

MRS Advances © 2018 Materials Research Society. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.
DOI: 10.1557/adv.2018.673



Neural Network Analysis of Dynamic Fracture in a Layered Material

Pankaj Rajak^{1,2}, Rajiv K. Kalia^{2,3,4,5}, Aiichiro Nakano^{2,3,4,5,6}, Priya Vashishta^{2,3,4,5}

¹Argonne Leadership Computing Facility, Argonne National Laboratory, Argonne, IL 60439, USA;

²Collaboratory for Advanced Computing and Simulations, ³Department of Physics, ⁴Department of Computer Science, ⁵Department of Chemical Engineering and Material Science, ⁶Department of Biological Sciences, University of Southern California, Los Angeles, CA 90089

Abstract

Dynamic fracture of a two-dimensional MoWSe_2 membrane is studied with molecular dynamics (MD) simulation. The system consists of a random distribution of WSe_2 patches in a pre-cracked matrix of MoSe_2 . Under strain, the system shows toughening due to crack branching, crack closure and strain-induced structural phase transformation from 2H to 1T crystal structures. Different structures generated during MD simulation are analyzed using a three-layer, feed-forward neural network (NN) model. A training data set of 36,000 atoms is created where each atom is represented by a 50-dimension feature vector consisting of radial and angular symmetry functions. Hyper parameters of the symmetry functions and network architecture are tuned to minimize model complexity with high predictive power using feature learning, which shows an increase in model accuracy from 67% to 95%. 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. Further *t*-SNE analyses of learned representation of these phases in the hidden layers of the NN model show that separation of all phases become clearer in the third layer than in layers 1 and 2.

INTRODUCTION

Molecular dynamics (MD) simulation of various physical and chemical phenomena of materials requires complex data analysis of the simulation results so as to identify different phases, chemical reaction and defects generated during the simulation. For example, during nanoindentation simulation, plastic deformation occurs inside the material due to dislocation nucleation.^[1] Similarly, under stress, materials like SiC, AlN show phase transformation.^[2, 3] Identification of different phases and defects generated during MD simulation requires complex structural analysis. These analyses range from calculation of nearest neighbours to shortest circuit analysis of atoms. In general, there is no single evaluation technique that can be applied to all simulation data since these analyses are system dependent. However, we can observe that structures generated in MD simulations are complex functional forms of their local environment. Hence, we can create a machine learning (ML) model that can automatically learn this functional form directly from the simulation data.

Recently ML methods in the material science domain have shown phenomenal success for high-throughput screening and property prediction.^[4-8] For example, statistical models built using a smaller fraction of the crystal structures (called training data) can accurately predict a wide range of material properties like band gap,^[9, 10] dielectric breakdown strength^[11, 12] and melting point^[10] for remaining crystal structure without doing any quantum molecular dynamics (QMD) simulation for them. ML methods like support vector regression^[9, 10], neural network^[13] and kernel ridge regression^[14] are shown to accurately predict the bandgap of double perovskites, polymers^[15, 16] and defects in FCC and disordered solid.^[17] Neural network and gaussian processes model (GP) are also used to create data driven force field for materials, where parameters of the model are learned using the QMD simulation trajectory.^[18-20]

In this work, we have studied mode I fracture of MoWSe₂ monolayer using MD simulation. At the onset of crack propagation, the system shows strain-induced phase transformation. To identify these phases, we have trained a three-layer feed-forward neural network (NN) model. The training data set consists of 36,000 atomic data examples, where each atom is represented by a 50-dimension feature vector consisting of radial and angular symmetry functions. The model classifies each atom in one of the six phases, which are either transition metal or chalcogen atoms in 2H phase, 1T phase and defects.

Traditional approach for structural analysis like common neighborhood analysis and centro-symmetry parameter calculation only work for mono-atomic systems like FCC, BCC and HCP crystals, do not distinguish 2H and 1T crystalline phases in transition-metal dichalcogenide layers considered here. Similarly, both 2H and 1T structures in our simulation have the same number of nearest neighbors (nn) – each Mo atom has 6 Se nn and Se atom has 3 Mo nn – due to which nearest-number analysis is not able to distinguish these structures either. 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 be used to analyse large data sets quickly.

METHOD

A. Model fracture of MoWSe₂ hetero-structure

Molecular dynamics (MD) simulations are performed to study mode I fracture of MoSe₂ monolayer consisting of a pre-crack and a single triangular patch of WSe₂ in front of the pre-crack. The system is first relaxed at 100 K and then subjected to tensile force perpendicular to the crack front. The system fractures at a strain of 2.16 %, which matches with the experimental value^[21] At the onset of crack propagation, we observe strain induced 2H to 1T phase transformation as shown in figure 1.^[21] The transformed 1T region also contains defects, which is shown in green colour. Movie S1 in supplementary material shows this phase transformation during crack propagation MoWSe₂ monolayer.

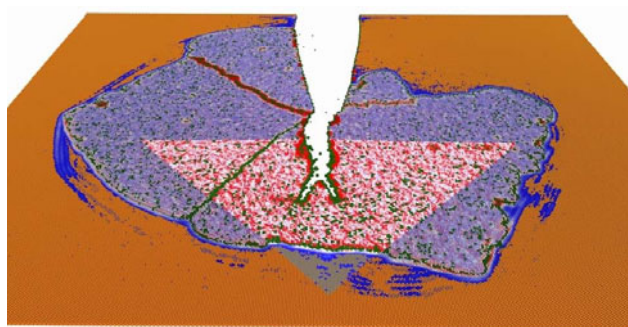


Figure 1: Strain induced 2H to 1T phase transformation in MoSe₂ monolayer with WSe₂ triangular patch. 2H regions are shown in orange, transformed 1T regions in blue and red, and defects are in green.

B. Feature Vector for Neural Network model for structural analysis

To build a NN model, we need to define a mathematical representation for each atom (feature vector) that captures its local environment. There are many choices for feature vector such as electronegativity (EN), ionization potential (IP), atomic displacement, atomic mass, local stress, radial and angular symmetry functions. In general, the choice of correct feature vector is highly problem specific. For example, EN, IP and atomic mass are suitable for build regression models for band gap for different materials,^[9, 10] whereas radial and annular symmetry functions are suitable to create NN and GP force fields.^[18-20] To build a NN model for structural analysis, we have represented each atom by 436-dimension feature vector, which consists of combinations of radial and angular symmetry functions. The equations for radial and angular feature are shown in equations 1 and 2 respectively.

$$f_c^R(R_{ij}) = \begin{cases} \left(\sum_{j=1}^N e^{\eta_1(R_{ij}-R_{cut})^2} \cdot f_c(R_{ij})\right) & R_{ij} < R_{cut} \\ 0 & R_{ij} \geq R_{cut} \end{cases} \quad (\text{Eq. 1})$$

$$f_c^A(R_{ij}) = \begin{cases} \left(2^{1-\zeta} \sum_{j=1}^N (1 + \lambda \cos \theta_{ijk})^\zeta \cdot e^{\eta_1(R_{ij}^2+R_{ik}^2)} \cdot f_c(R_{ij}) \cdot f_c(R_{ik})\right) & R_{ij} < R_{cut} \\ 0 & R_{ij} \geq R_{cut} \end{cases} \quad (\text{Eq. 2})$$

$$f_c(R_{ij}) = \begin{cases} 0.5 \times \left[\cos\left(\frac{\pi R_{ij}}{R_c}\right) + 1\right] & R_{ij} < R_{cut} \\ 0 & R_{ij} \geq R_{cut} \end{cases} \quad (\text{Eq. 3})$$

In equations 1-3, η_1 , η_2 , ζ , λ and R_{cut} are tunable parameters, which are also known as hyper-parameters of the model. For each atom i , contribution due to all its N neighbors that are within the cutoff distance R_{cut} (8Å) is computed for a specific value of radial (η_1) and angular (η_2 , ζ , λ) features, respectively. Effect of neighbor atoms that are further away from the central atom is damped using equation 3. The radial feature vector for each atom is created using the contribution from Mo-Mo, Mo-Se and Se-Se bond, where η_1 is varied between 0.0001 to 1000 on a log scale. For the angular features, contribution from Mo-Se-Mo, Se-Mo-Se, Mo-Mo-Se, Se-Se-Mo bonds are used. For the angular features, λ is taken as -1 or +1, ζ and η_2 between 0.1 to 10. In total, we have 113 radial feature and 323 angular features for each atom.

C. Neural Network model for structural analysis

A neural network model with three hidden layers is trained to identify the different phases of MoWSe₂ monolayer (see figure 2). Output layer of the NN model consists of 6 classes, which are Mo in 2H phase, Mo in 1T phase, Mo as defect, Se in 1H phase, Se in 1T phase and Se as defect. Here, input to NN model is the feature vector of the atoms, and the output is their probabilities in these six classes, and we choose the class with maximum probability as the phase label for the atoms. The first, second and third hidden layer has 350, 100 and 50 hidden units. In the first and second hidden layer, we have used Relu ($y = \max(0, x)$) activation function while in the third hidden layer sigmoid function ($\sigma(x) = (1 + e^{-x})^{-1}$) is used. The NN model is trained using 36,000 examples, which consists of equal distribution of atomic data from all six classes. The model is trained for 200 epoch, where in each epoch a batch size of 512 is used with a dropout probability of 0.25 for each hidden layer. After training, the performance of the model is validated on an unseen test simulation dataset.

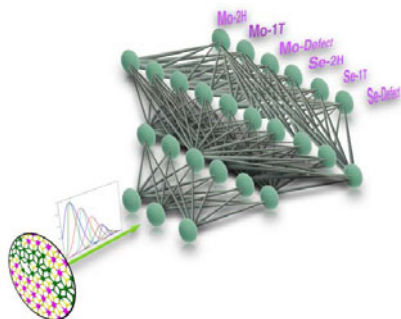


Figure 2: Schematic of the neural network model showing the mapping of atomic coordinates into sets of radial and angular feature vectors, NN layer architecture and the six out phases predicted by the model.

RESULTS

Visualization of the learned data representation by NN model

Figure 3 shows the learned representation of the data set by the first and third layer of the neural network and the original feature vector using principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE).^[22] Both PCA and t-SNE are dimensionality reduction technique and represents high-dimensional data in lower dimension. However, PCA preserves the global information like the distance between different clusters, nearest neighbor of clusters but in doing so loses the local variation among data. For example, in figure 3a-b, we observe that PCA is able to distinguish the two major phases, which are transition metal (Mo, W) and chalcogen atoms (Se) but it fails to capture the local variation in their structural representation, which are 2H crystal, 1T crystal and defects. The separation between 2H and 1T crystal structures become clearer in the third hidden layer as it contains four clusters compared to hidden layer 1,2 and the original feature space who contains only two clusters. Here, the axis of figure 3a-c corresponds to the largest two eigenvalues of the feature vector of the training data set.

Figure 3d-f shows the t-SNE representation of training data set. We observe that separation between all the six phases become clearer in the third hidden layer as compared to in hidden layer 1 and the original feature space. In the original feature vector representation of atoms, each atom is represented as some combination of radial and angular symmetry functions, which does not consider the non-linear interaction between these symmetry functions. Due to the absence of this nonlinear interaction among symmetry functions, there is no clear separation between all phases in the original feature vector (see figure 3d). Hidden layer 1 considers this nonlinear interaction between the symmetry function, as result of which it is able to understand the representation of 2H and 1T crystal structure (figure 3e). However, it is still not able to distinguish between defects and 1T crystal structure. Addition of more

layers alleviate this problem, and the third layer of the NN model is able to distinguish all the six phases (figure 3f).

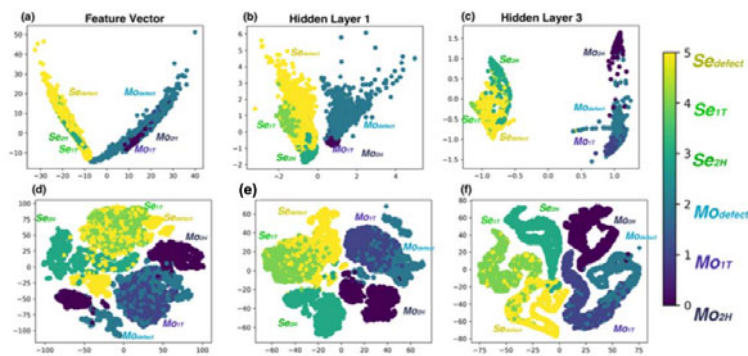


Figure 3: (a-c) shows PCA and Figure (d-f) shows t-SNE analysis of the input feature vector and the learned representation of data in hidden layer 1 and 3 of neural network

Neural Network training and Feature learning

The training accuracy of the model after 200 epoch is 93%, where each atom is represented by a 436-dimension vector. Since the hyper-parameters of the symmetry function are randomly chosen, many attributes of the feature vector will be highly correlated or irrelevant. Removing these irrelevant features will not only make the computation faster but also it will make the model more compact and increase its interpretability.

For multi-layer neural network, relevance of each feature can be computed by calculating its gradient with respect to the input feature. Figure 4a-b shows the gradient of the 10 features whose gradient either changed the most (figure 4a) or the least (figure 4b). We have re-trained another model using only a 50-dimension feature, which consists of only the most significant features with maximum gradient change or the least significant ones. The training accuracy of these two models is shown in figure 4c-d, where we can see that model constructed using only the 50 most significant features have an accuracy of 91% which is just 2% less than model accuracy created using entire feature vector. On the other hand, model constructed using less significant features have only a training accuracy of 58%.

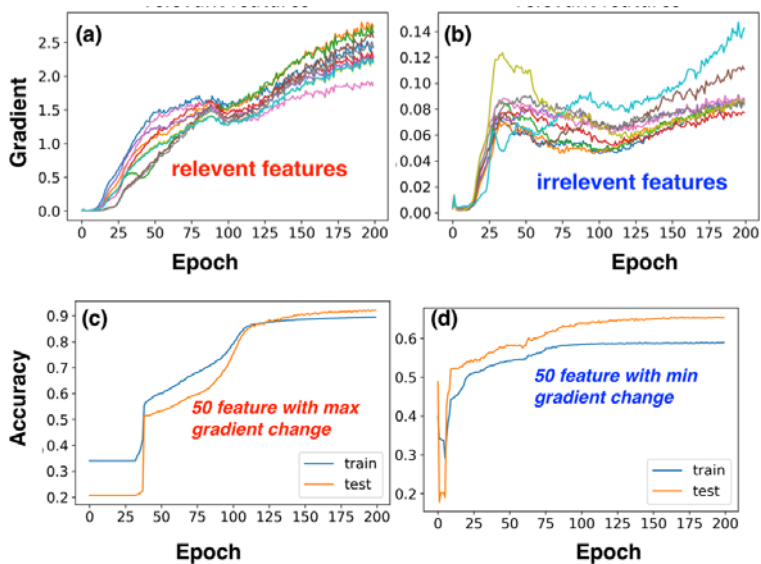


Figure 4: (a) 10 features whose gradient change is maximum during training using full feature vector set. (b) 10 features whose gradient change is minimum during training. (c) and (d) show training and validation accuracy of models, which are training with 50 features with maximum gradient change or minimum gradient change in the original feature space.

Hyper-parameter tuning and data visualization

Figure 5a visualizes the fracture data set where structure type for each atom is predicted by the NN model. We observe that it classifies the 1T and 2H crystal structure accurately; however, it mislabels some of the defects as 1T and vice versa. As we mentioned earlier, in the t-SNE visualization of the hidden layer (figure 3f), clusters of 2H and 1T crystal structures are completely separated whereas the boundary between the 1T phase and defects is not clear. Namely, for some atoms, their local neighbor environment is similar in 1T phase and defects, which causes this misclassification. For example, for some defects Mo-Se-Mo bond angle (BA) does not deviate much from the mean BA value of crystals as Mo-Se bond length is slightly above bond length of crystals, which is 2.9 Å. Since, in this model a cutoff distance of 8 Å is used for symmetry function calculation, and hence damping factor will be very small for these defects. This creates noise or ambiguity in the angular symmetry function value as it will exactly same in the 1T phase and the defects. Figure 5b visualizes the same fracture data but using a reduced cutoff distance (3 Å) for angular symmetry function. This tuning of the cutoff distance increased the model accuracy up to 95% as we can observe that misclassification between 1T phase and defects has significantly been reduced.

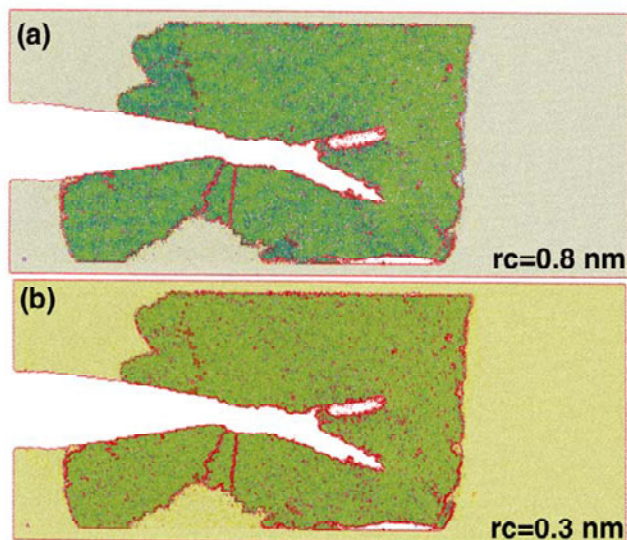


Figure 5: Visualization of validation set of a MoWSe₂ fracture data for modes tuned with two different cutoffs for angular symmetry function calculation (a) $r_{\text{cut}}=8 \text{ \AA}$ and (b) $r_{\text{cut}}=3 \text{ \AA}$.

CONCLUSIONS

In summary, we observed that neural network-based analysis code can automatically learn the relevant information from the simulation data. Multiple hidden layers help better separate the phases of MoWSe₂ compared to shallow network. A careful tuning of various model hyperparameters (*e.g.* number of radial and angular symmetry functions and cutoff distance for symmetry-function calculation) increases the model performance up to 95% while keeping the model complexity very small. The analysis of the learned features of the network gives us further information about the parameters of the symmetry functions that are important to separate different phases.

ACKNOWLEDGMENTS

This work was supported as part of the Computational Materials Sciences Program funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, under Award Number DE-SC0014607. All simulations were performed at the Center for High

References:

- [1] K. Nishimura, R. K. Kalia, A. Nakano and P. Vashishta, *Appl. Phys. Lett.* 92, 161904, (2008).
- [2] P. Vashishta, R. K. Kalia, A. Nakano and J. P. Rino, *J. Appl. Phys.* 101, 103515, (2007).
- [3] P. Vashishta, R. K. Kalia, A. Nakano and J. P. Rino, *J. Appl. Phys.* 109, 033514 (2011).
- [4] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev and A. Walsh, *Nature* 559, 547 (2018).
- [5] R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi and C. Kim, *npj Comput. Mater.* 3, 54, (2017).
- [6] K. Rajan, *Mater. Today* 8, 38 (2005).
- [7] R. LeSar, *Stat. Anal. Data Min.* 1, 372 (2009).
- [8] T. Mueller, A. G. Kusne and R. Ramprasad, *Rev. Comput. Chem.* (2015).
- [9] J. Lee, A. Seko, K. Shitara, K. Nakayama and I. Tanaka, *Phys. Rev. B* 93, 115104 (2016).
- [10] T. Gu, W. Lu, X. Bao and N. Chen, *Solid State Sci.* 8, 129 (2006).
- [11] C. Kim, G. Pilania and R. Ramprasad, *Chem. Mater.*, 28, 1304 (2016).
- [12] C. Kim, G. Pilania and R. Ramprasad, *J. Phys. Chem. C* 120, 14575 (2016).
- [13] Z. Zhaochun, P. Ruiwu and C. Nianyi, *Mater. Sci. Eng. B* 54, 149 (1998).
- [14] T. D. Huan, A. Mannodi-Kanakkithodi and R. Ramprasad, *Phys. Rev. B* 92, 014106 (2015).
- [15] G. Pilania, J. E. Gubernatis and T. Lookman, *Comput. Mater. Sci.* 129, 156 (2017).
- [16] A. Mannodi-Kanakkithodi, G. Pilania, T. D. Huan, T. Lookman and R. Ramprasad, *Sci. rep.* 6, 20952 (2016).
- [17] E. D. Cubuk et al., *Science* 358, 1033 (2017).
- [18] J. Behler and M. Parrinello, *Phys. Rev. Lett.*, 98, 146401 (2007).
- [19] Z. Li, J. R. Kermode and A. De Vita, *Phys. Rev. Lett.* 114, 096405 (2015).
- [20] T. D. Huan, R. Batra, J. Chapman, S. Krishnan, L. Chen and R. Ramprasad, *npj Comput. Mater.* 3, 37 (2017).
- [21] A. Apte, V. Kochat, P. Rajak, A. Krishnamoorthy, P. Manimunda, J. A. Hachtel, J. C. Idrobo, S. A. Syed Amanulla, P. Vashishta, A. Nakano, R. K. Kalia, C. S. Tiwary and P. M. Ajayan, *ACS Nano* 12, 3468 (2018).
- [22] L. van der Maaten and G. Hinton, *J. Mach. Learn. Res.* 9, 2579 (2008).