nature portfolio

Peer Review File

Scalable Crystal Structure Relaxation Using an Iteration-Free Deep Generative Model with Uncertainty Quantification

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

In this manuscript, based on the framework of PaEGNN, the authors developed a model named DeepRelax for the usage of crystal relaxation. Benchmarked by the MAE of the structural information between the predicted structure and relaxed structure on three datasets, this model shows great accuracy and efficiency compared to other available models of the same usage. While several issues need to be addressed before the manuscript published on the Nature Communications.

1. The benchmarked MAE of bond length showed in Table 2 and Table 3 are both an average of all types. However, for structural relaxation, usually a certain type of bond is crucial, which varies a lot during the relaxation and thus determine the final relaxed structure. For example, the inter-layer distance of the layered materials may have a significant change after being fully relaxed, while keeping the intra-layer bonds nearly unchanged. The bond length error of the “crucial” ones is very likely to be submerged by the average of all. The authors are suggested to test the model on some layered vdW crystals and have a more detailed bonding length MAE statistics of the representative bonding pairs.

2. According to Fig.5 (b), the energy profile of the predicted structure is fair close to the fully relaxed ones and apart from the unrelaxed ones, in order to demonstrate the effectiveness of the model. While, on one hand, this test is on the dataset of X-Mn-O, in which the structures contain a substitutional element X making the lattice parameter varies a lot after relaxation. It is curious if the model is still of high effectiveness on those structures with a relative rational initial unrelaxed structure. On the other hand, usually in a process of structural relaxation, the energy drops significantly at first few steps but takes many steps to minimize the energy after it. So that the authors are suggested to conduct the DFT validations on more kind of datasets and also take the residual optimizing steps for the predicted structures as an indicator as well as the energy.

3. Similar to the former question, I wonder if this model performs well on the meta-stable structures and those structures with flat potential energy surface (e.g. crystal with intrinsic vacancies but of different distribution).

Reviewer #2 (Remarks to the Author):

Atomic structural relaxation is usually the first step and foundation for further computational analysis of properties in computational chemistry, physics, materials science, and medicine.
In this work, the authors develop a scalable, universal, and trustworthy deep generative model for fast structural relaxation, DeepRelax, by learning from the structural relationships between unrelaxed and relaxed crystals. DeepRelax requires only the initial crystal structures to predict equilibrium structures in just a few hundred milliseconds on a single GPU. Furthermore, DeepRelax

can efficiently handle multiple crystal structures in parallel by organizing them into mini-batches for simultaneous processing. This capability is especially advantageous in large-scale virtual screening, where rapid assessment of numerous unknown crystal configurations is essential. Most importantly, DeepRelax employs an uncertainty quantification method to assess the reliability of the predicted structure, enabling the identification and exclusion of predictions that are likely to be unreliable. These results are very interesting and I would like to recommend the pulication of this work in Nat. Comm. after the authors address the following issues:
(1) "EDG" is not defined before use. I would suggest to define it also in the caption of Fig. 1.
(2) Yoon [Ref. 16] and Kim [Ref. 15] conceptually proposed direct ML approaches to predict the final relaxed structures directly from their unrelaxed versions. Can the authors compare the accuracy of the current method with the methods proposed by Yoon and Kim?

Reviewer #2 (Remarks on code availability):

Although I did not try it, it has detailed instructions.

1 Response to Reviewer 1

Comment 1.1:

In this manuscript, based on the framework of PaEGNN, the authors developed a model named DeepRelax for the usage of crystal relaxation. Benchmarked by the MAE of the structural information between the predicted structure and relaxed structure on three datasets, this model shows great accuracy and efficiency compared to other available models of the same usage. While several issues need to be addressed before the manuscript published on the Nature Communications.

The benchmarked MAE of bond length showed in Table 2 and Table 3 are both an average of all types. However, for structural relaxation, usually a certain type of bond is crucial, which varies a lot during the relaxation and thus determine the final relaxed structure. For example, the inter-layer distance of the layered materials may have a significant change after being fully relaxed, while keeping the intra-layer bonds nearly unchanged. The bond length error of the "crucial" ones is very likely to be submerged by the average of all. The authors are suggested to test the model on some layered vdW crystals and have a more detailed bonding length MAE statistics of the representative bonding pairs.

Our reply to 1.1:

We completely agree with you that in some unique materials, such as layered vdW crystals, a certain type of bond is crucial during relaxation. In the revision, we have relaxed 58 layered vdW crystals covering 29 elements using van der Waals corrections, parameterized within the DFT-D3 Grimme method. This dataset is divided into training, validation, and test sets in an 8:1:1 ratio. Given the small sample size, we employ transfer learning, utilizing a model pre-trained on the Materials Project dataset. Moreover, we have introduced new metrics for our analyses, focusing specifically on intra-layer chemical bond lengths and crucial vdW inter-layer gaps. The detailed results are as follows.

Table S1 shows the inter-layer vdW distances for the unrelaxed, DFT-D3 relaxed, and DeepRelax predicted structures within the six vdW crystals from the test set (S_4Ga_4, In_4Se_4, Te_4B_4, W_2Se_4, As_2I_6, and S_4Tl_4). As can be seen, the inter-layer vdW distances of the predicted structures closely match those of the DFT-D3 relaxed structures, highlighting the effectiveness of our transferred DeepRelax on layered vdW crystals.

Based on your insightful suggestion, "For structural relaxation, usually a certain type of bond is crucial, which varies a lot during the relaxation and thus determines the final relaxed structure," we have also developed new local metrics to demonstrate DeepRelax's performance in predicting structures of vdW crystals. This metric analysis is conducted on the bond length MAE for representative bonding pairs. As we know, the strong chemi-

cal bond within the intra-layer varies less, while the weak "vdW bond" varies significantly during relaxation. We define Δ as the bond length difference between the unrelaxed and relaxed structures. Thresholds are set at 0.1 Å, 0.2 Å, 0.3 Å, 0.4 Å, and 0.5 Å. Bonding pairs where Δ exceeds these thresholds are filtered out, effectively removing 67%, 77%, 84%, 85%, and 88% of the bonding pairs that show lesser changes during relaxation, respectively. As shown in Table S2, DeepRelax achieves lower bond length MAE across all thresholds compared to the Dummy model, also demonstrating its precision in predicting structural changes in layered vdW crystals.

Table S1 Inter-layer distances (Å) for unrelaxed, DFT-D3-relaxed, and DeepRelax-predicted structures of six vdW crystals from the test set

Formula	S_4Ga_4	In_4Se_4	Te_4B_4	W_2Se_4	As_2I_6	S_4Tl_4
Unrelaxed	3.600	3.800	4.120	3.800	3.960	3.600
DFT-D3-relaxed	3.147	3.078	3.855	3.175	2.875	2.794
DeepRelax-predicted	3.143	3.097	3.600	3.263	2.830	2.804

Table S2 MAE statistics for bond length of representative bonding pairs, with thresholds set at 0.1 Å, 0.2 Å, 0.3 Å, 0.4 Å, and 0.5 Å, filtering out 67%, 77%, 84%, 85%, and 88% of the less changing bonds during relaxation, respectively

Model	$\Delta > 0.1$	$\Delta > 0.2$	$\Delta > 0.3$	$\Delta > 0.4$	$\Delta > 0.5$
Dummy	0.493	0.535	0.685	0.707	0.794
DeepRelax	**0.152**	**0.163**	**0.146**	**0.144**	**0.164**

Changes made:

[Page 16, Main Text; Tables S1 and S2, Supplementary Materials]

Layered vdW crystals are of significant interest in the field of materials science and nanotechnology because of their unique tunable structures, such as twisting and sliding configurations. One notable characteristic of these crystals is that the weak inter-layer vdW force may significantly change upon full relaxation, while the strong intra-layer chemical bonds undergo relatively small changes.

To demonstrate the performance of our DeepRelax model on this type of crystal, we performed DFT relaxation of 58 layered vdW crystals covering 29 elements using van der Waals corrections, parameterized within the DFT-D3 Grimme method. Given the small sample size, we employ transfer learning, utilizing a model pre-trained on the Materials Project dataset.

Table S1 shows the inter-layer distances for the unrelaxed, DFT-D3-relaxed, and DeepRelax-predicted structures of six vdW layered crystals. The inter-layer distances of the predicted structures closely match those of the relaxed structures, highlighting the

effectiveness of transferred DeepRelax on layered vdW crystals. Furthermore, an analysis of the MAE in bond length for representative bonding pairs, detailed in **SM** Table S2, further demonstrates DeepRelax's precision in predicting structural changes in layered vdW crystals.

Comment 1.2:

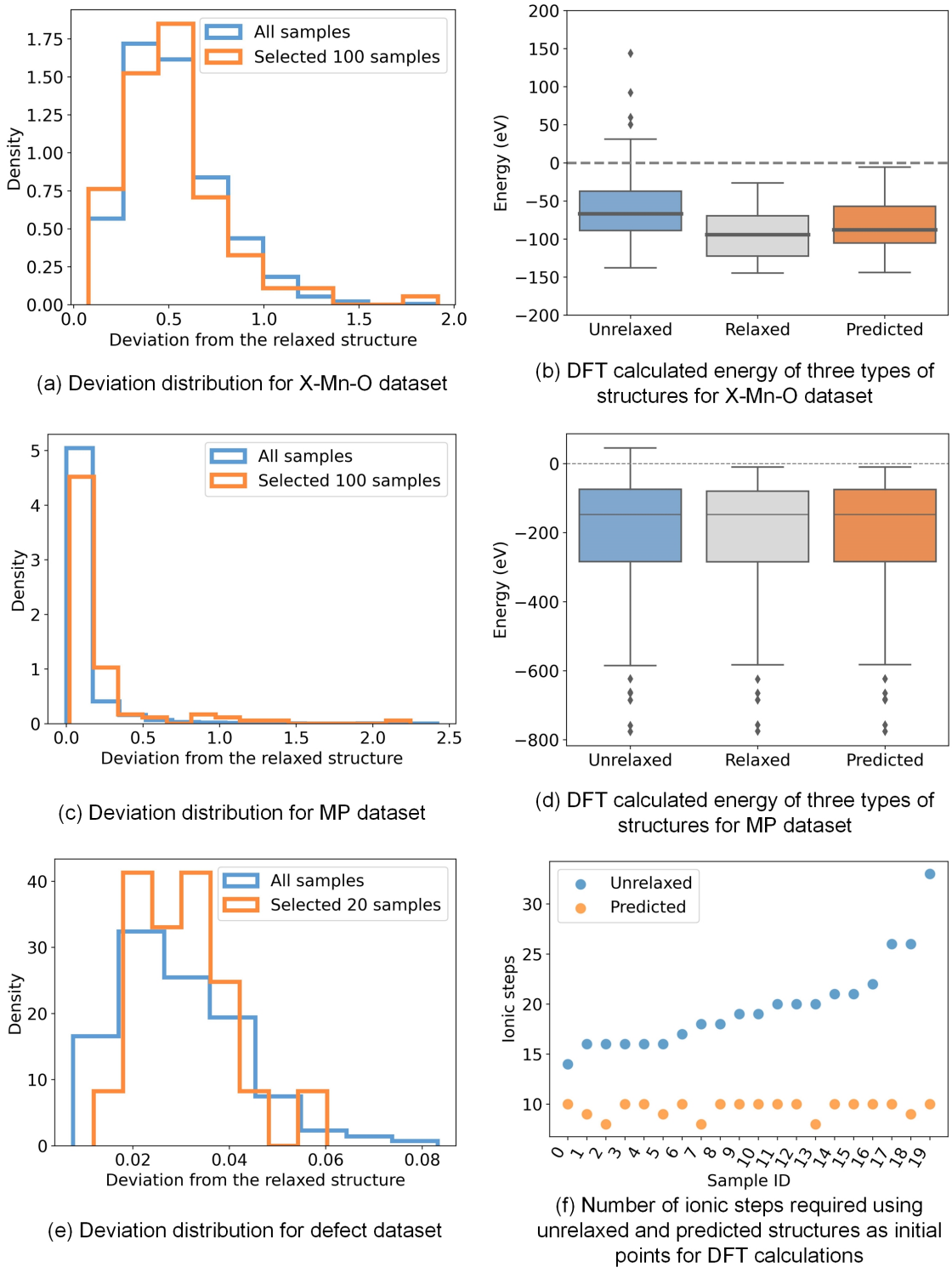
According to Fig.5 (b), the energy profile of the predicted structure is fair close to the fully relaxed ones and apart from the unrelaxed ones, in order to demonstrate the effectiveness of the model. While, on one hand, this test is on the dataset of X-Mn-O, in which the structures contain a substitutional element X making the lattice parameter varies a lot after relaxation. It is curious if the model is still of high effectiveness on those structures with a relative rational initial unrelaxed structure. On the other hand, usually in a process of structural relaxation, the energy drops significantly at first few steps but takes many steps to minimize the energy after it. So that the authors are suggested to conduct the DFT validations on more kind of datasets and also take the residual optimizing steps for the predicted structures as an indicator as well as the energy.

Our reply to 1.2:

These two insightful comments and suggestions have been well-taken in the revised manuscript. To answer your first concern about the model's effectiveness on structures with relatively stable initial unrelaxed configurations, we have tested our model on 100 randomly selected samples from the Materials Project database with relatively rational initial unrelaxed structures. Figures 5(c)-(d) show that our DeepRelax model can also provide accurate DFT-level structures when the initial unrelaxed structures are close to the final structures.

Regarding your second suggestion on "...taking the residual optimizing steps for the predicted structures as an indicator...", we have tested the DeepRelax model on 20 randomly selected samples from our 2D materials point-defect database. Figures 5(e)-(f) show the total ionic steps in DFT calculations, starting from unrelaxed structures (blue) and predicted structures (orange), respectively. It is clearly shown that DeepRelax can significantly speed up the process of structural relaxation.

Overall, the new DFT validation results in the revised Figure 5 show the robustness of DeepRelax's applicability across different scenarios. Specifically, this model demonstrates its advantage in scenarios involving complex systems, such as vdW crystals, defect materials, and surface/interfacial structures. Although DeepRelax cannot entirely replace DFT relaxation, starting from DeepRelax-predicted structures significantly reduces the number of DFT ionic steps required, especially for complex structures or systems.



(a) Deviation distribution for X-Mn-O dataset

(b) DFT calculated energy of three types of structures for X-Mn-O dataset

(c) Deviation distribution for MP dataset

(d) DFT calculated energy of three types of structures for MP dataset

(e) Deviation distribution for defect dataset

(f) Number of ionic steps required using unrelaxed and predicted structures as initial points for DFT calculations

Figure 5 (a) and (b) illustrate the distributions of deviations and energy for the randomly selected 100 samples from the X-Mn-O dataset, respectively, with deviations measured using MAE in coordinates between the unrelaxed and relaxed structures. (c) and (d) present corresponding analyses for the Materials Project dataset which has relatively rational initial unrelaxed structures. (e) and (f) illustrate the distributions of deviations and the number of DFT ionic steps required for the randomly selected 20 samples from the 2D materials defect dataset, starting from initial unrelaxed structures and DeepRelax-predicted structures, respectively. The samples are sorted based on the number of ionic steps required by the unrelaxed structures for better observation.

Changes made:

[Pages 16-18, Main Text; Figure 5, Main Text]

Usually, the initial crystal structure may deviate from or be close to the final relaxed structure. To demonstrate the efficacy and robustness of DeepRelax, we perform DFT validations on two types of initial structures: those from the X-Mn-O dataset, which exhibit large deviations from the relaxed state, and those from the MP dataset, which are generally closer to their final structures, as illustrated in **SM** Figure S1.

In the first experiment, we evaluated our model's predictive capability under challenging conditions using the X-Mn-O dataset. We filtered out unrelaxed structures from the X-Mn-O test set that are structurally similar to their final counterparts using Pymatgen's `Structure_matcher` function. From the remaining test set ($N = 1007$), we randomly selected 100 samples. Figure 5(a) shows the deviation distribution for the selected unrelaxed structures, which closely aligns with that of the complete test set, thus confirming the representativeness of the selected subset. Subsequently, we employed DeepRelax to predict the relaxed structures for these samples. The calculated energies of the unrelaxed, relaxed, and DeepRelax-predicted structures are given in **Subsection 5.2**. Figure 5(b) shows box plots of the energy distributions for the three structure categories. The energy distributions of the predicted and relaxed structures show similar medians and interquartile ranges, validating the model's accuracy in predicting energetically favorable structures. The MAE in energy is significantly reduced by an order of magnitude from 32.51 to 5.97.

In the second experiment, we tested whether DeepRelax remains effective with structures starting from a relatively rational initial unrelaxed state using the Materials Project dataset. Here, we again randomly selected 100 samples from the test set. Figure 5(c) shows the deviation distribution for these samples. The energies of the unrelaxed, relaxed, and DeepRelax-predicted structures were calculated using DFT. Figure 5(d) shows that the predicted structures feature an energy distribution nearly identical to that of the DFT-relaxed structures, demonstrating the model's effectiveness in handling relatively rational initial unrelaxed structures.

Besides the energy indicator, we further demonstrated our model's effectiveness using the number of residual optimizing steps required for DFT relaxation. Specifically, we randomly selected 20 structures from the test set of the 2D materials point-defect dataset, with their deviation distribution as shown in Figure 5(e). We then conducted DFT calculations starting from the unrelaxed and DeepRelax-predicted structures, respectively. As shown in Figure 5(f), starting DFT relaxation from the DeepRelax-predicted structures significantly reduces the number of required ionic steps.

Comment 1.3:

Similar to the former question, I wonder if this model performs well on the meta-stable structures and those structures with flat potential energy surface (e.g. crystal with intrinsic vacancies but of different distribution).

Our reply to 1.3:

This point has been well-taken in our revision. In response, we have expanded our experiments to include crystals with intrinsic point defects, specifically utilizing MoS2 structures with 5,933 defect configurations within an 8×8 supercell, as cataloged by Huang et al. [2023]. These configurations comprise single, double, and triple-site vacancy and substitution defects with different distributions. We divide the dataset into training, validation, and test sets in a ratio of 8:1:1. Given that structural variations between unrelaxed and relaxed defect structures are mainly localized near the defect sites, we adopt localized MAE statistics for a more precise evaluation of DeepRelax's performance. Specifically, we use Ax to denote atoms within an x Å radius of the defect sites and B_x for the associated bonds. We set x to 3, 4, 5, and 6 in our experiment. Figure S3 presents the experimental results, which demonstrate a notably lower MAE in both atom coordinates and bond lengths for DeepRelax compared to the Dummy model, thereby underscoring DeepRelax's efficacy in defect structural relaxation. Additional DFT calculations are conducted to further validate the speed of DFT relaxation from the DeepRelax-predicted structures, as illustrated in Figure 5(f) in our reply to your Comment 1.2.

Regarding meta-stable structures, as we currently lack a dataset containing meta-stable structures, testing and refining the model under these conditions remain objectives for future work.

Changes made:

[Pages 16, Main Text]

Most crystals have intrinsic defects. Moreover, controllable defect engineering facilitates property modifications and introduces new functionalities in crystalline materials[61-63]. However, DFT-level defect analysis is challenging due to the need for large supercells and a high level of theoretical precision[13]. Recent work by Mosquera-Lois et al.[13] and Jiang et al.[64] demonstrates that ML surrogate models can accelerate the relaxation of defect structures. Nonetheless, training or fine-tuning such ML models requires comprehensive data including energy, forces, stress, etc.

In this subsection, we briefly illustrate the potential application of our direct prediction model, DeepRelax, to crystal structures with neutral point defects. We employ MoS_2 structures with a low defect concentration, including 5,933 different defect configurations within an 8×8 supercell, as cataloged by Huang et al.[63], to evaluate DeepRelax. Figure S3 demonstrates a notably lower MAE in both atom coordinates and bond lengths for

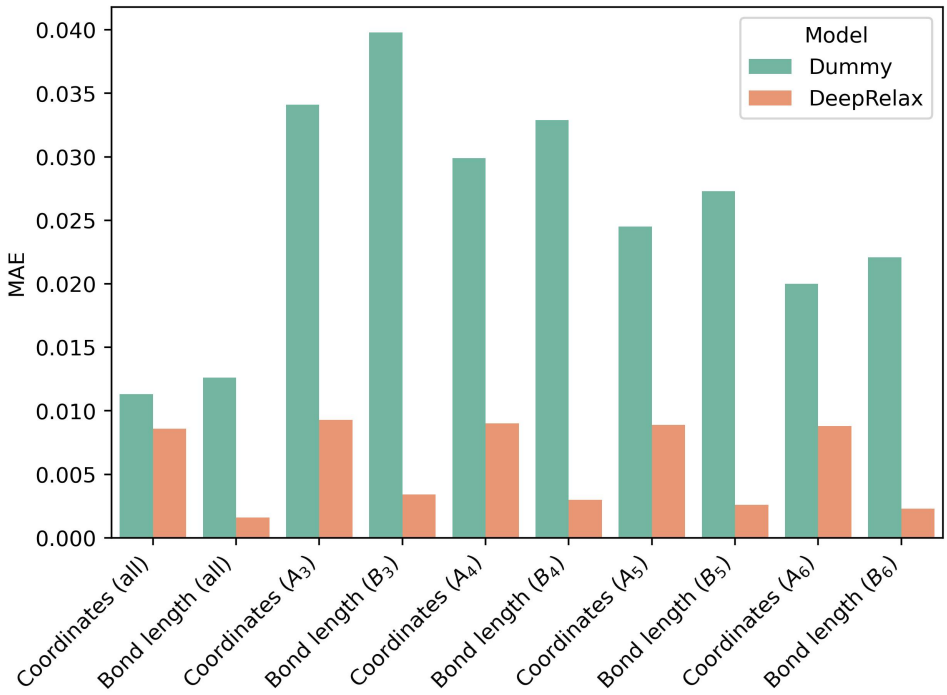


Figure S3 Comparative results of DeepRelax and the Dummy model on defect calculations, detailing MAE for all-atom coordinates, all-bond lengths, local-atom coordinates, and local-bond lengths.

DeepRelax compared to the Dummy model, thereby underscoring DeepRelax's efficacy in defect structure calculations.

2 Response to Reviewer 2

Comment 2.1:

Atomic structural relaxation is usually the first step and foundation for further computational analysis of properties in computational chemistry, physics, materials science, and medicine. In this work, the authors develop a scalable, universal, and trustworthy deep generative model for fast structural relaxation, DeepRelax, by learning from the structural relationships between unrelaxed and relaxed crystals. DeepRelax requires only the initial crystal structures to predict equilibrium structures in just a few hundred milliseconds on a single GPU. Furthermore, DeepRelax can efficiently handle multiple crystal structures in parallel by organizing them into mini-batches for simultaneous processing. This capability is especially advantageous in large-scale virtual screening, where rapid assessment of numerous unknown crystal configurations is essential. Most importantly, DeepRelax employs an uncertainty quantification method to assess the reliability of the predicted structure, enabling the identification and exclusion of predictions that are likely to be unreliable. These results are very interesting and I would like to recommend the pulication of this work in Nat. Comm.
after the authors address the following issues:
"EDG" is not defined before use. I would suggest to define it also in the caption of Fig. 1.

Our reply to 2.1:

Thank you for highlighting this oversight. We have explicitly defined "EDG" (Euclidean distance geometry) along with "PaEGNN" (periodicity-aware equivariant graph neural network) in the caption of the revised Figure 1. This addition will ensure that all terms are clearly understood, enhancing the readability of our figures.

Comment 2.2:

Yoon [Ref. 16] and Kim [Ref. 15] conceptually proposed direct ML approaches to predict the final relaxed structures directly from their unrelaxed versions. Can the authors compare the accuracy of the current method with the methods proposed by Yoon and Kim?

Our reply to 2.2:

We completely agree with you that comparing our DeepRelax model with closely related models, specifically Cryslator Kim et al. [2023] and DOGSS Yoon and Ulissi [2020], is beneficial for evaluating model performance. We have included a detailed comparison between DeepRelax and Cryslator in the revised Table 1, demonstrating the superior

performance of DeepRelax. Additionally, the developers of Cryslator reported that their model has limited performance on the Materials Project dataset in their original published paper Kim et al. [2023], while our DeepRelax maintains its effectiveness across this dataset (See in Table 2).

Regarding the comparison with DOGSS, we attempted to conduct a thorough comparative analysis as you suggested. However, we cannot access the necessary resources for this model published in 2020. The source code for DOGSS is currently not available via the provided link in the paper `https://github.com/Open-Catalyst-Project/baselines` or any other official repositories.

Moreover, DOGSS uses an approximation method by only considering the interatomic distance within a structure, while DeepRelax simultaneously predicts both interatomic distances and atomic displacements. This capability, combined with DeepRelax's equivariant representation, explicit handling of periodicity, uncertainty quantification, and ability to predict lattice parameters, significantly extends its applicability and accuracy compared to DOGSS.

We have detailed these comparisons, highlighting the distinct advantages and capabilities of DeepRelax over existing direct ML approaches, in the Supplementary Materials, Section 4.

Table 1 Comparative results of DeepRelax and baseline models on the X-Mn-O dataset, based on MAE of coordinates (Å), bond length (Å), lattice (Å), cell volume (Å^3), and match rate (%) between predicted and ground truth relaxed structures

Model	Coordinates	Bond length	Lattice	Cell volume	Match rate
Dummy	0.314	0.429	0.221	32.8	64.8
PAINN	0.159	0.175	0.066	3.8	81.2
EGNN	0.166	0.189	0.066	4.2	77.5
Cryslator*	0.127	-	-	6.2	83.7
DeepRelax	**0.116**	**0.136**	**0.063**	**3.4**	**84.7**

*The results of Cryslator are taken from Kim et al. [2023]. DeepRelax is evaluated on the same training, validation, and testing sets as Cryslator for a fair comparison.

Changes made:

[Page 3, Main Text]

To address this, Yoon et al.[16] developed a model called DOGSS, and Kim et al.[15] proposed a model named Cryslator. Both conceptually introduce *direct* ML approaches to predict the final relaxed structures from their unrelaxed counterparts.

[Page 21, Main Text]

While there are other *direct* structure-prediction ML methods, such as DOGSS Yoon and Ulissi [2020] and Cryslator Kim et al. [2023], the differences and advantages of our DeepRelax model over these are detailed in **SM Section 4**.

[Pages 5, Supplementary Materials]

Two direct structure-prediction ML models, DOGSS Yoon and Ulissi [2020] and Cryslator Kim et al. [2023], are published for predicting crystal structures. Here, we outline the differences and advantages of our DeepRelax model over these two as follows:

DOGSS, published in 2020, integrates differentiable optimization with graph neural networks to predict relaxed structures. It learns a simple harmonic force field, utilizing a GNN to predict the parameters: pairwise interatomic distances in the relaxed structure $\boldsymbol{r}_e \in \mathbb{R}^{N\times M\times 1}$, and spring constants $\boldsymbol{k} \in \mathbb{R}^{N\times M\times 1}$ for springs connecting atom pairs. Here, N represents the number of atoms, and M denotes the number of nearest neighbor atoms connected by distinct edges. Once these parameters are learned, the model constructs the harmonic potential energy surface (PES) as

$$\hat{E}(\boldsymbol{r}_o(\boldsymbol{x}_o), \boldsymbol{r}_e, \boldsymbol{k}) = \sum_i \sum_j \boldsymbol{k}_{i,j}(\boldsymbol{r}_{o,(i,j)}(\boldsymbol{x}_o) - \boldsymbol{r}_{e,(i,j)})^2 \tag{1}$$

where $\boldsymbol{r}_o(\boldsymbol{x}_o) \in \mathbb{R}^{N\times M\times 1}$ represents pairwise interatomic distances in the unrelaxed structure. DOGSS solves for optimal structures by minimizing this energy:

$$\boldsymbol{x}_{\text{pred}} = \arg \min_{\boldsymbol{x}_o \notin \boldsymbol{x}_{\text{con}}} \hat{E}(\boldsymbol{r}_o(\boldsymbol{x}_o), \boldsymbol{r}_e, \boldsymbol{k}), \tag{2}$$

where $\boldsymbol{x}_{\text{pred}}$ is a set of atomic positions at local minima in the harmonic PES, and $\boldsymbol{x}_{\text{con}}$ is a set of positions of constrained atoms that are not allowed to move. This method's limitation lies in its inability to predict correct structures in systems without constrained atoms, due to potentially infinite solutions that satisfy the predicted interatomic distances. In contrast, DeepRelax can be applied to systems where all atoms are allowed to move, as it predicts both interatomic distances and displacements simultaneously. This capability, combined with DeepRelax's equivariant representation, explicit handling of periodicity, uncertainty quantification, and ability to predict lattice parameters, significantly extends its applicability and accuracy compared to DOGSS.

Cryslator conceptualizes crystal relaxation as a structure translation task and developed a generative domain translation framework that maps unrelaxed structural domains to the relaxed domains. This model requires training several networks independently, increasing its complexity. Furthermore, the Cryslator's developers argued that Cyslator exhibits low prediction performance on the Materials Project dataset as reported in its original paper Kim et al. [2023]. In contrast, DeepRelax maintains its effectiveness on the complex MP dataset (see Table 2 in the main text), demonstrating robust performance

across various test scenarios.

References

Pengru Huang, Ruslan Lukin, Maxim Faleev, Nikita Kazeev, Abdalaziz Rashid Al-Maeeni, Daria V Andreeva, Andrey Ustyuzhanin, Alexander Tormasov, AH Castro Neto, and Kostya S Novoselov. Unveiling the complex structure-property correlation of defects in 2d materials based on high throughput datasets. *npj 2D Materials and Applications*, 7(1):6, 2023.

Sungwon Kim, Juhwan Noh, Taewon Jin, Jaewan Lee, and Yousung Jung. A structure translation model for crystal compounds. *npj Computational Materials*, 9(1):142, 2023.

Junwoong Yoon and Zachary W Ulissi. Differentiable optimization for the prediction of ground state structures (dogss). *Physical Review Letters*, 125(17):173001, 2020.

REVIEWERS' COMMENTS

Reviewer #1 (Remarks to the Author):

The authors have modified the manuscript according to the previous comments, making the manuscript now a more intact and satisfying one. The results in the manuscript are very interesting and I would like to recommend the publication of this work on Nature Communications.

Reviewer #2 (Remarks to the Author):

This paper can now be accepted for publication in NC.