# Supplementary Information: Modeling crystal defects using defect-informed neural networks

Ziduo Yang[1,2], Xiaoqing Liu[1], Xiuying Zhang[1], Pengru Huang[3*], Kostya S. Novoselov[3] and Lei Shen[1,4*]

[1]Department of Mechanical Engineering, National University of Singapore, 9 Engineering Drive 1, 117575, Singapore.
[2]Department of Electronic Engineering, College of Information Science and Technology, Jinan University, Guangzhou, 510632, China.
[3]Institute for Functional Intelligent Materials, National University of Singapore, 4 Science Drive 2, 117544, Singapore.
[4]National University of Singapore (Chongqing) Research Institute, Chongqing, 401123, China.

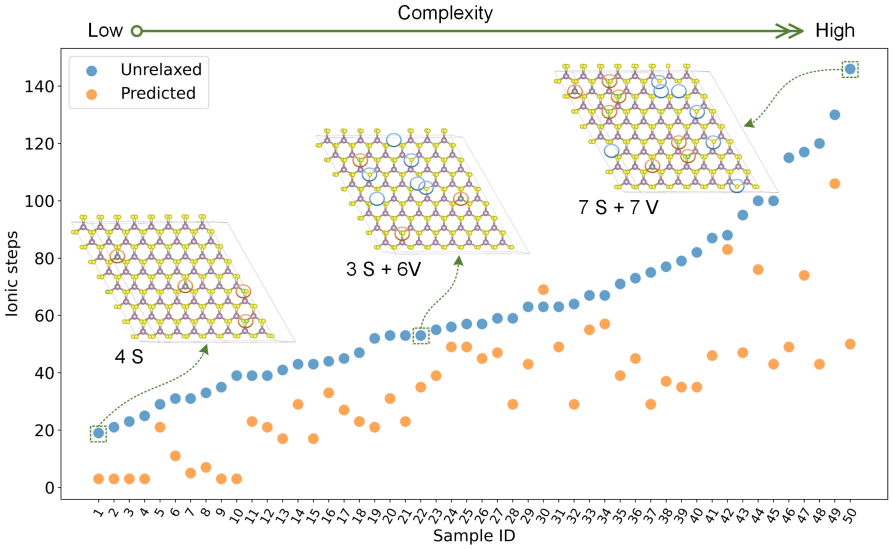
*Corresponding author(s). E-mail(s): pengru@nus.edu.sg; shenlei@nus.edu.sg;

**Suppl. Table 1** Comparison of computational cost between DefiNet and DeepRelax, measured in GPU time (seconds) using an NVIDIA RTX A6000

| Model | $MoS_2$ | $WSe_2$ | h-BN | GaSe | InSe | BP |
|---|---|---|---|---|---|---|
| DeepRelax | 0.4581 | 0.4378 | 0.4546 | 0.4534 | 0.4571 | 0.4585 |
| DefiNet | 0.0156 | 0.0167 | 0.0136 | 0.0128 | 0.0107 | 0.0347 |

# Suppl. Note 1. Computational cost

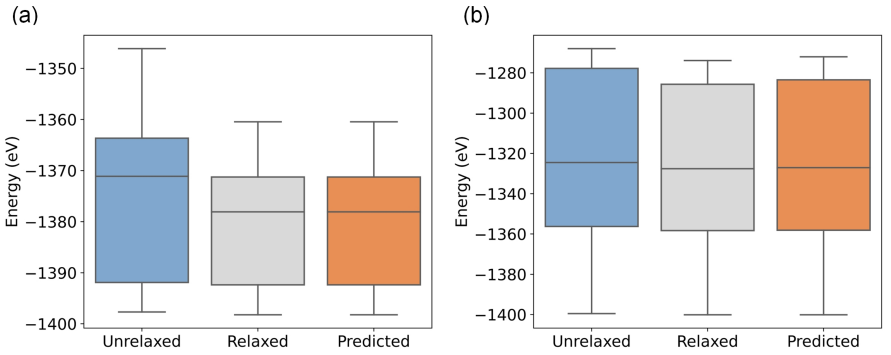
Suppl. Table 1 presents the GPU time comparison between DefiNet and DeepRelax, showing that DefiNet is approximately 26.2 times more computationally efficient than DeepRelax. The evaluation was conducted using an NVIDIA RTX A6000 GPU.




**Suppl. Fig. 1** Comparison of the number of DFT ionic steps required to relax structures starting from the initial unrelaxed configurations and from the DefiNet-predicted structures for high-density defects. The samples are sorted based on the number of ionic steps required by the unrelaxed structures for better observation.

# Suppl. Note 2. DFT validation

For the high-density defect scenario, the results shown in Suppl. Fig. 1 indicate that using DefiNet-predicted structures as starting points reduces the computational effort required for DFT relaxation by approximately 47%. While this reduction is less significant than in the low-density defect scenario, it can be attributed to two factors: First, the sample size for the high-density defect dataset is limited. Second, the patterns of high-density defects are more




**Suppl. Fig. 2** Box plots of energy distributions for unrelaxed, DFT-relaxed, and DefiNet-relaxed structures for (a) low-density defect configurations and (b) high-density defect configurations. Both DFT-relaxed and DefiNet-predicted structures show similar medians and spreads, demonstrating the accuracy of DefiNet in identifying energetically favorable structures.

complex than those of low-density defects. The large configuration space in high-density defect scenarios may limit DefiNet's generalization ability.

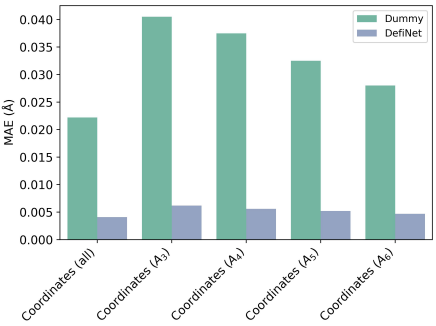
Moreover, Fig. 2(a) and (b) present box plots of energy distributions for low- and high-density configurations, respectively. In both scenarios, the total energies of the DefiNet-predicted structures closely match those of fully DFT-relaxed structures, indicating that DefiNet effectively predicts energetically favorable configurations.

## Suppl. Note 3. Scalability evaluation

To evaluate the scalability of DefiNet in high-density defect scenario, we generated 50 high-density defect configurations of $MoS_2$ using a $12 \times 12$ supercell containing 432 atoms. This represents a significant scale-up from the $8 \times 8$ supercells (192 atoms) used during training.

We assessed DefiNet's ability to predict the relaxed structures of these larger systems and compared its performance against the Dummy model, which serves as a baseline by outputting the initial unrelaxed structures. The results, presented in Suppl. Fig. 3, demonstrate that DefiNet maintains high accuracy even at this increased scale. Specifically, DefiNet achieves improvements in coordinate MAE over the Dummy model by 81.53% for all atoms, and similarly high improvements of 84.69%, 85.06%, 84.00%, and 83.21% for atoms within 3 Å, 4 Å, 5 Å, and 6 Å of the defect sites, respectively.

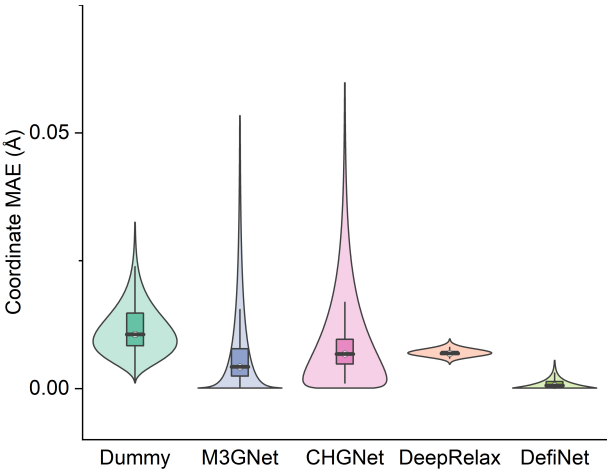
These substantial improvements indicate that DefiNet effectively captures local atomic interactions and defect-induced distortions, which are essential for accurate structural relaxation, regardless of the overall system size.



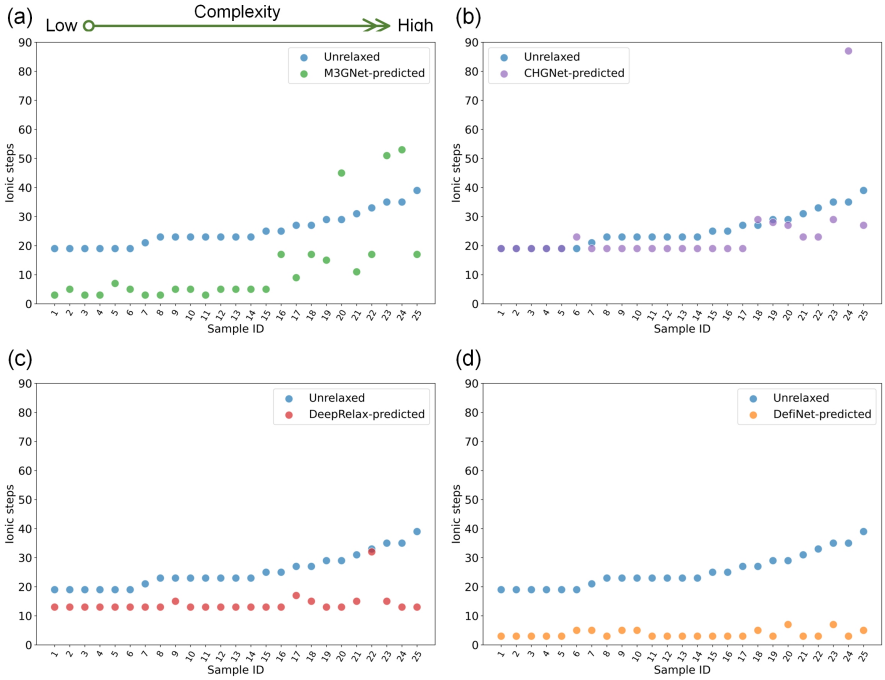

**Suppl. Fig. 3** Scalability evaluation of a model trained on an $8 \times 8$ supercell of $MoS_2$ with high-density defects, tested on a $12 \times 12$ supercell system. $A_3$, $A_4$, $A_5$, and $A_6$ represent MAE calculations using only atoms within 3Å, 4Å, 5Å, and 6Å radii around defect sites, respectively. DefiNet achieves improvements in coordinate MAE over the Dummy model by 81.53% for all atoms, and similarly high improvements of 84.69%, 85.06%, 84.00%, and 83.21% for atoms within 3 Å, 4 Å, 5 Å, and 6 Å of the defect sites, respectively.

**Suppl. Table 2** Performance of two ML-potential models measured by energy MAE (eV), forces MAE (eV/Å), and stress MAE (GPa). The evaluation is conducted on the test set.

| Model | Energy | Forces | Stress |
|---|---|---|---|
| M3GNet | 0.003 | 0.007 | 0.012 |
| CHGNet | 0.036 | 0.005 | 0.013 |




**Suppl. Fig. 4** Comparison of coordinate MAE (Å) between DefiNet and other models using violin plots.




**Suppl. Fig. 5** DFT validations using ML-predicted structures as initial configurations. (a) Unrelaxed vs. M3GNet-predicted, (b) Unrelaxed vs. CHGNet-predicted, (c) Unrelaxed vs. DeepRelax-predicted, and (d) Unrelaxed vs. DefiNet-predicted.

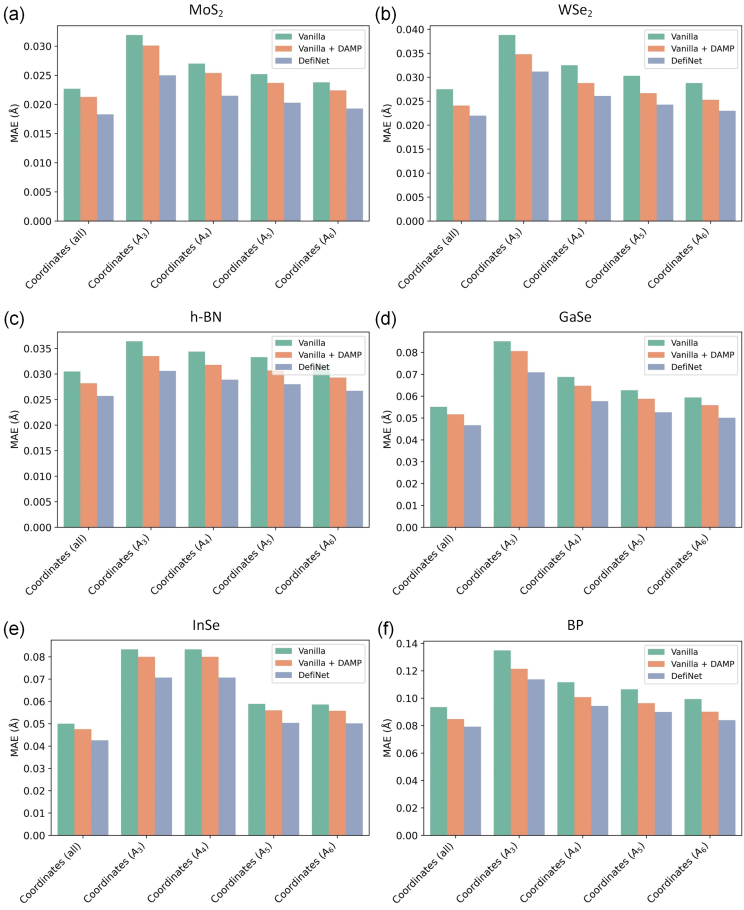
# Suppl. Note 4. Comparison to ML-potential relaxation

To demonstrate the superiority of DefiNet, we compare it with two popular ML-potential models, M3GNet [1] and CHGNet [2]. We implemented these models using the source code provided by the original researchers, available at https://github.com/materialsvirtuallab/matgl and https://github.com/CederGroupHub/chgnet/tree/main, respectively, with default settings. All models were trained, validated, and tested on the same dataset, with 4,746 training samples, 593 validation samples, and 594 test samples. For training the ML-potential models, we sampled three structures—initial, intermediate, and final ionic steps—from each trajectory. These structures include detailed information on energy, forces, and stress. The test set performance of the two models is listed in Suppl. Table 2. These trained models serve as surrogate models, iteratively approximating physical quantities such as energies, forces, and stresses.

Suppl. Fig. 4 presents the structural relaxation results, showing two key observations. First, DefiNet significantly outperforms the other methods ($p < 0.001$, t-test) in terms of coordinate MAE. Both single-step models, DefiNet and DeepRelax, demonstrate greater robustness than the ML-potential models,

exhibiting lower deviation in coordinate MAE. In contrast, the ML-potential models show higher deviations in coordinate MAE, likely due to their iterative nature, where repeated estimation of energy, forces, and stresses can lead to error accumulation and, in some cases, inaccurate relaxation.

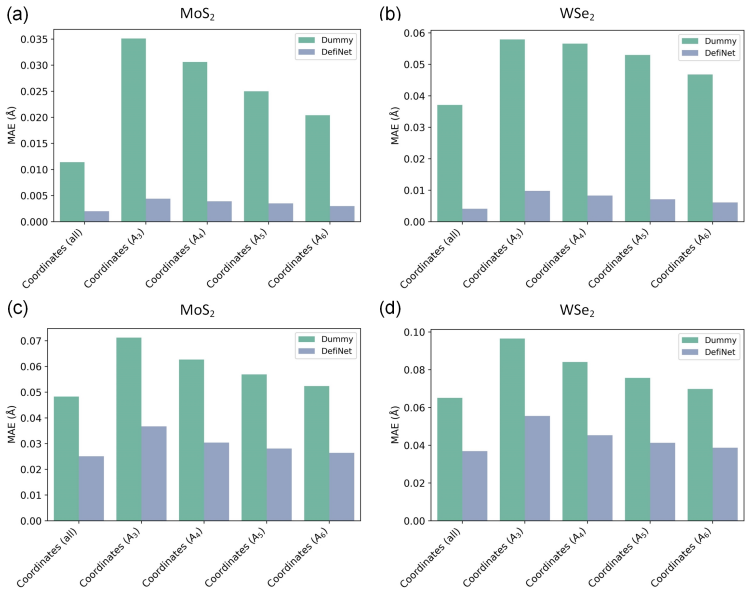
We further selected 25 samples from the test set for DFT validation. Using the predicted structures from each model as initial configurations significantly reduces the number of required ionic steps, as shown in Suppl. Fig. 5. The DFT validation results align with the performance trends observed in Suppl. Fig. 4, where DefiNet consistently outperforms other methods, maintaining excellent robustness regardless of structural complexity.




**Suppl. Fig. 6** Ablation study on high-density defect datasets evaluating the impact of Universal Defect Representation with Defect-Aware Message Passing (DAMP) and Defect-Aware Coordinate Updating (DACU). Results are shown for (a) $MoS_2$, (b) $WSe_2$, (c) h-BN, (d) GaSe, (e) InSe, and (f) BP, demonstrating the contributions of each component to the model's performance. $A_3$, $A_4$, $A_5$, and $A_6$ represent MAE calculations using only atoms within 3Å, 4Å, 5Å, and 6Å radii around defect sites, respectively.

# Suppl. Note 5. Ablation study

Suppl. Fig. 6 demonstrates that removing either component reduces DefiNet's performance on high-density datasets, underscoring their importance. The Universal Defect Representation with DAMP effectively models all defect types and their interactions, while the DACU modules (RPV2ΔCoord and Vec2ΔCoord) ensure precise prediction of atomic movements.




**Suppl. Fig. 7** Transferability evaluation of DefiNet between high-density and low-density defect datasets for $MoS_2$ and $WSe_2$. (a)-(b) DefiNet trained on high-density defect datasets and tested on low-density defect structures for (a) $MoS_2$ and (b) $WSe_2$. (c)-(d) DefiNet trained on low-density defect datasets and tested on high-density defect structures for (c) $MoS_2$ and (d) $WSe_2$. $A_3$, $A_4$, $A_5$, and $A_6$ represent MAE calculations using only atoms within 3Å, 4Å, 5Å, and 6Å radii around defect sites, respectively.

# Suppl. Note 6. Transferability evaluation

Transferability is a key aspect of a robust ML model, reflecting its ability to generalize across different datasets and conditions. In the context of defect structure relaxation, a transferable model can be trained on defects of a certain density and still perform reasonably when applied to structures with different defect densities. To assess DefiNet's transferability, we conducted experiments in two scenarios:

First, we trained DefiNet on high-density defect datasets and tested it on low-density defect structures. The high-density datasets contain complex

defect-defect interactions due to the proximity of multiple defects, while low-density datasets primarily involve isolated defects. As shown in Suppl. Fig. 7(a)-(b), DefiNet demonstrates excellent transferability in this scenario. The model maintains high accuracy when predicting the relaxed structures of low-density defect systems, achieving significant improvements over the Dummy model in coordinate MAE for all atoms and those near defect sites. This suggests that training on complex, high-density defect environments equips the model with a comprehensive understanding of defect interactions, which generalizes well to simpler, low-density cases.

Conversely, we trained DefiNet on low-density defect datasets and tested it on high-density defect structures. In this case, the model encounters defect configurations with interactions it has not seen during training. The results, presented in Suppl. Fig. 7(c)-(d), show that while DefiNet still outperforms the Dummy model, the improvements are more modest compared to the previous scenario. This result is expected, as the model trained on low-density defects lacks exposure to the complex defect-defect interactions present in high-density systems.

Overall, these experiments indicate that DefiNet trained on high-density defects generalizes well to low-density scenarios, but training on low-density defects provides limited transferability to high-density cases. This highlights the importance of including complex defect interactions during training to enhance the model's generalizability.

## Supplementary references

[1] Chen, C., Ong, S.P.: A universal graph deep learning interatomic potential for the periodic table. Nature Computational Science **2**(11), 718–728 (2022)

[2] Deng, B., Zhong, P., Jun, K., Riebesell, J., Han, K., Bartel, C.J., Ceder, G.: Chgnet as a pretrained universal neural network potential for charge-informed atomistic modelling. Nature Machine Intelligence, 1–11 (2023)