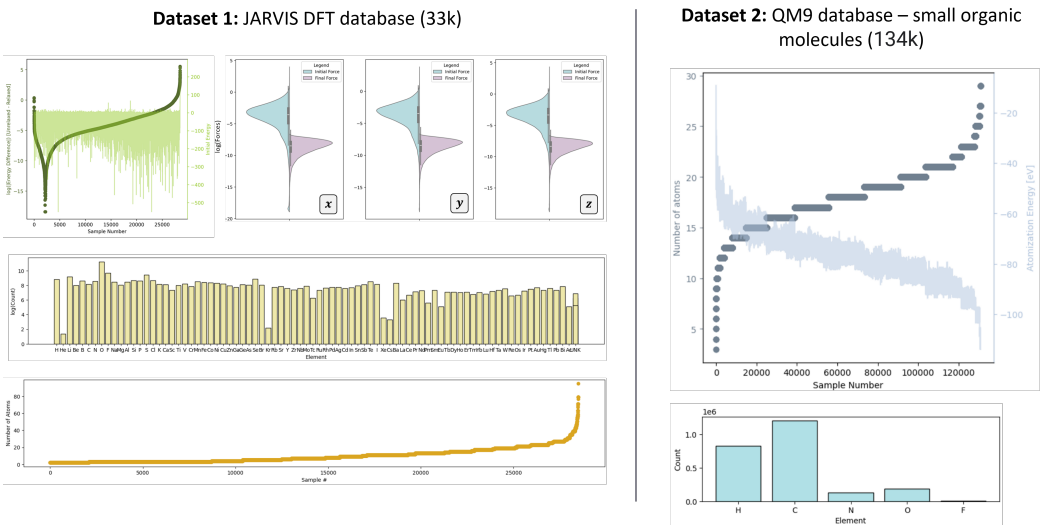

Figure A.1: The data distribution of the JARVIS DFT 3D database, featuring 33,000 different molecular structures, and the QM9 database, which consists of 134,000 small organic molecules. Only the initial and final structures' distributions are included in both datasets.