

Supplementary Material: Graph neural network force field for coarse-grain molecular crystal RDX

Brian H. Lee,¹ James P. Larentzos,² John K. Brennan,² Alejandro Strachan^{1, a)}

Author affiliations

¹*School of Materials Engineering and Birck Nanotechnology Center, Purdue University, West Lafayette, Indiana 47907, USA*

²*U.S. Army Combat Capabilities Development Command (DEVCOM) Army Research Laboratory, Aberdeen Proving Ground, Maryland 21005, USA*

Author email

^{a)}Author to whom correspondence should be addressed: strachan@purdue.edu

1. Model features

In training the GNNFF, we have experimented with two types of edge features and two types of node features. For the first node feature type, the intramolecular node features were two-dimensional one-hot vectors indicating whether the particle is nitro or triazine group. For the second node feature type, the node features had an additional dimension that recorded the local density of the particle, whose definition is given by Moore et al.¹,

$$\rho_{local,i}(r_{ij}) = \frac{84}{5\pi r_{cut}^3} \left(1 + \frac{3r_{ij}}{2r_{cut}}\right) \left(1 - \frac{r_{ij}}{r_{cut}}\right)^4 \quad (1)$$

Here, $\rho_{local,i}$, r_{ij} , r_{cut} are the local density of particle i , distance between particles i and j , and the cutoff distance corresponding to 9 Å in our model, respectively. We find that the density feature does not significantly affect the crystal and amorphous structures observed from the CG simulations. Therefore, we use this feature only for data generation as explained in Supplementary Section 2 of Supplementary Material (SM). All of the results in the main manuscripts are from the first node feature type without the local density as input.

The types of edge features were varied to either accept raw coordinates or an expansion of the coordinates in Gaussian bases. The edge features of intramolecular stretching interactions (e_{str}) with raw coordinates were three-dimensional vectors equal to $[d_{N_i-T}, \theta_{N_i-T-N_j}, \theta_{N_i-T-N_k}]$. The intramolecular angular features (e_{ang}) were $[\theta_{N_i-T-N_j}, d_{N_i-T}, d_{T-N_j}]$ and the intermolecular features (e_{inter}) were $[f_c(d)]$, where f_c is a smoothly decaying function defined by Behler et al.²,

$$f_c(d_{ij}) = \begin{cases} 0.5[\cos(\pi d_{ij}/r_{cut}) + 1] & \text{for } d_{ij} \leq r_{cut} \\ 0 & \text{for } d_{ij} > r_{cut} \end{cases} \quad (2)$$

Here, r_{cut} is the cutoff distance and is defined as 9 Å in this study to include the first three peaks of the radial distribution function of crystalline RDX.

For the edge features with Gaussian bases, we expand the raw features with four Gaussian functions. For example, the intermolecular edge features were $[G_{1,\text{inter}}(f_c(d)), G_{2,\text{inter}}(f_c(d)), G_{3,\text{inter}}(f_c(d)), G_{4,\text{inter}}(f_c(d))]$. The intramolecular edge features were 12-dimensional vectors that expand the raw features with Gaussian functions. The Gaussian functions are defined as,

$$G_n(x) = e^{-\eta_n(x-x_{0,n})^2} \quad (3)$$

The parameters for the Gaussian functions ($\eta_n, x_{0,n}$) are given in Supplementary Table 1. These values are chosen based on the inter- and intramolecular configurations in our training dataset described in Supplementary Section 2, Supplementary Figure 1.

Supplementary Table 1. Gaussian function parameters

Type	η	x_0
$G_{1,\text{str}}$	5	2 Å
$G_{2,\text{str}}$	20	2.7 Å
$G_{3,\text{str}}$	20	3.1 Å
$G_{4,\text{str}}$	5	4 Å
$G_{1,\text{ang}}$	10	1.047 rad
$G_{2,\text{ang}}$	20	1.395 rad
$G_{3,\text{ang}}$	5	2.094 rad
$G_{4,\text{ang}}$	10	2.793 rad
$G_{1,\text{inter}}$	0.15	0 Å

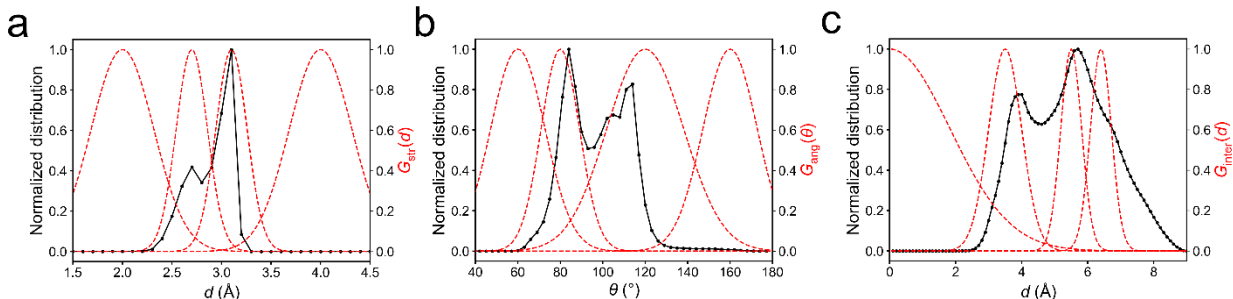
$G_{2,\text{inter}}$	2	3.5 Å
$G_{3,\text{inter}}$	5	5.5 Å
$G_{4,\text{inter}}$	5	6.4 Å

2. Training dataset

The training dataset is obtained from the iterative training approach explained in Section 2.2 of the main manuscript. We use three generations of data. Gen 0 is obtained from atomistic model simulations of crystalline RDX at ambient density (1.8 g/cm³) and temperatures of 250 K, 500 K, and 750 K. Gen 1 and Gen 2 datasets are obtained from CG model simulations with the same conditions as Gen 0. Gen 1 includes four sets of trajectories obtained from four models with varying edge and intramolecular node features outlined in Section S1. The Gen 2 dataset is obtained from two models that both utilize Gaussian functions for the input features, while varying the intramolecular node features. In total, Gen 0, Gen 1, and Gen 2 contained 300, 1073, and 540 frames, respectively, for a total of 1.6 million data points. We also note that some of the configurations obtained from the CG model simulations were not included in the dataset because the decoded atomistic model trajectory led to numerically unstable configurations.

Supplementary Figure 1 depicts an analysis of the configuration distributions within the training datasets. Here, the intramolecular configuration distributions are counts of the configurations normalized by the maximum number of counts. The intermolecular configuration distributions are counts of the intermolecular distances that are multiplied by the smoothing function (f_c) and normalized by the maximum count. We observe that the following configurations are under-represented in the dataset: intramolecular stretching configurations of ($d > 3.6$ Å and $d < 2.0$ Å),

angular configurations of ($\theta < 60^\circ$, $\theta > 160^\circ$), and intermolecular overlapping configurations of ($d < 2.5 \text{ \AA}$). Therefore, we utilize prior forces to avoid exploration of such configurations in the CG model simulations.



Supplementary Figure 1. Normalized distribution of configurations in the training dataset and the associated Gaussian functions. (a) Intramolecular distances, (b) intramolecular angles, (c) intermolecular distances. Black lines are the normalized counts of configurations. Red dashed lines are the Gaussian functions.

3. Prior forces

Simulations employing neural network force fields are prone to exploring configurational spaces that are significantly under-represented in the training data set. For these configurations, the extrapolation of a GNNFF is not guaranteed to be accurate and can lead to catastrophic failures of the model. To avoid these cases, we employ prior forces that restrict the exploration of such configurations following the approach by Wang et al.³ Specifically, we apply prior forces for the intramolecular stretching configurations ($d > 3.6 \text{ \AA}$ and $d < 2.0 \text{ \AA}$), angular configurations ($\theta < 60^\circ$, $\theta > 160^\circ$), and intermolecular overlapping configurations ($d < 2.5 \text{ \AA}$), as these configurations are not sampled sufficiently in the training set (see Supplementary Figure 1). The prior forces for these configurations are determined from the potential of mean force (PMF)

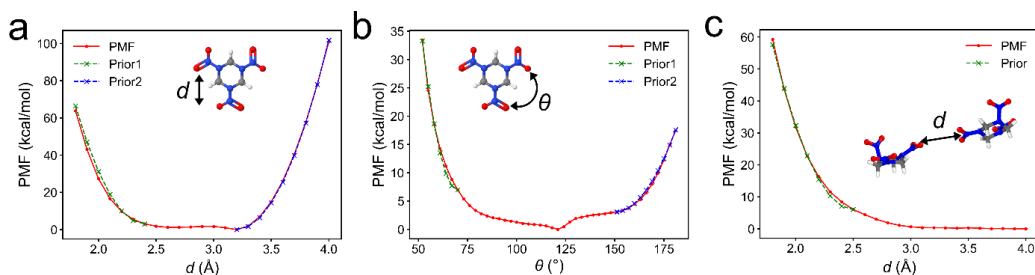
calculations obtained from steered molecular dynamics of the atomistic model, gas phase RDX using the Colvar package⁴ of LAMMPS.

For the intramolecular interaction prior forces, the nitro-triazine distance (Supplementary Figure 2a) and the nitro-triazine-nitro angle (Supplementary Figure 2b) of a single RDX molecule were used as the collective variables. For the intermolecular interactions (Supplementary Figure 2c), the distance between the nitro groups of two RDX molecules were used as the collective variable. In all cases, the PMF in the configurations of interest for prior forces can be described by harmonic potentials, $U = \frac{k}{2}(r - r_0)^2$. Therefore, we determine the prior forces as given by the following equations,

$$F_{\text{str}} = \begin{cases} 159.08(d - 3.6) & \text{for } d > 3.6 \text{ \AA} \\ 176.09(d - 2.0) & \text{for } d < 2.0 \text{ \AA} \end{cases} \quad (4)$$

$$F_{\text{ang}} = \begin{cases} 0.0812(\theta - 60) & \text{for } \theta < 60^\circ \\ 0.0151(\theta - 160) & \text{for } \theta > 160^\circ \end{cases} \quad (5)$$

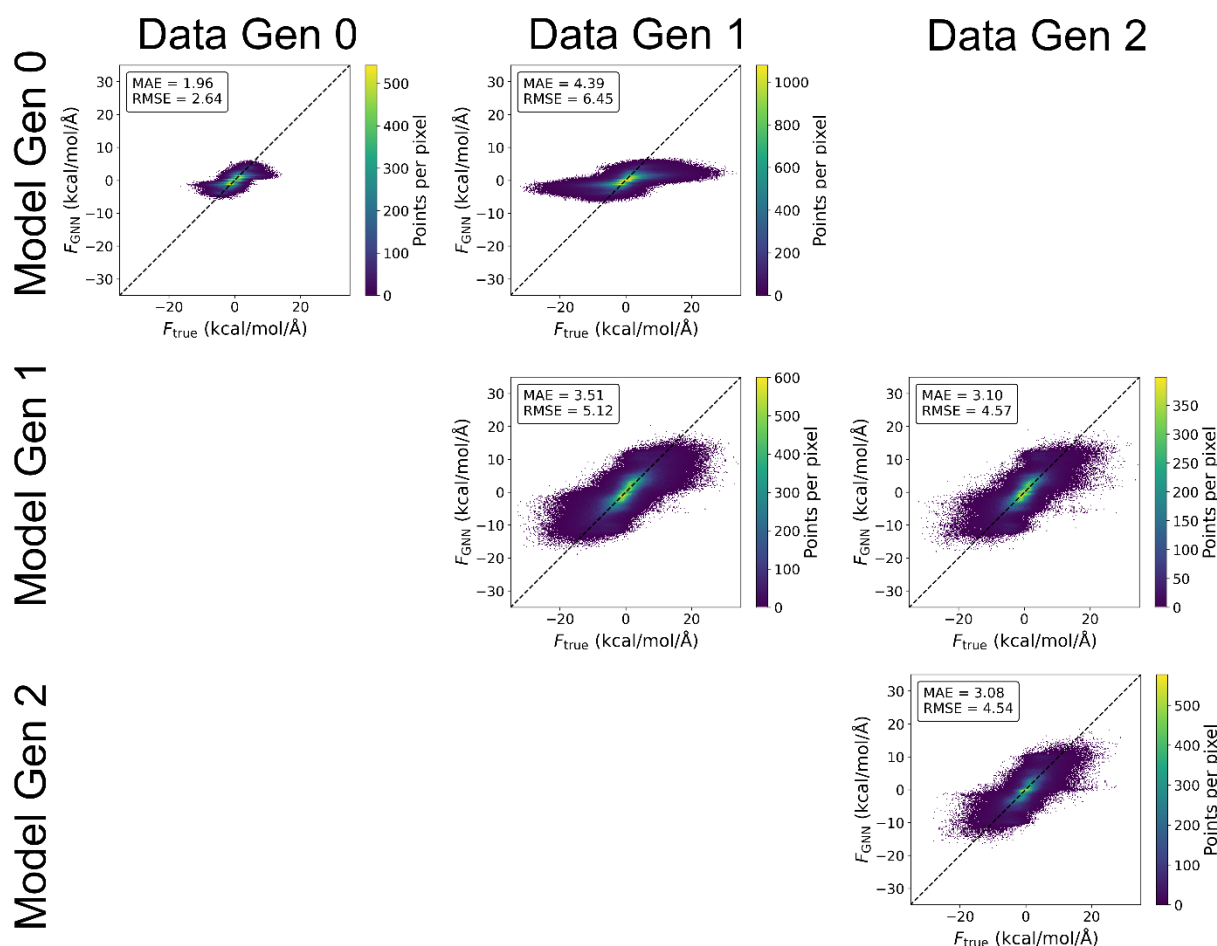
$$F_{\text{inter}} = 105.22(d - 2.5) \text{ for } d < 2.5 \text{ \AA} \quad (6)$$



Supplementary Figure 2. PMF from the atomistic model and corresponding prior energies for (a) intramolecular stretching, (b) intramolecular bending, and (c) intermolecular overlapping configurations.

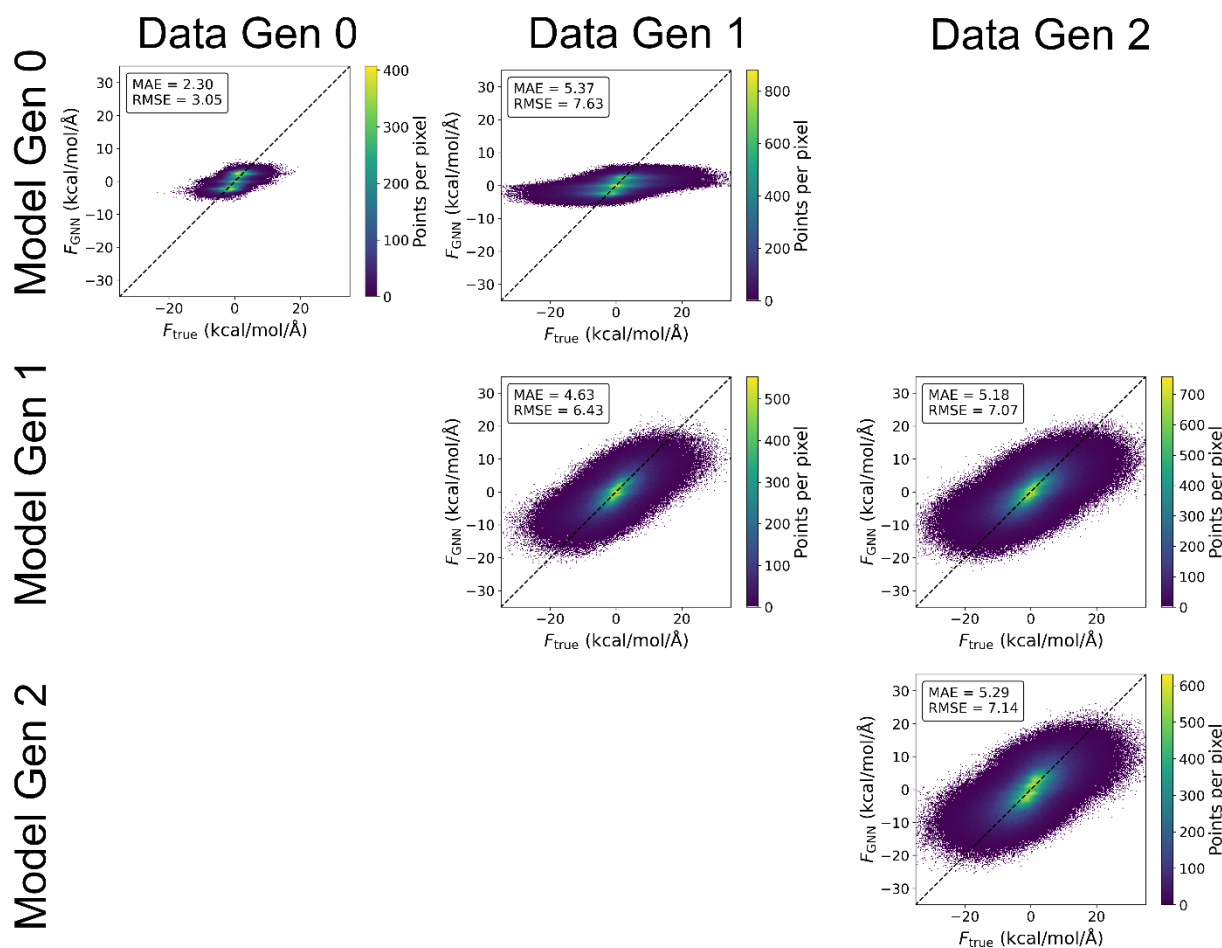
4. Generation-dependent inference

Generation-dependent inference performance of the intra- (Supplementary Figure 3) and inter-molecular (Supplementary Figure 4) forces are depicted. Here, the rows correspond to the generation dataset that the model was trained on. For example, second row plots correspond to inference of the model that was trained with all data from Gen 0 and Gen 1. The columns correspond to the generation dataset that the model was used for inference. For example, the middle column plots correspond to inference on Gen 0 and Gen 1 datasets with model trained on Gen 0 data (first row) and Gen 0 and Gen 1 data (second row). Only test set results are plotted as the model performs similarly for the train set.



Supplementary Figure 3. Generation dependent inference performance for intramolecular forces. The columns correspond to the dataset generation used for inference, while the rows correspond to the dataset used for training the model.

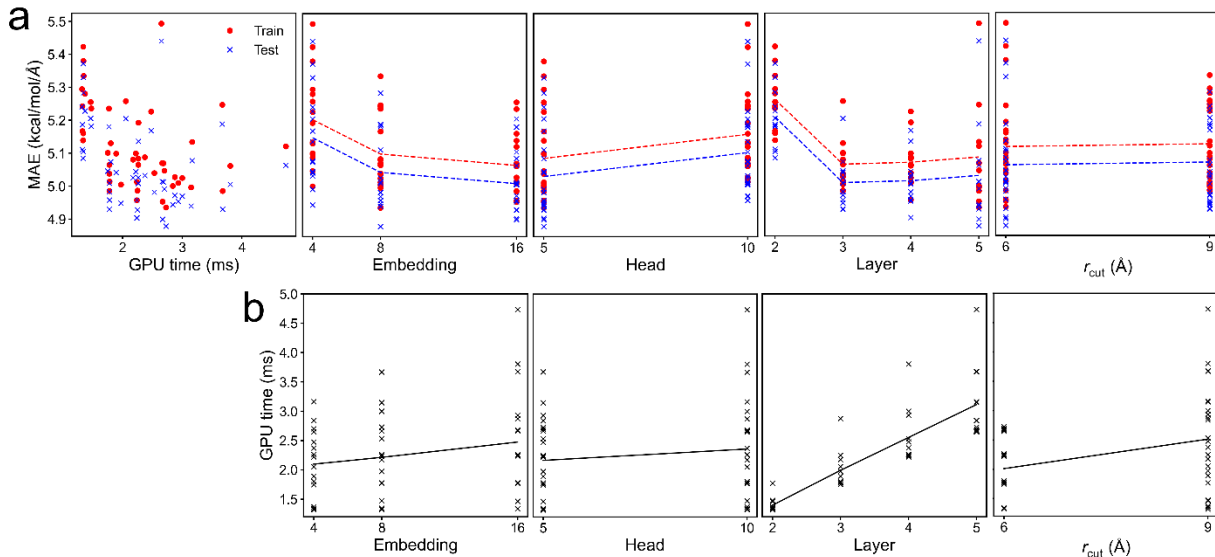
The results demonstrate that the Gen 0 force magnitudes are smaller compared to later generations. This is because the molecular configurations are from the crystalline configuration of the atomistic model simulations that do not deviate significantly from the local energy minima configurations. However, as the generations are added, configurations with large forces are added. The model trained on Gen 0 does not perform well on such configurations and predicts them to have small force magnitudes. However, the models trained with those iteratively added configurations are able to predict the large magnitude forces.



Supplementary Figure 4. Generation dependent inference performance for intermolecular forces. The columns correspond to the dataset generation used for inference while the rows correspond to the dataset used for training the model.

5. Hyperparameter dependent model performance

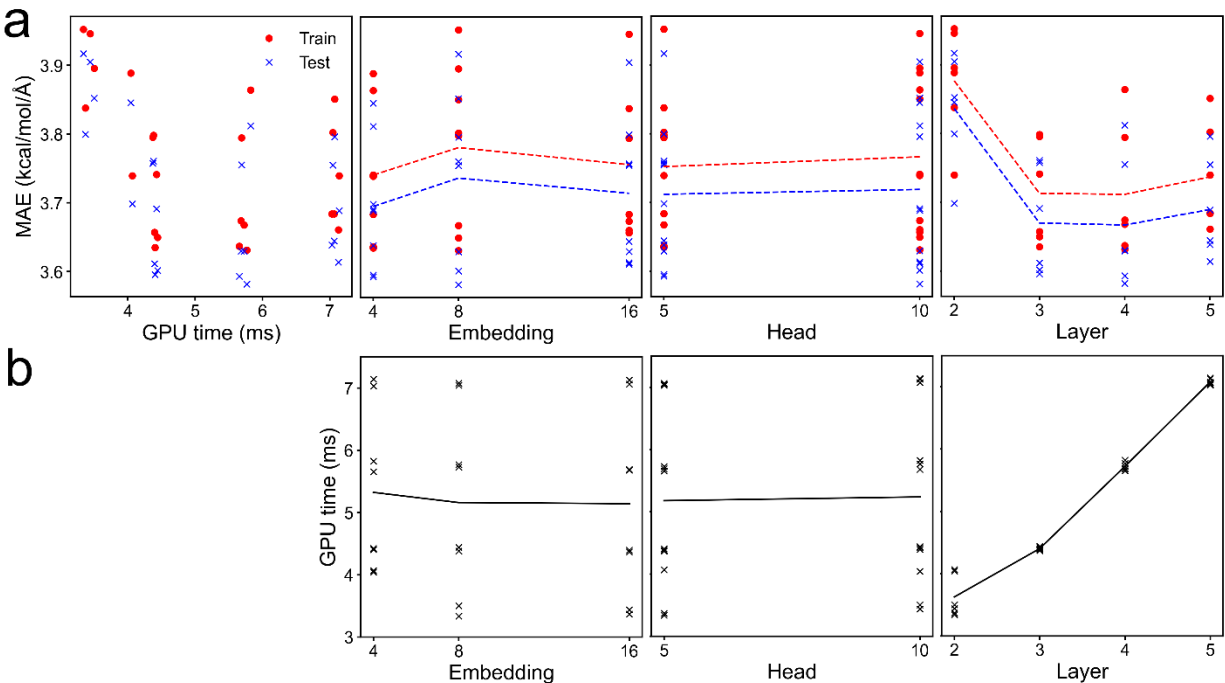
The accuracy and computational efficiency of the GNNFF was tested as a function of the hyperparameters. For the intermolecular forces, we have trained the model for varying number of embeddings ($E=4, 8, 16$), attention heads ($H=5, 10$), layers ($L=2, 3, 4, 5$) and cut-off distances ($r_{\text{cut}}=6 \text{ \AA}, 9 \text{ \AA}$). For each model, we computed the mean absolute error (MAE) and the GPU time per force calculation. The GPU time was measured by computing the forces on an ambient density crystal 10000 times with NVIDIA A100 GPU. In terms of computational efficiency, we observe that L (average increase in GPU time: 0.57 ms/L) was more significant compared to the E (0.031 ms/E), H (0.039 ms/H), and r_{cut} ($0.0168 \text{ ms/\AA}$). For accuracy, L led to significant improvement in MAE up to the third layer, but further increase did not improve the accuracy. Increase in E lead to better accuracy, although with diminishing returns. H and r_{cut} did not improve our model.



Supplementary Figure 5. Hyperparameter testing results for inter-molecular interactions. (a) Mean absolute error as function of hyperparameters and GPU time. (b) GPU time as function of hyperparameters. The scatters correspond to the values for each model while the dashed lines represent the average values for all models pertaining to the hyperparameter value.

For the intramolecular forces, the MAE and GPU times were measured for models with varying E , H , and L . Similar to the results from intermolecular force models, E (-0.015 ms/E) and H (0.012 ms/H) do not significantly affect the GPU time while L (1.15 ms/L) was important. The

accuracy was also improved significantly as layers were increased from 2 to 3 layers. However, increasing other hyperparameters did not lead to significant improvement in accuracy. For our model, we have chosen the model with the best accuracy in both inference and structural properties during simulations.

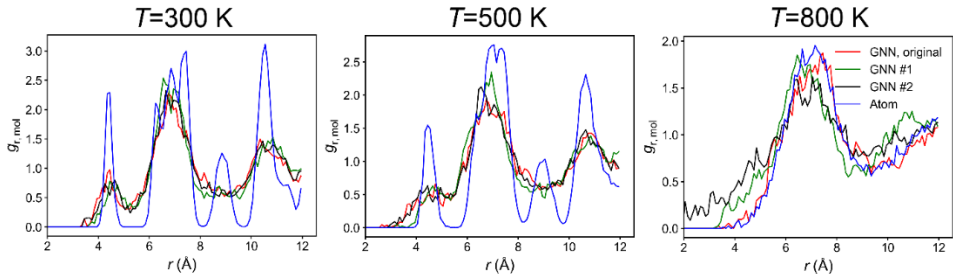


Supplementary Figure 6. Hyperparameter testing results for intra-molecular interactions. (a) Mean absolute error as function of hyperparameters and GPU time. (b) GPU time as function of hyperparameters. The scatters correspond to the values for each model while the dashed lines represent the average values for all models pertaining to the hyperparameter value.

For both intra- and inter-molecular interactions, we find the best performing model has ~10% reduced MAE compared to the worst model. In terms of computational efficiency, the differences were 114% (intra) and 259% (inter).

A difficulty in hyperparameter optimization for GNNFF is that the model with best inference accuracy does not necessarily lead to best structural properties during simulations, as reported by a recent paper⁵ that evaluated the performance of six atomistic GNNFFs on these criteria. The hyperparameter of our main manuscript was chosen after training models with various hyperparameters on Gen 0 data and picking the model with best overall structural and force

prediction accuracy. As additional results, we report the structural properties from two models with different hyperparameters trained on the entire 3 generations of data. Model 1 was chosen as one of the best performing models in terms of MAE and had the intermolecular hyperparameters of (16 E , 5 H , 3 L , 9 Å) and intramolecular hyperparameters of (8 E , 10 H , 4 L). Model 2 was the most computational efficient and had intermolecular hyperparameters of (8 E , 5 H , 2 L , 9 Å) and intramolecular hyperparameters of (4 E , 5 H , 2 L). As shown in Supplementary Figure 7, we obtain similar structural properties for T up to 500 K. However, the melted state at 800 K leads to significant differences, including appearance of peak around 11 Å (GNN #1) or appearance of molecules with distances less than 4 Å (GNN #2). One reason for such behavior is that these models were not trained with iteratively collected dataset. A future study for finding model with optimal balance of computational efficiency and accuracy may be valuable.



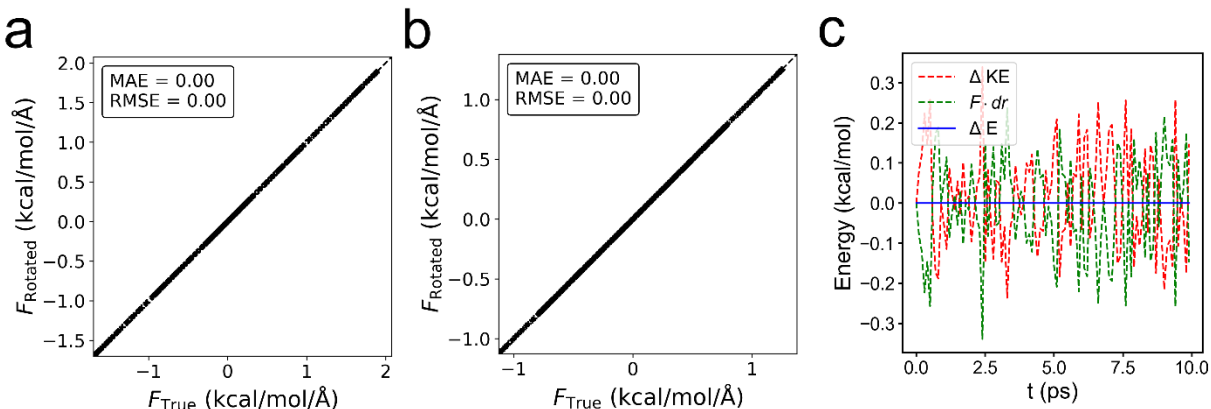
Supplementary Figure 7. $g_{r,mol}$ for atomistic and CG models with 3 different hyperparameters at ambient density and temperatures of 300, 500, and 800 K.

6. Rotational covariance and energy conservation

Similar to the approaches taken by previous direct covariant force models,^{6,7} we have ensured the rotational covariance of our model by assigning the direction vector of the forces to be covariant with system rotation. As a demonstration of the rotational covariance, we compare the forces from rotated systems versus forces rotated by the same operation. Specifically, we computed the forces on ambient crystal configuration of CG RDX. Then, we picked one molecule and rotated the system around the center-of-mass of the molecule. The forces on the chosen molecule obtained from the rotated systems are compared with the forces from original coordinates that are rotated in the same manner. Parity plots of the intra- and inter-molecular forces are plotted

on Supplementary Figure 8a (intra) and Supplementary Figure 8b (inter). The comparison shows that our model ensures rotational covariance of the forces.

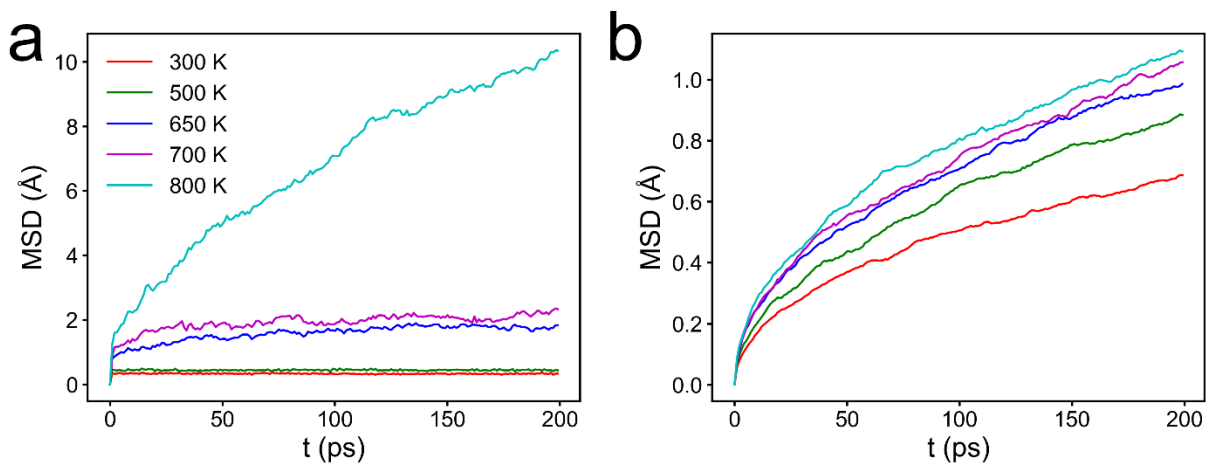
As our model uses direct covariant force approach, our model does not output the potential energies. However, we have made several design choices to ensure energy conservation. The intermolecular forces decay to 0 at the cutoff distance as we multiply the $u_{ij,k}$ with the f_c , a smoothly decaying function equal to 0 at the cutoff distance. This is equivalent to multiplying the forces by f_c , ensuring that there is no discontinuity in the forces at the cutoff. The prior forces are also set to 0 at their respective cutoff distances. Finally, the intermolecular forces are formulated to obey Newton’s third law of motion by assigning the mean value of the output from the interacting particles as explained in the Methods section. As a test of energy conservation, we have run NVE simulations and measured the change in the kinetic energy of the system and the work measured by the change in position of the particles multiplied by their forces on each time step. The results in Supplementary Figure 8c show that the total energy is conserved.



Supplementary Figure 8. (a, b) Comparison of rotated forces (F_{Rotated}) versus forces on rotated systems (F_{True}) for intra- (a) and inter-molecular (b) forces. (c) Change in total kinetic energy and work over time from NVE simulation.

7. Dynamic properties

The mean squared displacements (MSD) of the center-of-mass of molecules were measured for bulk system atomistic and CG model simulations. As reported for previous CG models,^{8–11} there is significant difference in the dynamic properties that is not resolved by our CG model.



Supplementary Figure 9. MSD from bulk simulations at ambient densities and varying temperatures for the (a) atomistic and (b) CG models.

Supplementary References

- ¹ J.D. Moore, B.C. Barnes, S. Izvekov, M. Lísal, M.S. Sellers, D.C.E. Taylor, and J.K. Brennan, A coarse-grain force field for RDX: Density dependent and energy conserving. *J. Chem. Phys.* **144**(10), 104501 (2016).
- ² J. Behler, and M. Parrinello, Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**(14), 146401 (2007).
- ³ J. Wang, S. Olsson, C. Wehmeyer, A. Pérez, N.E. Charron, G. De Fabritiis, F. Noé, and C. Clementi. Machine Learning of Coarse-Grained Molecular Dynamics Force Fields. *ACS Cent. Sci.* **5**(5), 755–767 (2019).
- ⁴ G. Fiorin, M.L. Klein, and J. Hénin. Using collective variables to drive molecular dynamics simulations. *Mol. Phys.* **111**(22–23), 3345–3362 (2013).
- ⁵ V. Bihani, S. Mannan, U. Pratiush, T. Du, Z. Chen, S. Miret, M. Micoulaut, M.M. Smedskjaer, S. Ranu, and N.M.A. Krishnan. EGraFFBench: evaluation of equivariant graph neural network force fields for atomistic simulations. *Digital Discovery* **3**(4), 759–768 (2024).
- ⁶ C.W. Park, M. Kornbluth, J. Vandermause, C. Wolverton, B. Kozinsky, and J.P. Mailoa, Accurate and scalable graph neural network force field and molecular dynamics with direct force architecture. *npj Comput. Mater.* 2021 7:1 7(1), 1–9 (2021).
- ⁷ J.P. Mailoa, M. Kornbluth, S. Batzner, G. Samsonidze, S.T. Lam, J. Vandermause, C. Ablitt, N. Molinari, and B. Kozinsky. A fast neural network approach for direct covariant forces prediction in complex multi-element extended systems. *Nat. Mach. Intell.* **1**(10), 471–479 (2019).
- ⁸ R. Baron, A.H. de Vries, P.H. Hünenberger, and W.F. van Gunsteren. Configurational Entropies of Lipids in Pure and Mixed Bilayers from Atomic-Level and Coarse-Grained Molecular Dynamics Simulations. *J. Phys. Chem. B* **110**(31), 15602–15614 (2006).
- ⁹ S.J. Marrink, H.J. Risselada, S. Yefimov, D.P. Tieleman, and A.H. de Vries. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **111**(27), 7812–7824 (2007).

¹⁰ D. Roy, N. Patel, S. Conte, and M. Maroncelli. Dynamics in an Idealized Ionic Liquid Model. *J. Phys. Chem. B* **114**(25), 8410–8424 (2010).

¹¹ J. Ruza, W. Wang, D. Schwalbe-Koda, S. Axelrod, W.H. Harris, and R. Gómez-Bombarelli. Temperature-transferable coarse-graining of ionic liquids with dual graph convolutional neural networks. *J. Chem. Phys.* **153**(16), 164501 (2020).