

# Reactive molecular dynamics simulations and machine learning

Aravind Krishnamoorthy<sup>1</sup>, Pankaj Rajak<sup>1, 2</sup>, Sungwook Hong<sup>1</sup>, Ken-ichi Nomura<sup>1</sup>, Subodh Tiwari<sup>1</sup>, Rajiv K Kalia<sup>1</sup>, Aiichiro Nakano<sup>1</sup>, and Priya Vashishta<sup>1</sup>

<sup>1</sup> Collaboratory for Advanced Computing and Simulations, Department of Physics & Astronomy, Department of Chemical Engineering & Materials Science, and Department of Computer Science, University of Southern California, Los Angeles, CA 90089, USA

<sup>2</sup> Argonne Leadership Computing Facility, Argonne National Laboratory, Lemont, IL 60439, USA

Email: priyav@usc.edu

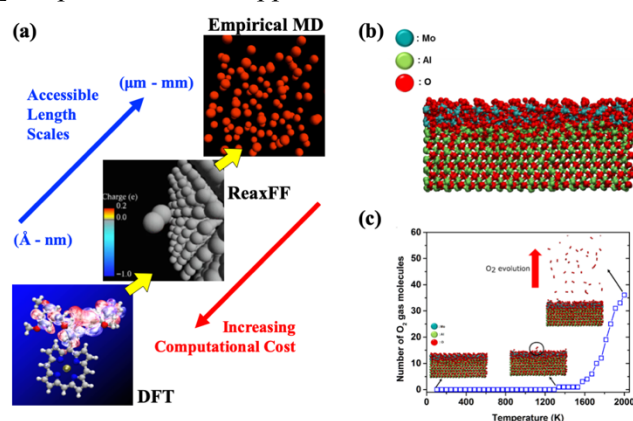
**Abstract.** Machine learning (ML) is revolutionizing scientific and engineering disciplines owing to its ability to capture hidden patterns in large amounts of data. The recent success of ML can be attributed to increasing amount of data, simulation resources, and improving understanding of statistical inference. For these reasons computational materials science is undergoing a paradigm shift. The main reason is that trial-and-error approach to materials design is inefficient: laboratory trials require a lot of time, and the results of previous trials are not utilized in a systematic fashion. A data-driven approach, which draws upon all relevant data from experiments, and reactive and quantum molecular dynamics simulations, can address these issues. The MAGICS (Materials Genome Innovation for Computational Software) Center develops to aid the synthesis of stacked layered materials by chemical vapor deposition, exfoliation, and intercalation. The identification of different phases is a classification problem and can be solved using ML techniques. We have used feed-forward neural network with three hidden layers to identify the different phases present during computational synthesis of MoSe<sub>2</sub>. Our goal is to carry out exascale MD simulations using forces for multimillion-atom quantum dynamics simulations from neural networks and deep learning using small QMD simulations.

## 1. Reactive Molecular Dynamics Simulations of Computational Synthesis

The reactive molecular dynamics (RMD) method has enabled large-scale simulations of chemical events in complex materials involving multimillion atoms [1-3]. In particular, RMD simulations based on first principles-informed reactive force fields (ReaxFF) [4] describe chemical reactions (i.e., bond breakage and formation) through a bond-order/distance relationship that reflects each atom's coordination change. ReaxFF-RMD simulations describe full dynamics of chemical events at the atomic level with significantly reduced computational cost compared with quantum-mechanics (QM) calculations. ReaxFF consists of a number of empirical force-field parameters in its functional form, which are optimized mainly against a QM-based training set that includes not only energies of small clusters (e.g., full bond dissociation, angle distortion and torsion energies) and reaction energies/ barriers for key chemical reactions, but also bulk properties of crystal systems. ReaxFF has shown its ability to successively study chemical, physical and mechanical properties of a wide range of complex materials such as hydrocarbons, high energy materials and metal/transition-metal systems. These qualities, most notably, the near-quantum accuracy, ability to model very large materials systems, while retaining

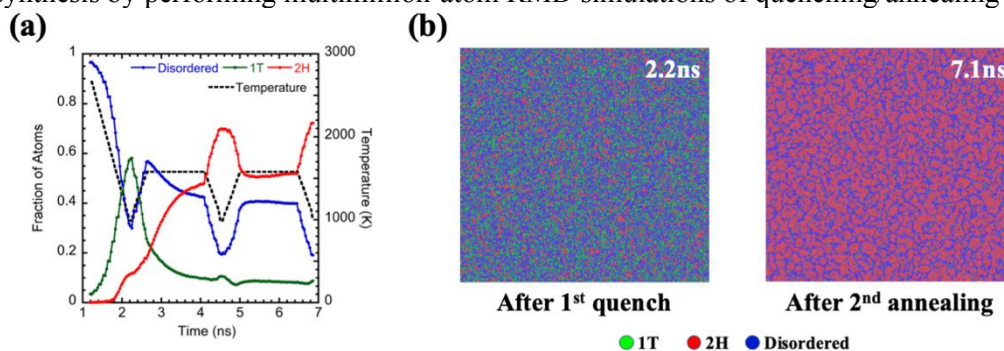


atomic resolution have made ReaxFF-RMD an invaluable method for investigating atomic-scale mechanisms and reaction pathways encountered in chemical reactions involved in synthesis of functional engineering materials. Such computational modelling, particularly reactive molecular dynamics (RMD) simulations, can provide useful insights into interfaces [5] and surface-gas interactions [6] on model systems down to atomic length scales. At MAGICS Center, we have performed extensive ReaxFF-RMD simulations of a technologically important chemical reaction – that of computational synthesis of MoS<sub>2</sub> structures using MoO<sub>3</sub> surfaces and gaseous S<sub>2</sub>. Molybdenum disulfide (MoS<sub>2</sub>), one of archetypal transition metal dichalcogenide (TMDC) materials, has received much attention owing to its unique physical, chemical, and mechanical properties.[7] MoS<sub>2</sub> monolayer has a high band gap, high carrier mobility, exceptional mechanical strength, compared to conventional Si. [8, 9] for potential use in applications such as ultra-thin channel materials for transistors.[10, 11]



**Figure 1.** (a) Hierarchy of molecular dynamics methods demonstrating the tradeoff between computational cost and accuracy and accessible length scales. (b) and (c) show the prototypical configuration and results of computational synthesis of MoS<sub>2</sub> from MoO<sub>3</sub> precursors deposited on Al<sub>2</sub>O<sub>3</sub> substrates. Specifically, ReaxFF-RMD simulations accurately describe the O<sub>2</sub> evolution and self-reduction of the O-rich MoO<sub>3</sub> layer at high temperatures characteristic of CVD synthesis.

To bring MoS<sub>2</sub> into mass production, numerous effort has been made, including physical vapor deposition, mechanical exfoliation, hydrothermal synthesis, and CVD.[10, 12] Among these techniques, CVD is the most practical and scalable way to synthesize large-scale and high-quality MoS<sub>2</sub> layer on the target substrate.[13, 14] However, optimal conditions for CVD synthesis of highly-crystalline MoS<sub>2</sub> layers have yet to be fully investigated. Deciphering selection rules for different growth scenarios (e.g. temperature cycles and annealing schedules) to make predictions of optimized environmental parameters and growth factors has remained unclear due to a lack of understanding of mechanistic processes for CVD growth of MoS<sub>2</sub>. At MAGICS Center, we performed multimillion-atom RMD simulations, which allows an atomic scale insight into reaction dynamics of nanostructured materials,[5] to investigate grain formation and evolution of line defects and their healing process. Our RMD simulations are based on the first-principle-informed ReaxFF [15] reactive force field, which was previously used and validated for CVD synthesis reactions.[16, 17] Specifically, we investigate the effect of temperature cycling and annealing schedules on the local atomic structures encountered during CVD synthesis by performing multimillion-atom RMD simulations of quenching/annealing steps.



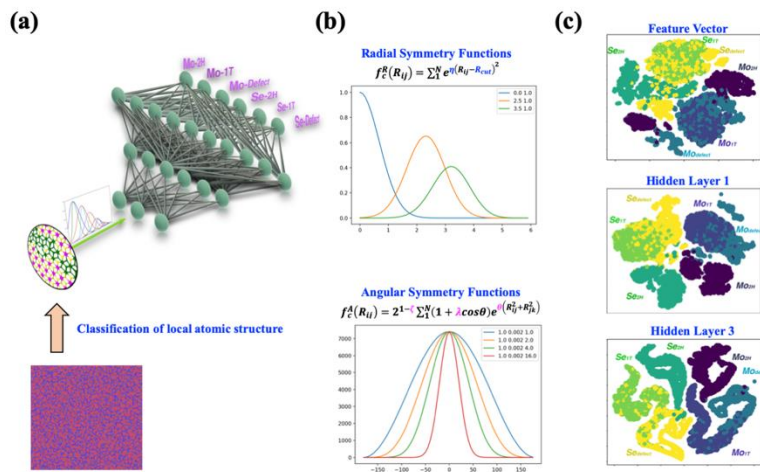
**Figure 2.** Annealing schedule and phase composition in the RMD simulation shows increase in fraction of crystalline 1T and 2H with lower annealing temperature (b) This is also reflected in the phase fraction maps plotted at  $t = 2.2$  ns and  $t = 7.1$  ns. These

maps are plotted by assigning a colour to each atom in the reacting substrate based on local atomic configurations.

Our computational model for RMD simulations consists of 4,305,600 atoms. The system dimensions are  $211.0 \times 196.3$  (nm<sup>2</sup>) in the x-, and y-directions. ReaxFF reactive force field parameters were taken from Hong *et al.* [17] which can reproduce critical reaction steps for CVD synthesis of MoS<sub>2</sub> (i.e., O<sub>2</sub> evolution and SO/SO<sub>2</sub> formation), validated through experimental literature and DFT calculations [16]

## 2. Machine Learning for Analysis and Classification of Local Atomic Structures

MD simulation of physical and chemical phenomena of materials requires complex data analysis of the simulation results to identify different phases, chemical reaction and defects. Identification of different phases and defects requires complex structural analysis ranging from calculation of nearest neighbours to shortest circuit analysis of atoms. Traditional approaches for structural analysis like common neighbourhood analysis and centro-symmetry parameter calculation only work for mono-atomic systems and can not distinguish 2H and 1T phases in TMDC crystals contains transition metal, TM and chalcogen, X. Both 2H and 1T structures have the same number of nearest neighbours due to which nearest-number analysis is not able to distinguish these structures. Compared to these traditional approaches, our NN model is able to define a unique order parameter that identifies these phases with high accuracy, is highly scalable and can used to analyse large data set quickly. To build a NN model for structural analysis, we have represented each atom by 436-dimension feature vector, made up of radial and angular symmetry functions (Figure 3b). Figure 3 shows the learned representation of the data set at the first and third layer of the neural network and the original feature vector using t-distributed stochastic neighbour embedding (t-SNE) [22], which is a dimensionality reduction technique that represents high-dimensional data in a lower dimension. The data is plotted for an example MoSe<sub>2</sub> TMDC structure [18] that contains 2H, 1T and defective regions. The separation between the six phases become clearer in the third hidden layer as compared to in hidden layer 1 and the original feature space.

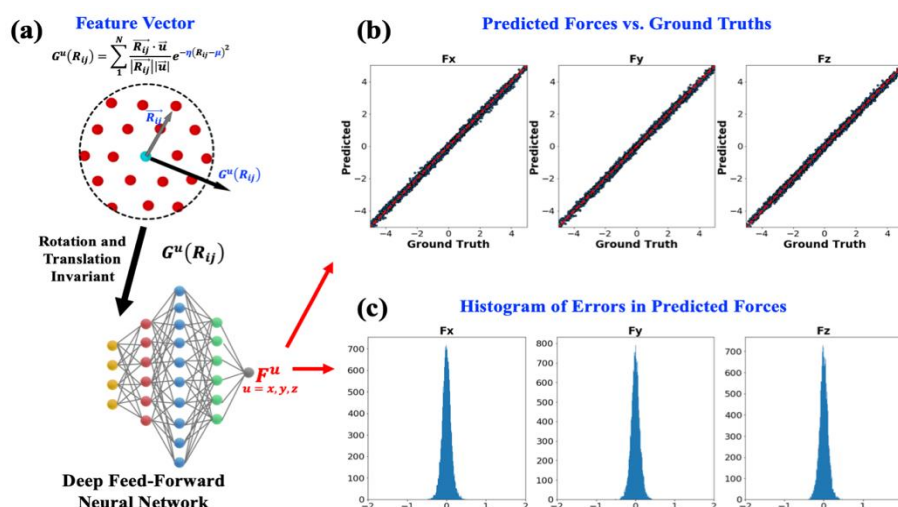


**Figure 3:** (a) Schematic of a deep fully connected neural network used for structure and phase classification of an example MoSe<sub>2</sub> TMDC structure (b) Radial (2-body) and Angular (3-body) symmetry functions for constructing feature vectors from local configurations around each atom. (c) t-SNE representation of different atomic configurations showing that the neural network has learned to discriminate different phases in the input data.

## 3. Machine Learning of Interatomic Forces for Molecular Dynamics Simulations

Machine Learning models have proven to be capable of learning highly complex non-linear functions and relationships between input functions and output observables [7–19]. Recently several techniques have been implemented that can effectively learn the potential energy surface (PES) of a complex collection of atoms, which, in turn, describes their structural and dynamical properties. These ML methods trained on *ab initio* calculations have been shown to have *ab initio* accuracy and are comparable to empirical FFs in cost. The two primary techniques for learning arbitrary PES are Gaussian Process methods and Artificial Neural Networks. The former method, pioneered by Csanyi *et al.* [19, 20], involves the use of non-parametric Gaussian Kernels centred on training data, primarily energies of atomic configurations and atomic forces. These Gaussian Approximation Potentials (GAP) have been demonstrated to be highly accurate for a wide range of mono- and diatomic systems but possess a computational complexity of  $O(n^3)$ , which restricts their speed and applicability. In contrast, the Neural

Network approach for approximating PES, proposed by Parrinello et al. [21, 22], uses multi-layer neural networks and non-linear activation functions to learn the dependence of the PES on the (featurized) atomic coordinates. These networks are commonly trained on system energies, with interatomic forces being obtained as gradients with respect to the feature-vector during back-propagation. The speed of force prediction in these models is limited by the relatively slow gradient computation process. To overcome this limitation, we have constructed a deep neural network model trained exclusively on interatomic forces, using feature vectors qualitatively similar to those of Behler et al [22]. The primary advantage of this model is the use of forward propagation during force prediction, which significantly improves prediction speed. Figure 4 (a) shows the schematic of a 3-layer neural network model for force prediction for a system of crystalline aluminium along with the 2-body radial feature vector. Figures 4(b) and 4(c) highlight the quality of the trained forcefield and demonstrate the predicted forces are consistent, both in magnitude and direction, with quantum-mechanical forces for the Al system.



**Figure 4.** (a) Radial feature vector and schematic of the neural network used for computing force in the AI model system. (b) NN-predicted forces compared with DFT ground truth values for a test set shows a well-trained model with relatively

#### 4. Conclusion

MAGICS Center has performed several multi-million-atom ReaxFF-RMD simulations on computational synthesis of MoS<sub>2</sub> monolayers from MoO<sub>3</sub> precursors to identify atomic-scale mechanisms and reaction pathways for guided design of synthesis conditions. Our RMD simulations show a strong dependence of 2H, 1T and defective structures as well as grain nucleation and growth on multiple heating/cooling cycles on the CVD growth may play an important role in synthesis of high-quality MoS<sub>2</sub> layers. The analysis of local atomic structures to identify crystalline and defective phases is highly non-trivial. We constructed a three-layer, feed-forward neural network (NN) model to analyze different structures generated during MD simulation. The NN model classifies each atom in one of the six phases which are either as transition metal or chalcogen atoms in 2H phase, 1T phase and defects. t-SNE analyses of learned representation of these phases in the hidden layers of the NN model show that separation of all phases becomes clearer in the third layer. Machine Learning can be used to learn highly complex relationships including those between local atomic configurations and energies and forces.

#### References

- [1] Nomura K-i, Kalia R K, Nakano A and Vashishta P 2008 A scalable parallel algorithm for large-scale reactive force-field molecular dynamics simulations *Computer Physics Communications* **178** 73-87
- [2] Li Y, Nomura K, Insley J, Morozov V, Kumaran K, Romero N, W. Goddard I, Kalia R, Nakano A and Vashishta P 2018 Scalable Reactive Molecular Dynamics Simulations for Computational Synthesis *Computing in Science & Engineering* 1-

- [3] Nomura K, Small P E, Kalia R K, Nakano A and Vashishta P 2015 An extended-Lagrangian scheme for charge equilibration in reactive molecular dynamics simulations *Computer Physics Communications* **192** 91-6
- [4] van Duin A C T, Dasgupta S, Lorant F and A G W 2001 ReaxFF: A Reactive Force Field for Hydrocarbons *Journal of Physical Chemistry A* **105** 9396-409
- [5] Mo Y, Turner K T and Szlufarska I 2009 Friction laws at the nanoscale *Nature* **457** 1116-9
- [6] Mueller J E, van Duin A C T and Goddard W A 2010 Development and Validation of ReaxFF Reactive Force Field for Hydrocarbon Chemistry Catalyzed by Nickel *Journal of Physical Chemistry C* **114** 4939-49
- [7] Lee Y H, Zhang X Q, Zhang W, Chang M T, Lin C T, Chang K D, Yu Y C, Wang J T W, Chang C S and Li L J 2012 Synthesis of Large-Area MoS<sub>2</sub> Atomic Layers with Chemical Vapor Deposition *Adv. Mater.* **24** 2320-5
- [8] Lembke D and Kis A 2012 Breakdown of high-performance monolayer MoS<sub>2</sub> transistors *ACS Nano* **6** 10070-5
- [9] Ganatra R and Zhang Q 2014 Few-layer MoS<sub>2</sub>: a promising layered semiconductor *ACS Nano* **8** 4074-99
- [10] Venkata Subbaiah Y, Saji K and Tiwari A 2016 Atomically Thin MoS<sub>2</sub>: A Versatile Nongraphene 2D Material *Adv. Funct. Mater.* **26** 2046-69
- [11] Gupta A, Sakthivel T and Seal S 2015 Recent development in 2D materials beyond graphene *Prog. Mater. Sci.* **73** 44-126
- [12] Lv Z, Mahmood N, Tahir M, Pan L, Zhang X and Zou J-J 2016 Fabrication of zero to three dimensional nanostructured molybdenum sulfides and their electrochemical and photocatalytic applications *Nanoscale* **8** 18250-69
- [13] Yu J, Li J, Zhang W and Chang H 2015 Synthesis of high quality two-dimensional materials via chemical vapor deposition *Chem. Sci.* **6** 6705-16
- [14] Zhan Y, Liu Z, Najmaei S, Ajayan P M and Lou J 2012 Large-area vapor-phase growth and characterization of MoS<sub>2</sub> atomic layers on a SiO<sub>2</sub> substrate *Small* **8** 966-71
- [15] van Duin A C T, Dasgupta S, Lorant F and Goddard W A 2001 ReaxFF: a reactive force field for hydrocarbons *J. Phys. Chem. A* **105** 9396-409
- [16] Hong S, Krishnamoorthy A, Rajak P, Tiwari S C, Misawa M, Shimojo F, Kalia R K, Nakano A and Vashishta P 2017 Computational Synthesis of MoS<sub>2</sub> Layers by Reactive Molecular Dynamics Simulations: Initial Sulfidation of MoO<sub>3</sub> Surfaces *Nano Lett.* **17** 4866-72
- [17] Hong S, Sheng C, Krishnamoorthy A, Rajak P, Tiwari S C, Nomura K-i, Misawa M, Shimojo F, Kalia R K and Nakano A 2018 Chemical Vapor Deposition Synthesis of MoS<sub>2</sub> Layers from the Direct Sulfidation of MoO<sub>3</sub> Surfaces Using Reactive Molecular Dynamics Simulations *The Journal of Physical Chemistry C* **122** 7494-503
- [18] Apte A, Kochat V, Rajak P, Krishnamoorthy A, Manimunda P, Hachtel J A, Idrobo J C, Amanulla S A S, Vashishta P, Nakano A, Kalia R K, Tiwary C S and Ajayan P M 2018 Structural Phase Transformation in Strained Monolayer MoWSe<sub>2</sub> Alloy *Acs Nano* **12** 3468-76
- [19] Bartok A P, Payne M C, Kondor R and Csanyi G 2010 Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons *Physical Review Letters* **104** 136403
- [20] Mocanu F C, Konstantinou K, Lee T H, Bernstein N, Deringer V L, Csanyi G and Elliott S R 2018 Modeling the Phase-Change Memory Material, Ge<sub>2</sub>Sb<sub>2</sub>Te<sub>5</sub>, with a Machine-Learned Interatomic Potential *J Phys Chem B* **122** 8998-9006
- [21] Behler J 2017 First Principles Neural Network Potentials for Reactive Simulations of Large Molecular and Condensed Systems *Angew Chem Int Edit* **56** 12828-40
- [22] Behler J and Parrinello M 2007 Generalized neural-network representation of high-dimensional potential-energy surfaces *Physical Review Letters* **98**

### Acknowledgments

This work supported as part of the Computational Materials Sciences Program, U.S. Department of Energy, Basic Energy Sciences, Award DE-SC0014607.