

Machine Learning Assisted Hybrid ReaxFF Simulations

Dundar E. Yilmaz, William Hunter Woodward and Adri C. T. van Duin*

Department of Mechanical Engineering, The Pennsylvania State University, University Park, PA 16802

AUTHOR INFORMATION

Corresponding Author

* duy42@psu.edu

Abstract

We have developed a Machine Learning Assisted Hybrid ReaxFF Simulation method ("Hybrid/Reax"), which alternates reactive and non-reactive molecular dynamics simulations with the assistance of machine learning (ML) models to simulate phenomena that require longer time scales and/or larger systems than are typically accessible to ReaxFF. Hybrid/Reax uses a specialized tracking tool during the reactive simulations to further accelerate chemical reactions. Non-reactive simulations are used to equilibrate the system after the reactive simulations stage. ML models are used between reactive and non-reactive stages to predict non-reactive force field parameters of the system based on the updated bond topology. Hybrid/Reax simulation cycles can be continued until the desired chemical reactions are observed. As a case study, this method was used to study the crosslinking of a polyethylene matrix analogue (decane) with crosslinking agent dicumyl-peroxide. We were able to run relatively long simulations (>20 million MD steps) on a small test system (4660 atoms) to simulate crosslinking reactions of PE in the presence of dicumyl peroxide. Starting with 80 PE molecules, more than half of them crosslinked by the end of the Hybrid/Reax cycles on a single Xeon processor in under 48 hours. This simulation would take approximately one month if run with pure ReaxFF MD on the same machine.

KEYWORDS: ReaxFF, OPLS, Molecular Dynamics, Machine Learning

Introduction

Molecular Dynamics (MD) is a widely adopted method to study diverse molecular systems at an atomistic level, ranging from biophysics to chemistry to materials science,¹⁻³ by calculating the potential energy surface (PES) of the system to predict various observables. The PES of the system can be accurately calculated with Quantum Mechanical (QM) methods⁴ or efficiently with classical empirical force fields⁵⁻⁸ which are simplified descriptions of interactions between atoms. While QM models provide highly accurate results, they are of limited applicability in terms of spatial and temporal scales due to their high computational expense. In its most conventional form (i.e. classical MD), MD simulations usually employ parameterized force fields that enable the study of large systems, with millions to billions of degrees of freedom, using atomistic models that are computationally tractable and scalable on large computer systems. MD simulations can thus be used to address a wide range of challenging scientific and engineering problems, which would have been very hard using experimental techniques or computationally intractable using a full QM approach. Example applications of MD range from two dimensional materials,⁹⁻¹¹ polymeric systems,¹²⁻¹⁷ ferroelectric systems¹⁸⁻¹⁹ and the computational design of new battery materials.²⁰⁻²³

Modern force fields can be grouped into two categories, reactive and non-reactive.⁶ Reactive force fields allow the breaking and creation of chemical bonds during the simulation while non-reactive force fields use a fixed bond scheme to provide more efficient computation. Balancing accuracy with computational efficiency in MD simulations depends on the force field used to describe the PES of the system. ReaxFF is a reactive force field, originally developed to simulate hydrocarbons and later extended to a wide range of materials.^{11-14, 16, 18-19, 24-27} Ideally, reactive force fields would be used in MD simulations to directly study phenomena where chemical reactions play crucial roles. However, chemical reactions, for example the hydrolysis and crosslinking of polymers,^{12, 17, 28-29} hydrocarbon oxidation,³⁰ chemisorption,³¹ defect formation¹¹ and diffusion,³² and growth processes,⁹ typically require long reaction times – beyond microseconds - that are often infeasible with reactive force fields, which typically are computationally limited to tens of nanosecond time scales. Accelerating reactive MD simulations can facilitate observation of these reactions within a reasonable simulation time. In this regard, there have been efforts such as developing custom algorithms to efficiently utilize the latest GPU hardware,^{12, 33-34} and introducing probability space to dynamical simulations.³⁵⁻³⁷

Non-reactive force fields often use a fixed bond scheme approach to improve computational efficiency by calculating bonded interactions over a predetermined neighbor list. For instance Optimized Potential for Liquid Simulations (OPLS)⁸—a non-reactive force field designed to simulate organic liquids—uses this approach and runs approximately 50 times faster compared to ReaxFF with the caveat of being unable to handle chemical reactions. Thus, OPLS can access to long simulation times, beyond microseconds, for large systems.

Machine learning (ML),³⁸⁻³⁹ a subset of artificial intelligence (AI) methods, has recently become popular for use in MD simulations.⁴⁰ ML can be defined as an effort to give computers the ability to learn without being explicitly programmed. ML methods have been widely used as a classification tool in many disciplines.³⁸ In computational materials science, it has been used in learning relations between carefully designed descriptors and their corresponding observables (binding energies, diffusion speeds, reaction efficiency, etc.).⁴¹⁻⁴⁵ A more fundamental area in which ML methods are employed in the field of computational materials science is prediction of the PES.^{41-42, 44, 46-47} ML-predicted PES's are often called ML force fields. In recent years, with advances in computer hardware and performance, they have gained popularity over DFT-based methods. However, ML force fields are, in general, at least an order of magnitude slower than empirical force fields.⁴⁰

In this paper, we report the development of a method to accelerate reactive MD simulations: running reactive force field steps and then non-reactive steps in a loop, with an ML model gluing these two methods together. We dubbed this method ML assisted Hybrid Reax Simulations (Hybrid/Reax). As a case study, we used Hybrid/Reax to simulate polyethylene (PE) crosslinking chemistry, which involved the creation of carbon-carbon bonds between decane molecules using the common peroxide crosslinking agent dicumyl-peroxide (DCP). The set of reactions which leads to the formation of interconnected PE chains involves breaking the O-O bond of the DCP molecule to form two radicals, said radicals abstracting two H atoms from two different PE chains, and the resulting -CH- radicals forming a crosslink between the two chains. This process requires both the chemistry of bond cleavage/formation and the migration of radicals into the vicinity of hydrogen atoms or other radicals, and thus the computational expense via traditional methods would be high. In this case study, Hybrid/Reax simulates these reactions using a reactive and a non-reactive force field in a hybrid fashion to take advantage of the time scales of the non-reactive force field without losing the ability to carry out chemical reactions.

Method

The Hybrid/Reax simulation cycle composed of three stages (Figure 1). After relaxing the system with a non-reactive MD simulation, the Hybrid/Reax simulation cycles starts with the first stage, where a non-reactive MD simulation is run to relax the system, and continues with the second stage in which the reactive MD simulation along with the tracking tool was employed to conduct chemical reactions based on preset conditions.⁴⁸ In the final stage, positions and bonding information are transferred to the ML model to predict non-reactive force field parameters and charges. After prediction of non-reactive force field parameters and charges, the Hybrid/Reax returns to the first stage and continues these cycles until desired chemical reactions are completed. In the following, we will describe these stages in detail.

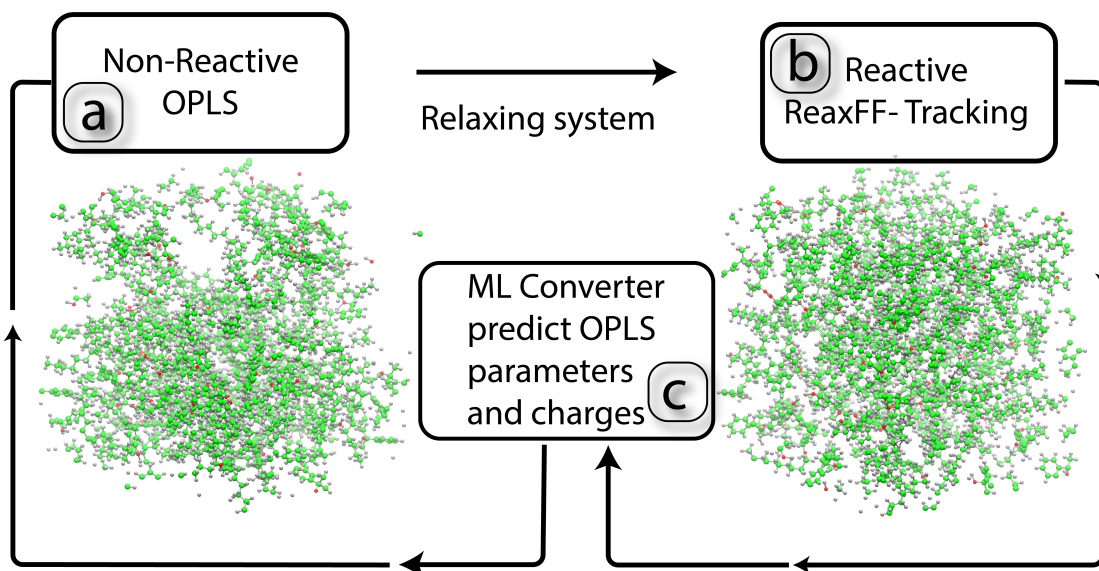


Figure 1. Flow chart of ML assisted Hybrid Reax (Hybrid/Reax) simulation. a) Non-reactive MD simulation relaxes the system. b) Reactive MD simulation with tracking tool conducts chemical reactions and c) a ML based converter predicts OPLS parameters and charges based on the bond topology.

Non-Reactive MD Simulations

We used the Optimized Potential for Liquid Simulations (OPLS) force field for non-reactive MD simulations.⁸ OPLS was developed for the simulations of organic liquids.⁸ The energy of the system consists of non-bonded and bonded interactions. Coulomb and Lennard Jones type potentials were defined for non-bonded interactions:

$$\text{Eq. 1.} \quad E = \sum_i^N \sum_j^N \left[\frac{q_i q_j e^2}{r_{ij}} + 4\epsilon_{ij} \left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^6}{r_{ij}^6} \right) \right] f_{ij}$$

In Eq. 1, q_i are charges, and ϵ_i and σ_i are Lennard Jones parameters for each atom. The geometric mean is used as mixing rule to calculate ϵ_{ij} and σ_{ij} . Finally, f_{ij} is a scaling factor so that we can use the same parameters for inter- and intra-molecular interactions. Bond bending and stretching interactions are defined as:

$$\text{Eq. 2. } E_{bonded} = \sum_{bonds} K_r (r - r_0)^2 + \sum_{angles} K_\theta (\theta - \theta_0)^2$$

Where r_0 and K_r are equilibrium bond lengths and bond stretching coefficients, and θ_0 and K_θ are equilibrium bond angle and bending coefficients, respectively. Torsional energy is defined as a fourth order Fourier expansion over dihedral angle ϕ_i :

$$\text{Eq. 3. } E_{torsion} = \sum_i \frac{C_1^i}{2} [1 + \cos(\phi_i)] + \frac{C_2^i}{2} [1 - \cos(2\phi_i)] + \frac{C_3^i}{2} [1 + \cos(3\phi_i)] + \frac{C_4^i}{2} [1 - \cos(4\phi_i)]$$

Where $C_{j=1,4}^i$ are Fourier coefficients.⁴⁹ The OPLS force field may assign different atom types to a particular element based on the bond topology. For instance the parameters of a carbon atom in a benzene ring differ from those of a carbon atom in an alkane group; similarly, the parameters of a carbon-carbon bond in a butane molecule may differ from those of a carbon-carbon bond in a carboxyl group.⁸ Furthermore, the OPLS force field uses a fixed bond scheme which does not allow bond forming/breaking reactions. Thus, initial bond topology along with coordinates of atoms should be supplied at the beginning of the simulation.

ReaxFF with Tracking Tool

For reactive MD simulations, we employed ReaxFF which is a bond-order based potential with energy described by bonded and non-bonded interactions:

$$\text{Eq. 4. } E = E_{self} + E_{Coulomb} + E_{Van\ der\ Waals} + E_{bond} + E_{angle} + E_{torsion} + E_{conjugation} + E_{H-bond} + E_{lone-pair} + E_{over} + E_{under}$$

In Eq. 4, E_{self} , $E_{Coulomb}$ and $E_{Van\ der\ Waals}$ represent non-bonded interactions and the rest are all functions of bond order which is a metric of chemical bonding between two interacting atoms. Bond orders are calculated at every step and corrected for over- and under-coordination cases. This enables ReaxFF to simulate chemical reactions during the simulation. The relatively complex functional form of ReaxFF allows transferability with the caveat of having a larger parameter set. A typical ReaxFF force field for a binary system contains around 40 parameters to

be trained with experimental data or higher-order calculations. Interested readers may consult the literature for detailed information.⁶

Experimental time scales for certain chemical reactions, especially those with high barrier energies or involving intermediary steps, may range from seconds to hours, which are inaccessible with MD simulations. Methods such as the ‘bond-boost’ and the “bond restraint” were applied to ReaxFF simulations to observe certain chemical reactions. We employed a recently developed “tracking tool” which tracks atom positions during the MD simulation and identifies early stages of a chemical reaction to apply appropriate restraints to drive the system across the reaction barrier.⁴⁸

Predicting OPLS Parameters and Charges with ML Model

In general, functional forms of force fields which define the PES of a system contain specific coefficients (e.g. force field parameters for each atom, bond, angle, etc.) to be fitted against a database are created by compiling experimental data and density functional theory calculations.⁵⁰⁻⁵¹ The desire to use the same force field for a wide range of materials leaves two choices: designing a fairly complex force field with a large number of parameters to capture different materials or assigning different atom types for different materials (i.e. individual ligands or molecules). While developers of ReaxFF follow the former route, as mentioned earlier OPLS force field parameters for bonded interactions depend on bond topology of the system. This choice along with fixed bonds accelerates the computational speed of OPLS approximately 50 times compared to ReaxFF.

In the second stage we expect chemical reactions to occur during reactive MD simulations with tracking tool. Hence the bond topology of the system will be different at the end of this stage. To continue the simulation after the reactive MD stage, the OPLS parameters’ dependence on bond topology necessitates identification of local bond topology to assign proper parameters for each interaction. Thus, we developed ML models to predict OPLS parameters and charges based on the local bond topology. One of the key factors to the success of ML models is designing a feature vector/descriptor which represents distinctive features of each data point in the database. Since the OPLS force field defines atom types based on the neighbors of each atom, a descriptor based on neighbors should work best. In *Supp. Fig. 1* we present a sample molecule and provide descriptors based on this molecule. For instance, a descriptor for a pairwise parameter would be

a vector with three components, the mass of the particular atom, its number of neighbors, and the total mass of its neighbors (*Supp. Table 1*). This descriptor should identify a particular atom with its local bond topology. Bond stretching (*Supp. Table 2*), bond angle (*Supp. Table 3*) and dihedral parameters (*Supp. Table 4*) were designed with this logic.

Since electrostatic interactions are long ranged, we designed a descriptor to not be limited with first neighbors and included second and third neighbors of each atom. Thus, the first component of descriptor for predicting charges would be the mass of a particular atom followed by the masses of its first, second, and third neighbors. In this case we fixed the length of the descriptor vector to 16 components and padded with zeros where needed.

Collecting Data

The quality of the training data is another key factor for the success of the ML model. Thus we aimed to compile relevant small molecules as much as possible to construct the training database starting with those shipped with Biochemical and Organic Simulation System (BOSS) software.⁵² To extend the database, a Python script was created to search and scrape molecule information from the ChemSpider database,⁵³ a website that hosts a chemical information database of more than 77 million molecules. Keywords such as “di-cumyl,” “peroxide,” and “alkane” were used to find relevant molecules. SMILES codes of around 10,000 individual molecules were downloaded, and OPLS parameters and charges for each of these molecules were generated using BOSS. For each molecule, a data file containing positions and OPLS parameters in LAMMPS format was created. Next, descriptors were created for each type of interaction (pairwise, bond-stretching, bond-bending, and dihedral), and corresponding parameters were matched to construct the ML training dataset. BOSS software employs the CM1A⁵⁴ method to calculate charges, but we opted to train a ML model to predict charges along with force field parameters in this project.

Predicting Charges

We trained a Dense Neural Network (DNN) of one input and five hidden layers with 512 artificial neurons on each layer using TensorFlow suite⁵⁵ to predict charges. Rectifier Linear Unit was used as the activation function for artificial neurons. The training dataset contains 83,985 individual data points and is split 80/20 for training and validation, respectively. The DNN converged to a loss of 0.000778 and a mean absolute error of 0.0164 after 200 epochs (Figure

2a). Charges predicted by the DNN are compared to those calculated with the CM1A method in Figure 2b. The performance of DNN in predicting charges of individual atoms was reasonably good. However, when predicting the total charge of a molecule, since DNN has no information on the total charge state, this information should be included by rescaling the charges (e.g. total charge of the molecule equals zero).

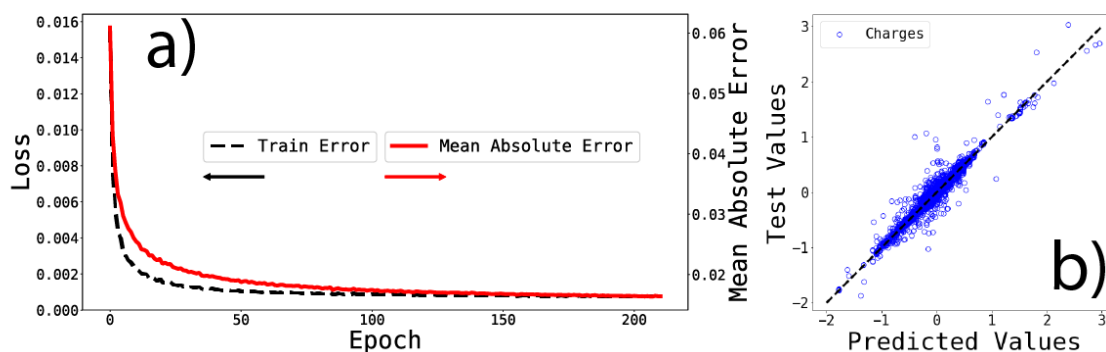


Figure 2. a) Training the DNN for 200 epochs for predicting charges and b) validation of predicted charges.

Predicting Pairwise Parameters

Unlike charges, analyzing the database shows that pairwise parameters, ϵ and σ of Eq. 4, were not distributed evenly as displayed in Figure 3 a and b. This distribution renders regression models inefficient: instead, classifier-type models should perform better. Thus, unique values of ϵ and σ in the training dataset were selected as class labels and a DNN with one input layer, two hidden layers, and one output layer were trained. The dimensions of the output layers were 22 and 19 for ϵ and σ , respectively. The dataset split 80/20 as training and validation sets, respectively. After 200 epochs both models were converged to 0.97 and 0.96 accuracy values respectively. Figure 3c and d compares predicted pairwise parameters with the ones generated by BOSS.

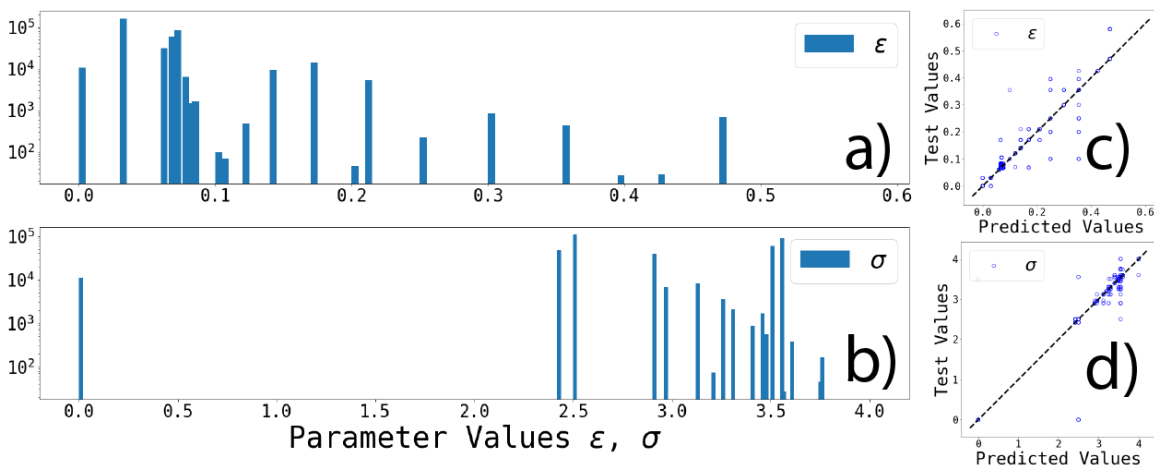


Figure 3 Histogram plots of number of occurrences of parameter values ϵ (a) and σ (b) in the training set. Validation of pairwise parameters ϵ (c) and σ (d) predicted with DNN.

Predicting Bond Stretching Parameters

Bonded interactions in OPLS force field consist of bond stretching and bond bending terms (Eq. 2). There are two parameters for bond stretching interaction, r_0 and K_r : equilibrium bond length and bond stretching coefficients. There were over 300,000 bond type descriptors and their corresponding parameters in the training dataset; analyzing the training data set shows the correlations between the mass of an atom in the bond with the equilibrium bond length, r_0 and also between equilibrium bond length and the bond stretching coefficient K_r (Figure 4).

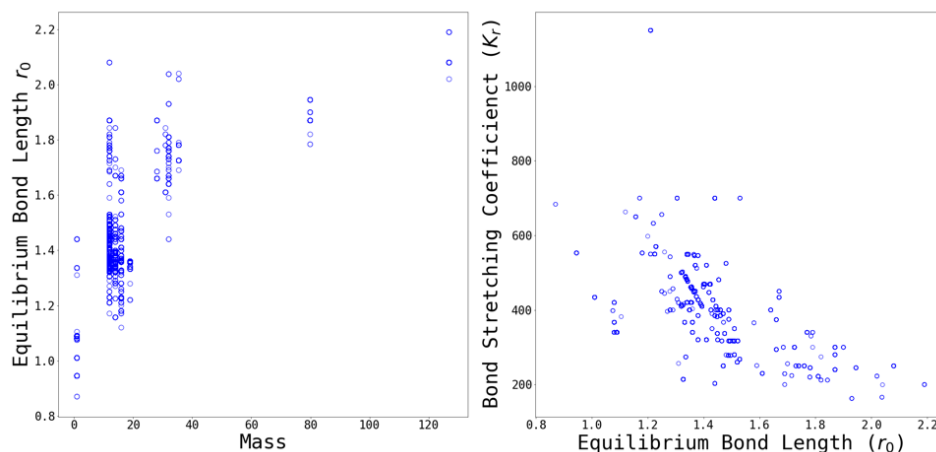


Figure 4. Correlation between masses and bond stretching parameters.

To take advantage of these correlations, a two-step process was designed to predict equilibrium bond lengths and bond stretching parameters. In the first step, a DNN was trained with bond

descriptors (*Supp. Table 2*) to predict equilibrium bond lengths. Then, in the second step, those equilibrium bond lengths were appended to the bond descriptors to train a similar DNN to predict bond stretching coefficients. Figure 5 summarizes training the DNN for bond stretching parameters. Due to correlation between equilibrium bond lengths and masses of bond atoms, the DNN for equilibrium bond lengths quickly converged in under 100 iterations/epochs (Figure 5a). Conversely, the DNN trained to predict bond stretching parameters converged in 1000 epochs (Figure 5b). The training database was split 80/20 for training and verification, respectively (Figure 5c and d).

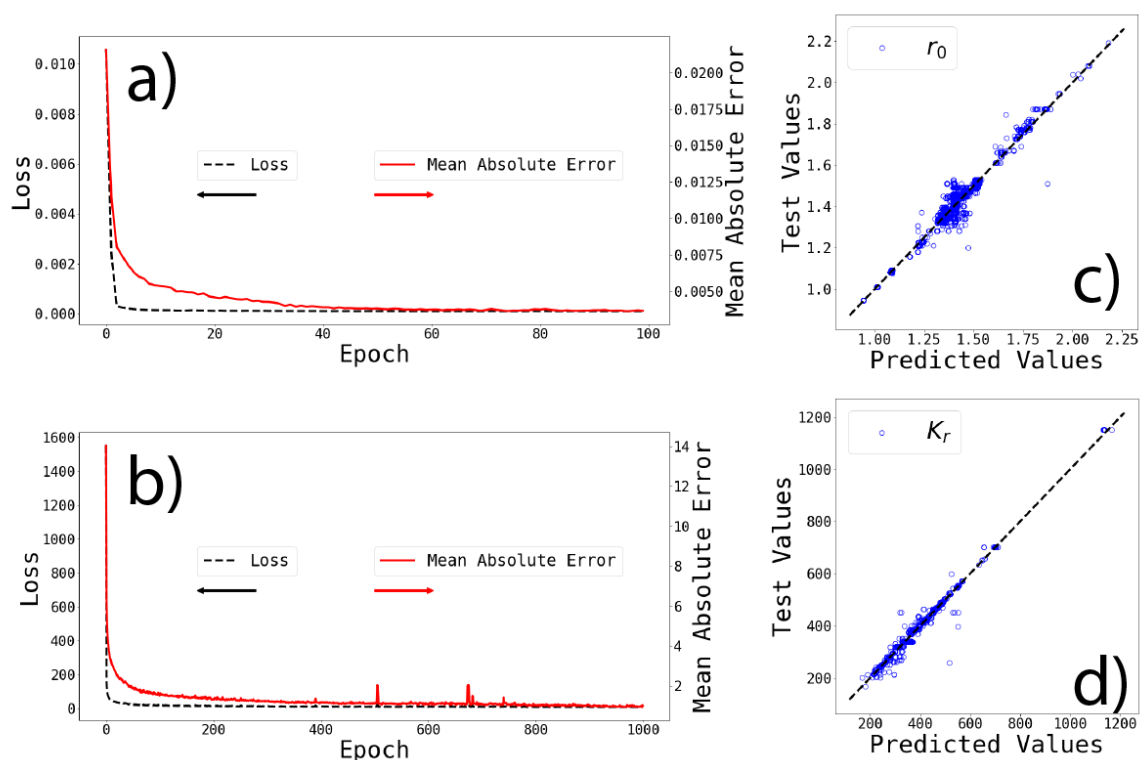


Figure 5. Predicting bond stretching parameters. Loss and mean absolute error of Dense Neural Network (DNN) trained for equilibrium bond length (a) and bond stretching coefficients (b) and verification of trained DNN for equilibrium bond length (c) and bond bend stretching coefficients (d).

Predicting Bond Bending Parameters

Bond bending energy in OPLS has two terms (Eq. 2)—equilibrium bond angle and bond bending coefficient—which are defined by atom types and the neighbors in the bond angle. The descriptor for bond bending parameters was designed as follows: The first three components are masses of angle atoms, the next three components are number of neighbors of angle atoms, and finally the last three components are the sum of masses of neighbors of angle atoms (*Supp. Table*

3Error! Reference source not found.). There were over 500,000 bond angle descriptors and their corresponding bond bending parameters in the training database. DNN's were trained to predict bond bending parameters and both DNN's were converged in 250 epochs (Figure 6 a and b). The training database was split 80/20 for training and verification, respectively. Comparison of predicted values with values generated by BOSS are presented in Figure 6 c and d.

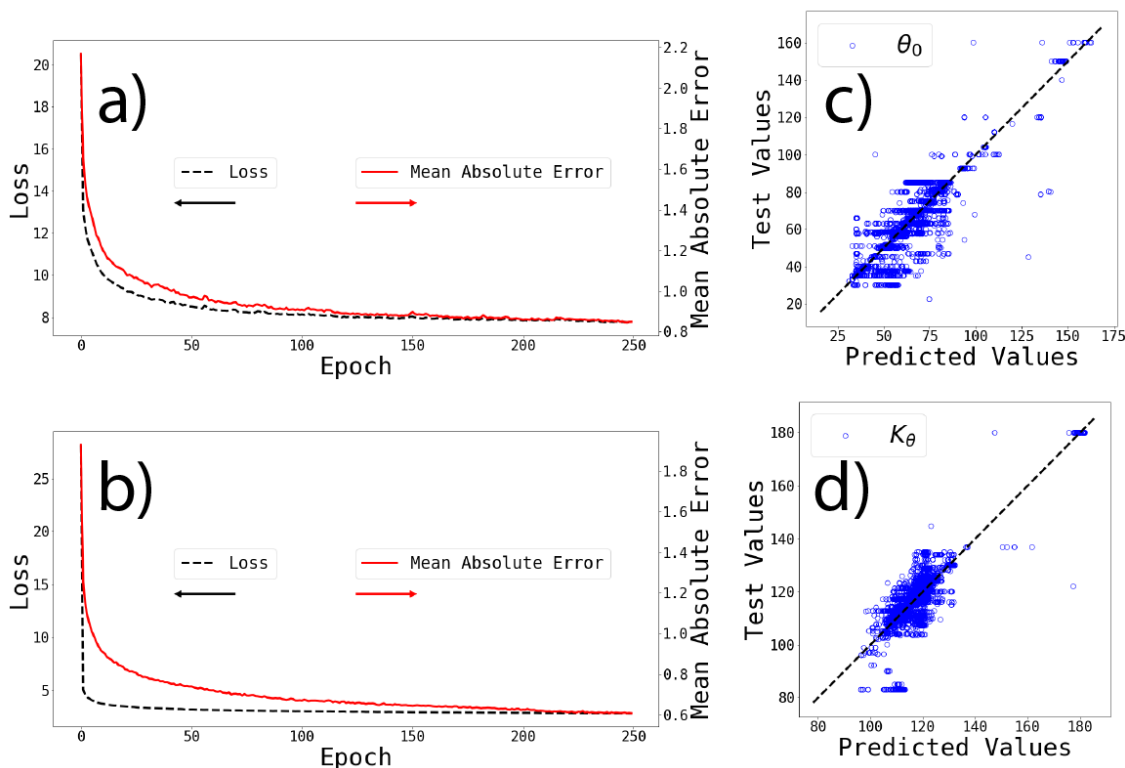


Figure 6. Predicting bond bending parameters. Loss and mean absolute error of Dense Neural Network (DNN) trained for a) equilibrium bond angle and b) bond bending coefficients and verification of trained DNN for c) equilibrium bond angle and d) bond bending coefficients.

Predicting Dihedral Parameters

Like pairwise parameters, dihedral parameters were also not distributed evenly, thus instead of regression, DNN classifiers were used. The Dihedral energy function has four parameters. For each parameter in Eq. 3 a separate DNN classifier was trained. The descriptor vectors for dihedral angle parameters were designed as follows: The first four components are masses of atoms in the dihedral angle, the next four components are number of atoms bonded to each dihedral atom, and the last four components are the total mass of bonded atoms to each dihedral atom (*Supp. Table 4*). There were close to 1,000,000 dihedral descriptors with corresponding dihedral angle parameters in the training database.

The number of unique values found in the training database for each parameter C_i determined the dimension of the output layer. There were 3 hidden layers with 512 artificial neurons in each DNN. The training of DNN's for dihedral angle parameters are summarized in *Supp. Fig. 2*. All four DNN's were converged to >99% accuracy. The training database was split into 80/20 for training and verification, respectively. Overlaid histogram plots of predicted values and corresponding test values of parameters are presented in *Supp. Fig. 3*.

Case Study: Crosslinking of Polyethylene

Polymer simulations require intensive computational resources due to the necessity of a relatively large simulation box. ReaxFF has been previously utilized to investigate certain chemical reactions in polymer systems.¹³⁻¹⁷ Here we present the dynamical simulation of decane/peroxide chemical reactivity as a case study to demonstrate effectiveness of the newly developed ML Assisted Hybrid simulation technique to model complex polymer systems (e.g. polyethylene crosslinking). We build the system by randomly placing 80 decane ($C_{10}H_{22}$) and 50 dicumyl-peroxide ($C_{18}H_{22}O_2$, DCP) molecules in a simulation box of length 20 nm each dimension. In total there were 4660 atoms in the system. We used Polybuild⁵⁶, an in-house tool, to build the systems. Before starting the hybrid simulation cycle, we gradually compressed the simulation box to 4.0 nm in each dimension over 100 ps to achieve final density of 0.65 g/cm³.

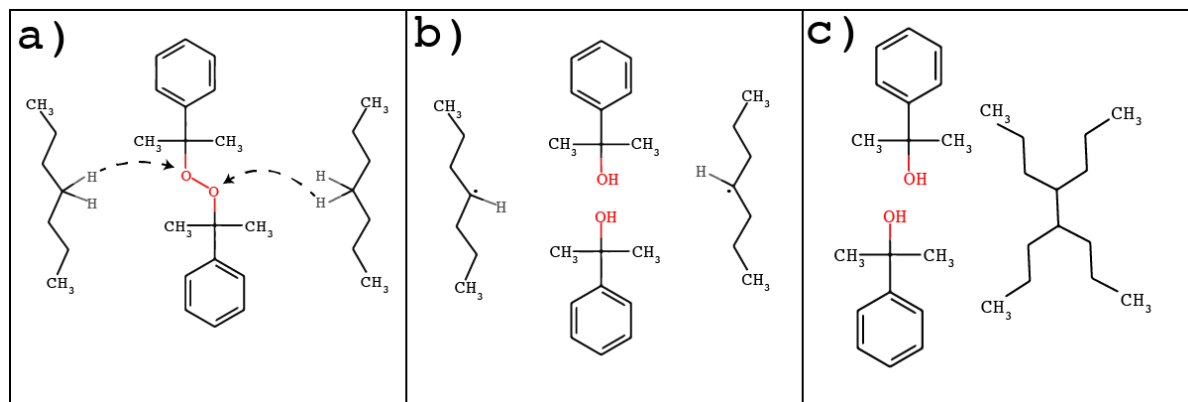


Figure 7. Schematic diagram for the basic crosslinking reaction of decane (polyethylene) with dicumyl-peroxide. a) Dicumyl peroxide extracts hydrogens from two CH_2 monomers and b) forms two cumyl alcohols and PE chains with radical sites. c) Two PE chains form a crosslink via radical sites.

In a recent paper, Akbarian *et al.* studied the crosslinking of decane with dicumyl-peroxide using ReaxFF.¹⁷ The crosslinking reaction starts with breaking up the O-O bond, then O atoms of the peroxide radicals capture H atoms from two alkane molecules resulting radical sites on each PE, followed by the formation of a crosslink between two decane molecules (Figure 7). This reaction often involves intermediate steps, and its observation requires relatively long simulation times. Thus, we setup the “tracking tool” mentioned earlier, to track molecules in order to identify configurations of molecules that have the potential to be part of this reaction. Once the tracking tool detects such a configuration described in Figure 7, it applies attractive forces between O atoms, H atoms, and C atoms from each PE molecule, and repulsive force between O atoms of the DCP molecule to drive the system into the crosslinking reaction. After completing this reaction, ML Assisted Hybrid/Reax circles back to non-reactive MD simulation to relax the system. For each cycle, typically 20,000 MD steps were run by ReaxFF with the tracking tool whereas 1,000,000 MD steps were run by the non-reactive MD simulation. After 20 cycles with ML Assisted Hybrid/Reax simulation, more than half of the initial 80 decane molecules had undergone crosslinking reactions. During the simulations almost all the crosslinking agents, DCP molecules, were consumed. At the end there were around 20 polymer molecules with more than 10 C atoms (Figure 8). In total, 20 million MD steps were simulated on a 16 core Intel Xeon E5 processor in 24 hours. This simulation would take approximately one month if it were run with the ReaxFF technique on the same machine.

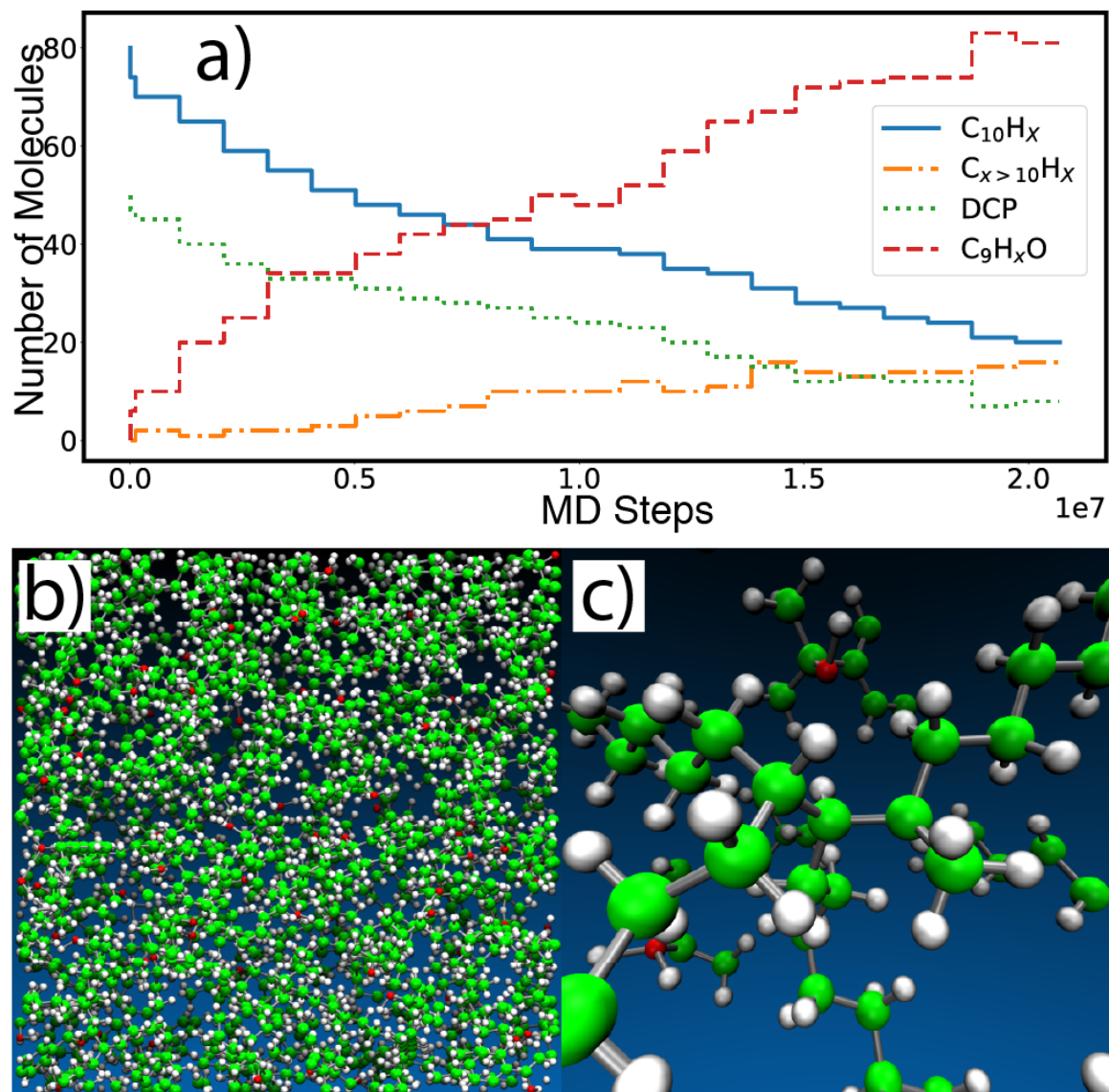


Figure 8 a) Molecule composition of the system during ML Assisted Hybrid Reax Simulations. b) Snapshot of the system and c) focus on the crosslinked decane after 20 million MD steps.

Conclusions

In this paper we developed a hybrid simulation method making use of non-reactive, Optimized Potential for Liquid Simulations (OPLS) and reactive (ReaxFF) force fields. The method uses a tandem approach, cycling back and forth between these two force fields with the help of Dense Neural Networks (DNNs) trained to predict OPLS parameters and atomic charges after ReaxFF simulations. Since the OPLS force field assigns parameters based on the local topology, descriptors for each parameter were designed based on the topology to train DNNs. The “tracking tool” was used during the ReaxFF simulations to accelerate chemical reactions. This

approach reduced the computation time significantly. Finally, we applied this method to a well-known crosslinking reaction of dicumyl peroxide and decane to demonstrate the efficiency of our hybrid approach. We expect that the developed method will be useful in areas where reactive simulations of large systems with longer simulation times are needed.

Acknowledgement

We acknowledge funding from DOW, grant number UPI 225559AM and funding from NSF, CDS&E grant no 1807622.

REFERENCES

1. Smit, D. F. a. B., *Understanding Molecular Dynamics Simulation: From algorithms to applications*. Academic Press: 2001; Vol. 1.
2. Alder, B. J.; Wainwright, T. E., Studies in Molecular Dynamics .1. General Method. *J. Chem. Phys.* **1959**, *31* (2), 459-466.
3. Rahman, A., Correlations in the Motion of Atoms in Liquid Argon . *Physical Review* **1964**, *136* (2A), 405.
4. Mattsson, T. R.; Lane, J. M. D.; Cochrane, K. R.; Desjarlais, M. P.; Thompson, A. P.; Pierce, F.; Grest, G. S., First-principles and classical molecular dynamics simulation of shocked polymers. *Phys. Rev. B* **2010**, *81* (5), 9.
5. Abell, G. C., Emprical chemical pseudopotential theory of molecular and metallic bonding. *Phys. Rev. B* **1985**, *31* (10), 6184-6196.
6. Liang, T.; Shin, Y. K.; Cheng, Y. T.; Yilmaz, D. E.; Vishnu, K. G.; Verners, O.; Zou, C. Y.; Phillpot, S. R.; Sinnott, S. B.; van Duin, A. C. T., Reactive Potentials for Advanced Atomistic Simulations. In *Annual Review of Materials Research, Vol 43*, Clarke, D. R., Ed. Annual Reviews: Palo Alto, 2013; Vol. 43, pp 109-129.
7. van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. A., ReaxFF: A reactive force field for hydrocarbons. *Journal of Physical Chemistry A* **2001**, *105* (41), 9396-9409.
8. Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J., Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J Am Chem Soc* **1996**, *118* (45), 11225-11236.
9. Hickey, D. R.; Yilmaz, D. E.; Chubarov, M.; Bachu, S.; Choudhury, T. H.; Miao, L. X.; Qian, C. H.; Redwing, J. M.; van Duin, A. C. T.; Alem, N., Formation of metal vacancy arrays in coalesced WS2 monolayer films. *2d Mater* **2021**, *8* (1).
10. Momeni, K.; Ji, Y. Z.; Wang, Y. X.; Paul, S.; Neshani, S.; Yilmaz, D. E.; Shin, Y. K.; Zhang, D. F.; Jiang, J. W.; Park, H. S.; Sinnott, S.; van Duin, A.; Crespi, V.; Chen, L. Q., Multiscale computational understanding and growth of 2D materials: a review. *Npj Comput Mater* **2020**, *6* (1).
11. Yilmaz, D. E.; Lotfi, R.; Ashraf, C.; Hong, S.; van Duin, A. C. T., Defect Design of Two-Dimensional MoS2 Structures by Using a Graphene Layer and Potato Stamp Concept. *Journal of Physical Chemistry C* **2018**, *122* (22), 11911-11917.

12. Dasgupta, N.; Yilmaz, D. E.; van Duin, A., Simulations of the Biodegradation of Citrate-Based Polymers for Artificial Scaffolds Using Accelerated Reactive Molecular Dynamics. *J. Phys. Chem. B* **2020**, *124* (25), 5311-5322.
13. Yilmaz, D. E.; van Duin, A. C. T., Investigating structure property relations of poly (p-phenylene terephthalamide) fibers via reactive molecular dynamics simulations. *Polymer* **2018**, *154*, 172-181.
14. Yilmaz, D. E., Modeling failure mechanisms of poly(p-phenylene terephthalamide) fiber using reactive potentials. *Comput. Mater. Sci.* **2015**, *109*, 183-193.
15. Mao, Q.; Rajabpour, S.; Kowalik, M.; van Duin, A. C. T., Predicting cost-effective carbon fiber precursors: Unraveling the functionalities of oxygen and nitrogen-containing groups during carbonization from ReaxFF simulations. *Carbon* **2020**, *159*, 25-36.
16. Kowalik, M.; Ashraf, C.; Damirchi, B.; Akbarian, D.; Rajahpour, S.; van Duin, A. C. T., Atomistic Scale Analysis of the Carbonization Process for C/H/O/N-Based Polymers with the ReaxFF Reactive Force Field. *J. Phys. Chem. B* **2019**, *123* (25), 5357-5367.
17. Akbarian, D.; Hamed, H.; Damirchi, B.; Yilmaz, D. E.; Penrod, K.; Woodward, W. H. H.; Moore, J.; Lanagan, M. T.; van Duin, A. C. T., Atomistic-scale insights into the crosslinking of polyethylene induced by peroxides. *Polymer* **2019**, *183*, 9.
18. Kelley, K. P.; Yilmaz, D. E.; Collins, L.; Sharma, Y.; Lee, H. N.; Akbarian, D.; van Duin, A. C. T.; Ganesh, P.; Vasudevan, R. K., Thickness and strain dependence of piezoelectric coefficient in BaTiO₃ thin films. *Phys. Rev. Mater.* **2020**, *4* (2).
19. Akbarian, D.; Yilmaz, D. E.; Cao, Y.; Ganesh, P.; Dabo, I.; Munro, J.; Van Ginhoven, R.; van Duin, A. C. T., Understanding the influence of defects and surface chemistry on ferroelectric switching: a ReaxFF investigation of BaTiO₃. *Physical Chemistry Chemical Physics* **2019**, *21*, 18240.
20. Verners, O.; Thijssse, B. J.; van Duin, A. C. T.; Simone, A., Salt concentration effects on mechanical properties of LiPF₆/poly(propylene glycol) diacrylate solid electrolyte: Insights from reactive molecular dynamics simulations. *Electrochimica Acta* **2016**, *221*, 115-123.
21. Islam, M. M.; Kolesov, G.; Kaxiras, E.; Van Duin, A., Treatment of explicit electrons in the ReaxFF reactive molecular dynamics and applications to battery interfaces. *Abstr Pap Am Chem S* **2015**, 250.
22. Raju, M.; Ganesh, P.; Kent, P. R. C.; van Duin, A. C. T., Reactive Force Field Study of Li/C Systems for Electrical Energy Storage. *J Chem Theory Comput* **2015**, *11* (5), 2156-2166.
23. Bedrov, D.; Smith, G. D.; van Duin, A. C. T., Reactions of Singly-Reduced Ethylene Carbonate in Lithium Battery Electrolytes: A Molecular Dynamics Simulation Study Using the ReaxFF. *Journal of Physical Chemistry A* **2012**, *116* (11), 2978-2985.
24. El-Zein, A.; Mohamed, K.; Talaat, M., Water trees in polyethylene insulated power cables: Approach to water trees initiation mechanism. *Electr. Power Syst. Res.* **2020**, *180*, 5.
25. Ashraf, C.; van Duin, A. C. T., Extension of the ReaxFF Combustion Force Field toward Syngas Combustion and Initial Oxidation Kinetics. *Journal of Physical Chemistry A* **2017**, *121* (5), 1051-1068.
26. Ostadhossein, A.; Rahnamoun, A.; Wang, Y. X.; Zhao, P.; Zhang, S. L.; Crespi, V. H.; van Duin, A. C. T., ReaxFF Reactive Force-Field Study of Molybdenum Disulfide (MoS₂). *J Phys Chem Lett* **2017**, *8* (3), 631-640.
27. Shin, Y. K.; Kwak, H.; Vasenkov, A. V.; Sengupta, D.; van Duin, A. C. T., Development of a ReaxFF Reactive Force Field for Fe/Cr/O/S and Application to Oxidation of Butane over a Pyrite-Covered Cr₂O₃ Catalyst. *Acs Catalysis* **2015**, *5* (12), 7226-7236.

28. Warshel, A., Computer simulations of enzyme catalysis: Methods, progress, and insights. *Annu. Rev. Biophys. Biomolec. Struct.* **2003**, *32*, 425-443.
29. Weismiller, M. R.; van Duin, A. C. T.; Lee, J.; Yetter, R. A., ReaxFF Reactive Force Field Development and Applications for Molecular Dynamics Simulations of Ammonia Borane Dehydrogenation and Combustion. *Journal of Physical Chemistry A* **2010**, *114* (17), 5485-5492.
30. Grimme, S., Exploration of Chemical Compound, Conformer, and Reaction Space with Meta-Dynamics Simulations Based on Tight-Binding Quantum Chemical Calculations. *J Chem Theory Comput* **2019**, *15* (5), 2847-2862.
31. Hu, Q. K.; Wang, H. Y.; Wu, Q. H.; Ye, X. T.; Zhou, A. G.; Sun, D. D.; Wang, L. B.; Liu, B. Z.; He, J. L., Two-dimensional Sc₂C: A reversible and high-capacity hydrogen storage material predicted by first-principles calculations. *International Journal of Hydrogen Energy* **2014**, *39* (20), 10606-10612.
32. Lee, G. D.; Wang, C. Z.; Yoon, E.; Hwang, N. M.; Kim, D. Y.; Ho, K. M., D(i)ffusion, coalescence, and reconstruction of vacancy defects in graphene layers. *Phys. Rev. Lett.* **2005**, *95* (20), 4.
33. Hughey, S.; Alsnayyan, A.; Aktulga, H. M.; Gao, T.; Shanker, B., Fast and scalable evaluation of pairwise potentials. *Comput Phys Commun* **2020**, 255.
34. O'Hearn, K. A.; Alperen, A.; Aktulga, H. M., FAST SOLVERS FOR CHARGE DISTRIBUTION MODELS ON SHARED MEMORY PLATFORMS. *Siam Journal on Scientific Computing* **2020**, *42* (1), C1-C22.
35. Ganeshan, K.; Hossain, M. J.; van Duin, A. C. T., Multiply accelerated ReaxFF molecular dynamics: coupling parallel replica dynamics with collective variable hyper dynamics. *Molecular Simulation* **2019**, *45* (14-15), 1265-1272.
36. Jing, Z. F.; Xin, L.; Sun, H., Replica exchange reactive molecular dynamics simulations of initial reactions in zeolite synthesis. *Physical Chemistry Chemical Physics* **2015**, *17* (38), 25421-25428.
37. Joshi, K. L.; Raman, S.; van Duin, A. C. T., Connectivity-Based Parallel Replica Dynamics for Chemically Reactive Systems: From Femtoseconds to Microseconds. *J Phys Chem Lett* **2013**, *4* (21), 3792-3797.
38. Jain, A. K.; Duin, R. P. W.; Mao, J. C., Statistical pattern recognition: A review. *Ieee Transactions on Pattern Analysis and Machine Intelligence* **2000**, *22* (1), 4-37.
39. Schmidhuber, J., Deep learning in neural networks: An overview. *Neural Networks* **2015**, *61*, 85-117.
40. Behler, J., Perspective: Machine learning potentials for atomistic simulations. *J. Chem. Phys.* **2016**, *145* (17).
41. Behler, J., Constructing high-dimensional neural network potentials: A tutorial review. *Int. J. Quantum Chem.* **2015**, *115* (16), 1032-1050.
42. Deringer, V. L.; Csanyi, G., Machine learning based interatomic potential for amorphous carbon. *Phys. Rev. B* **2017**, *95* (9).
43. Hansen, K.; Biegler, F.; Ramakrishnan, R.; Pronobis, W.; von Lilienfeld, O. A.; Muller, K. R.; Tkatchenko, A., Machine Learning Predictions of Molecular Properties: Accurate Many-Body Potentials and Nonlocality in Chemical Space. *J Phys Chem Lett* **2015**, *6* (12), 2326-2331.
44. Lahey, S. L. J.; Rowley, C. N., Simulating protein-ligand binding with neural network potentials. *Chem. Sci.* **2020**, *11* (9), 2362-2368.
45. Ramprasad, R.; Batra, R.; Pilania, G.; Mannodi-Kanakkithodi, A.; Kim, C., Machine learning in materials informatics: recent applications and prospects. *Npj Comput Mater* **2017**, *3*.

46. Zuo, Y. X.; Chen, C.; Li, X. G.; Deng, Z.; Chen, Y. M.; Behler, J.; Csanyi, G.; Shapeev, A. V.; Thompson, A. P.; Wood, M. A.; Ong, S. P., Performance and Cost Assessment of Machine Learning Interatomic Potentials. *Journal of Physical Chemistry A* **2020**, *124* (4), 731-745.
47. Behler, J., Perspective: Machine learning potentials for atomistic simulations. *J. Chem. Phys.* **2016**, *145* (17), 9.
48. Vashisth, A.; Ashraf, C.; Zhang, W. W.; Bakis, C. E.; van Duin, A. C. T., Accelerated ReaxFF Simulations for Describing the Reactive Cross-Linking of Polymers. *Journal of Physical Chemistry A* **2018**, *122* (32), 6633-6642.
49. Watkins, E. K.; Jorgensen, W. L., Perfluoroalkanes: Conformational analysis and liquid-state properties from ab initio and Monte Carlo calculations. *Journal of Physical Chemistry A* **2001**, *105* (16), 4118-4125.
50. Martinez, J. A.; Yilmaz, D. E.; Liang, T.; Sinnott, S. B.; Phillpot, S. R., Fitting empirical potentials: Challenges and methodologies. *Curr Opin Solid St M* **2013**, *17* (6), 263-270.
51. Martinez, J. A.; Chernatynskiy, A.; Yilmaz, D. E.; Liang, T.; Sinnott, S. B.; Phillpot, S. R., Potential Optimization Software for Materials (POSMat). *Comput Phys Commun* **2016**, *203*, 201-211.
52. Jorgensen, W. L.; Tirado-Rives, J., Molecular modeling of organic and biomolecular systems using BOSS and MCPRO. *Journal of Computational Chemistry* **2005**, *26* (16), 1689-1700.
53. Pence, H. E.; Williams, A., ChemSpider: An Online Chemical Information Resource. *Journal of Chemical Education* **2010**, *87* (11), 1123-1124.
54. Dodda, L. S.; Vilseck, J. Z.; Tirado-Rives, J.; Jorgensen, W. L., 1.14*CM1A-LBCC: Localized Bond-Charge Corrected CM1A Charges for Condensed-Phase Simulations. *J. Phys. Chem. B* **2017**, *121* (15), 3864-3870.
55. Martín Abadi, A. A., Paul Barham, Eugene Brevdo,; Zhifeng Chen, C. C., Greg S. Corrado, Andy Davis,; Jeffrey Dean, M. D., Sanjay Ghemawat, Ian Goodfellow,; Andrew Harp, G. I., Michael Isard, Rafal Jozefowicz, Yangqing Jia,; Lukasz Kaiser, M. K., Josh Levenberg, Dan Mané, Mike Schuster,; Rajat Monga, S. M., Derek Murray, Chris Olah, Jonathon Shlens,; Benoit Steiner, I. S., Kunal Talwar, Paul Tucker,; Vincent Vanhoucke, V. V., Fernanda Viégas,; Oriol Vinyals, P. W., Martin Wattenberg, Martin Wicke,; Yuan Yu, a. X. Z. TensorFlow: Large-scale machine learning on heterogeneous systems.
56. Yilmaz, D. E., PolyBuild: Building Polymer Ceramic Composites,. 2020.