Chinese Physics Letters **42**, 066101 (2025) **Express Letter**

Predicting Macroscopic Properties of Amorphous Monolayer Carbon via Pair Correlation Function

Mouyang Cheng[1†], Chenyan Wang[1†], Chenxin Qin[1†], Yuxiang Zhang[1†], Qingyuan Zhang[1], Han Li[1], and Ji Chen[1,2,3*]

[1]*School of Physics, Peking University, Beijing 100871, China*
[2]*Interdisciplinary Institute of Light-Element Quantum Materials and Research Center for Light-Element Advanced Materials, Peking University, Beijing 100871, China*
[3]*Frontiers Science Center for Nano-Optoelectronics, Peking University, Beijing 100871, China*

(Received 27 March 2025; accepted manuscript online 19 May 2025)

Establishing the structure-property relationship in amorphous materials has been a long-term grand challenge due to the lack of a unified description of the degree of disorder. In this work, we develop SPRamNet, a neural network based machine-learning pipeline that effectively predicts structure-property relationship of amorphous material via global descriptors. Applying SPRamNet on the recently discovered amorphous monolayer carbon, we successfully predict the thermal and electronic properties. More importantly, we reveal that a short range of pair correlation function can readily encode sufficiently rich information of the structure of amorphous material. Utilizing powerful machine learning architectures, the encoded information can be decoded to reconstruct macroscopic properties involving many-body and long-range interactions. Establishing this hidden relationship offers a unified description of the degree of disorder and eliminates the heavy burden of measuring atomic structure, opening a new avenue in studying amorphous materials.

DOI: 10.1088/0256-307X/42/6/066101 CSTR: 32039.14.0256-307X.42.6.066101

Unraveling the connection between material structure and physical property is the foundation for discovering new principles in condensed matter physics. While recent artificial intelligence foundation models of materials have been rapidly developed,[1–4] in non-crystalline materials the structure-property relationship (SPR) can be extremely subtle due to their complex and chaotic nature.[5] Two-dimensional (2D) amorphous materials appear as the most promising platform to tackle this challenge because of the possibility to visualize the structure at the atomic level in experiments via advanced transmission electron microscopy (TEM).[6–11] Among 2D amorphous materials, amorphous monolayer carbon (AMC) stands out, providing critical opportunities to establish the complex SPR.[8,10,12,13] On one hand, the degree of disorder of AMC can be controlled by growing conditions. On the other hand, macroscopic properties such as electric conductivity and thermal transport properties can be measured and tuned over a wide range.

Early theoretical attempts focused on direct computational modeling of the atomic structure and the calculation of physical properties.[12,14–18] These studies offer important insights into the understanding of AMC, indicating that SPR exists in an unconventional manner. For instance, Tian *et al.* propose that the electric conductivity of AMC is linked simultaneously to two order parameters, namely the medium range order and the density of conducting sites.[10] However, to establish SPR of AMC, there are still a few major limitations to overcome. The first limitation arises from the inadequate sampling of atomic configurations of AMC. Theoretical works mostly assumed continuous random networks for AMC structure,[19] but recent experimental works revealed that large-scale voids and other complex defects are important structural features.[10] Secondly, accurate computations of properties such as electric conductivity and thermal conductivity are rather time consuming and could not be applied to large-scale realistic amorphous materials. Last but not least, it is unclear how to find an optimal descriptor to encode the structural information. Conventional physics studies would prefer simple scalar order parameters such as the variance of ring and bond,[14] the Steinhardt order parameters[20] and the medium range order parameter,[10] which could lead to huge information loss. Another idea is to adopt the brute-force approach widely used in the machine learning community, i.e. employing the full atomic structure as the input and utilizing the graph neural network to encode information directly.[21–23] However, it is questionable whether such brute-force approaches are optimal for SPR prediction and whether they could provide useful physical insights into help the understanding of AMC. In addition, measuring the atomic structure is in fact one of the most difficult tasks for amorphous materials including AMC, which again limits the usefulness

†These authors contributed equally to this work.
*Corresponding author. Email: ji.chen@pku.edu.cn
© 2025 Chinese Physical Society and IOP Publishing Ltd. All rights, including for text and data mining, AI training, and similar technologies, are reserved.

of existing machine learning architectures. In summary, establishing SPR of amorphous materials remains a grand challenge.

To break through all these limitations, in this work we introduce SPRamNet, standing for Structure Property Relationship for amorphous materials with networks, a machine-learning framework that links macroscopic electronic and thermal properties of amorphous systems (Fig. 1). Using AMC with varying degrees of disorder, we demonstrate the predictive power of SPRamNet. The overall goal is to train a model capable of predicting AMC spectral and transport properties based on physical descriptors as simple as possible, such as the pair correlation function (PCF). PCF can be obtained when the atomic structure is determined, and it can also be measured experimentally without knowing the full atomic structure. We train networks separately on density of states for electrons (el-DoS) and phonons (ph-DoS), electrical conductivity and thermal conductivity. Below we focus on AMC and in Section V of the SI we discuss results on silicon carbide to demonstrate that our method can be extended to three dimensional amorphous materials.

We generate 4,000 AMC configurations, each containing between 90 and 260 carbon atoms per supercell, for electronic property training. Additionally, we generate 2,000 configurations, each containing between 700 and 925 carbon atoms per supercell for phonon property training [Fig. 1(a)]. Details on the number of atoms per unit cell are shown in Supplementary Information (SI, with additional references [24–32]) II. Starting from pristine graphene, we employ two typical techniques for creating diverse AMC structures: one is based on the Monte Carlo bond switching algorithm, [33,34] while the other is the quenching algorithm. [35] For the phonon dataset, all 2,000 structures are generated strictly in the two-dimensional plane, with half created using quenching algorithms and the other half using the Monte Carlo bond switching algorithm. For the electron dataset, 2,000 structures are generated by two-dimensional quenching, while the other 2,000 structures are produced by bond switching in three-dimensional space to allow the natural out-of-plane oscillation. Subsequently, we perform high-throughput density functional theory (DFT) calculations to obtain el-DoS and electric conductivity. ph-DoS and thermal conductivity are calculated using classical molecular dynamics (MD) with the Tersoff [36] force field [Fig. 1(d)]. More details about data preparation can be found in Section I of the SI.

The key to our workflow is to construct an appropriate descriptor to encode the structural information [Fig. 1(b)]. A suitable descriptor should both be scalable – applicable for arbitrary sizes of input supercells – and allow E(3) & permutation-invariant outputs, meaning the neural network prediction remains unchanged under atomic permutations and E(3) transformations such as rotation, translation, and inversion. As mentioned above, a brute-force approach is to build a message passing graph neural network (MPNN) [21–23] with multiple localized descriptors on each nodal atoms. Scalable and permutation-invariant output features can thus be guaranteed via global aggregation. However, the usefulness of localized descriptors heavily depends on the availability of atomic configurations. In SPRamNet we choose global descriptors such as PCF and the angular distribution function (ADF) to encode structural information of AMC. In particular, PCF describes how density varies as a function of distance from a reference particle, and is of special physical significance, as it is directly related to the system's structure factor via inverse Fourier transform. Pair correlation function can be obtained in experiments using X-ray or neutron diffraction, [37] bypassing quite involved measurements of atomic configurations, hence choosing it as a descriptor overcomes the greatest challenge in experimental studies of amorphous materials. Moreover, it is straightforward to check that PCF ensures both scalability and invariance while simplifying the computational complexity, making it an efficient and physically insightful choice for modeling

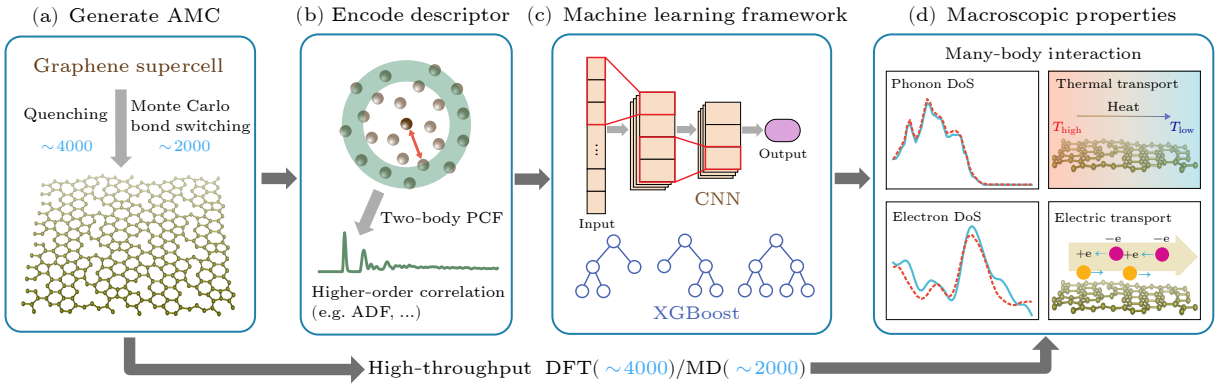

Fig. 1. Schematic of the SPRamNet workflow for predicting electronic and thermal properties of amorphous monolayer carbon (AMC) from pair correlation function (PCF). (a) AMC structures are generated by quenching or Monte Carlo bond switching algorithm from the pristine graphene supercell. The electronic and phonon properties for AMC are simulated via high-throughput density functional theory (DFT) and molecular dynamics (MD), respectively. (b) The structural information is encoded using PCF highlighted as the green curve, while other higher-order correlation inputs are optional. (c) These global descriptors serve as the input to a machine learning pipeline based on either convolutional neural network (CNN) or Extreme Gradient Boosting (XGBoost). (d) Subsequently, the machine learning models can be trained to predict macroscopic electronic and thermal properties.

AMC structures. The performance of a graph neural network architecture is discussed in the Section IV of the SI.

With PCF readily calculated from structural input, in SPRamNet we discretize PCF and target physical properties as fixed-size direct input and output of the network, respectively [Fig. 1(c)]. In principle, the network can be subsequently trained to reconstruct macroscopic properties of AMC.* For different property prediction tasks, we make minor adjustments to the network dimensions to make them compatible to the outputs, either vectorized el-DoS or ph-DoS, or simply scalar conductivity values σ or κ. One can also truncate PCF at different distance to further simplify the input information.

Before presenting the machine learning results, let's have a deeper analysis of the AMC structures in the dataset, which covers a wide range of degree of disorder. Figures 2(a)–2(c) show three AMC structures with increasing degree of disorder and their corresponding spectral properties. We set the frequency range for ph-DoS within $f \in [0, 100]$ THz, and the energy scale for el-DoS within $E \in [-2, 2]$ eV w.r.t the Fermi level E_F, since it covers the energy window where most important physical processes occur. PCF is shown within a cutoff radius of $r_{max} = 8$ Å. We note that the PCF descriptor $g(r)$ should be normalized w.r.t the radial distance r and number of atoms N, so that it becomes scalable and converges for increasing r. Also, additional care is required for a properly normalized $g(r)$ for quasi-2D structures in the electron dataset. More details on normalizing the $g(r)$ is shown in Section I.4 of the SI.

As expected, the degree of disorder has substantial impacts on spectral properties of AMC. For the structure with minor defects [Fig. 2(a)], long range crystalline order is clearly present in $g(r)$, where multiple sharp peaks persist for $r > 4$ Å. The el-DoS is near zero at the Fermi level, which also resembles the pristine graphene behavior, and a sharp peak emerges at $f = 48$ THz, corresponding to the G peak frequency and vibrational modes in graphene. As disorder gradually dominates in the AMC structure [from Fig. 2(a) to 2(c)], peaks for $r > 4$ Å become suppressed and gradually vanish, indicating the absence of long range order in strongly disordered AMC structures. Such loss of long range order induces great change in measurable physical properties. This change is evident in el-DoS, where a centered peak appears at the Fermi level and then smears out as degree of disorder increases. Additionally, while the peak positions in ph-DoS remain largely unchanged, the heights and widths of these peaks are significantly altered, indicating dominance of other vibrational modes in AMC as disorder starts to play an important role. Apart from these spectral properties, transport properties also vary in three orders of magnitude within our dataset, which will be shown below and in Section II of the SI.

Evidence above hints the possibility of a structure property relation between PCF descriptor and electronic and thermal properties, but to what extent can PCF reconstruct these macroscopic properties remains to be explored. Figure 3 presents the results on thermal properties of AMC using a convolutional neural network (CNN). We note that though the AMC structure is not inherently structured as 2D images, the correlation function can be naturally encoded as a sequential spatial correlation function, which is a perfect fit for CNN. The network architecture is based on a variant of LeNet[38] with optional residual connection.[39]

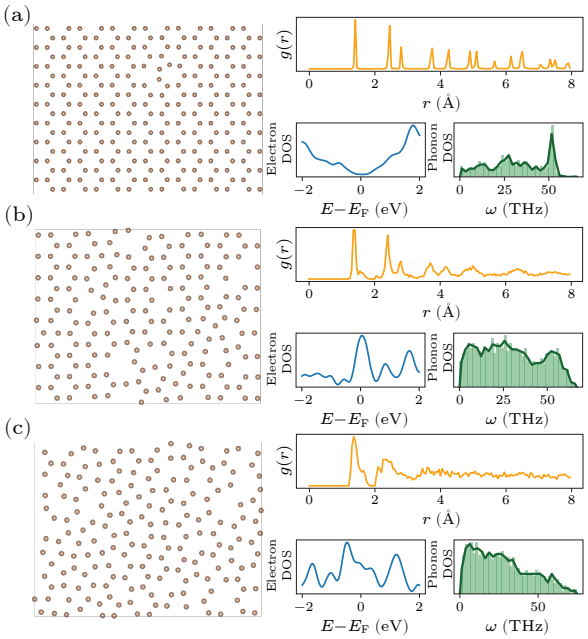

Fig. 2. Display of dataset generated for SPRamNet training. Three typical types of AMC structures with different degrees of disorder: (a) graphene structure with point defects; (b) near-pristine graphene structure with minor perturbation and defects; and (c) highly amorphous carbon structure, along with their corresponding PCF $g(r)$, density of states (el-DoS), and phonon density of states (ph-DoS).

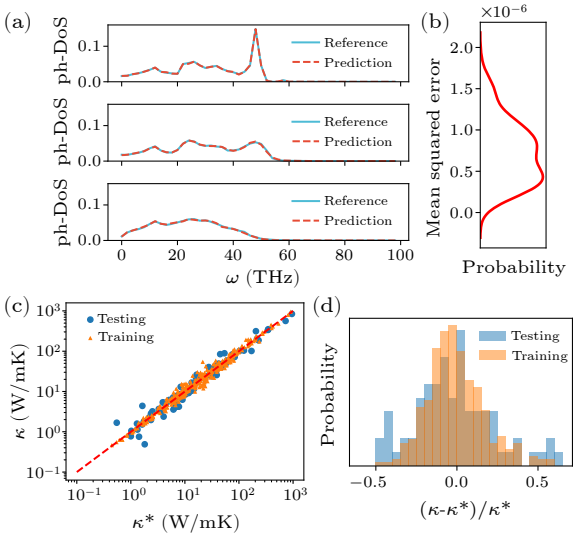

Fig. 3. SPRamNet Prediction of AMC thermal properties. (a) Comparison between the calculated and the predicted ph-DoS in the testing set. (b) Distribution of MSE loss for testing results. (c) Predicted thermal conductivity κ at $T = 300$ K plotted against the calculated reference (κ^*) on the logarithmic scales for comparison. (d) Distribution of relative error $(\kappa - \kappa^*)/\kappa^*$ for training and testing results.

*The code and generated dataset used for training in this work are available in https://github.com/vipandyc/SPRamNet.

More details on the architecture and hyperparameter settings are shown in Section III.1 of the SI. Figure 3(a) displays the comparison of ph-DoS for three typical structures with different disorder levels, and we observe a prefect agreement between MD and machine learning results. More quantitatively, the expectation of mean squared error (MSE) loss is as low as $\sim 10^{-6}$ in Fig. 3(b), which is negligible compared to the ph-DoS signal. To further test whether PCF contains sufficient information as the structural input, we incorporate the three-body correlation function, the angular distribution function (ADF), as additional input (Fig. S8). We find that such addition gives ∼3% improvement in terms of MSE loss, indicating a marginal contribution of higher order correlation functions.

Similarly, SPRamNet shows good performance on the thermal conductivity. Reference value of thermal conductivity κ is calculated using MD at 300 K employing the Green-Kubo (GK) formalism. In Fig. 3(c), the predicted thermal conductivity (κ) is plotted against the reference value (κ^*) for both the training and the testing data. In Fig. 3(d), the error distributions are plotted. Overall, these findings above suggest that SPRamNet receiving PCF as descriptor alone can almost fully predict the thermal properties of AMC. One can show the existence of SPR for model systems with pairwise interaction,[40] but the success of SPRamNet reveals that more SPRs are hidden in realistic materials containing interactions beyond pairwise types.

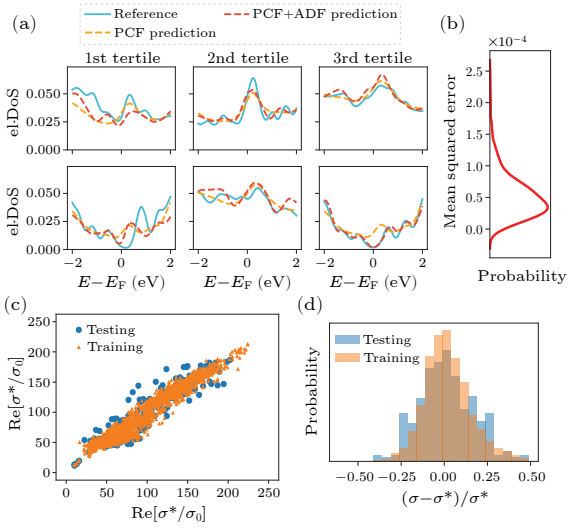

Fig. 4. SPRamNet prediction of electronic properties. (a) Examples comparing the calculated el-DoS with the predicted el-DoS in the testing set, for both PCF and PCF+ADF prediction. (b) Distribution of MSE loss for PCF+ADF testing results. The three columns represent examples with different tertiles of MSE error, from the 1st column (left) with highest error, to the 3rd (right) with lowest error. (c) Predictions of diagonal real electric conductivity Re[σ] at zero temperature. Predicted σ are plotted against DFT calculated reference σ^* for comparison. $\sigma_0 = 1$ kS/m. (d) Distribution of relative error $(\sigma - \sigma^*)/\sigma^*$.

Moving to electronic properties, the model training immediately becomes more difficult with the same network used for thermal properties, indicating that the SPR for electronic structure involves more degrees of freedom in AMC structure as expected. For el-DoS, we find it is necessary to improve the depth of the network to improve the expressiveness. As shown in Fig. 4(a), a deeper network can readily improve the prediction (red and orange curves) of el-DoS to a satisfactory level compared with the DFT results (blue curves). The expected MAE loss is lower than 5×10^{-5} in Fig. 4(b), and SPRamNet can still well reconstruct the main features of DFT calculated el-DoS. Going through all the data, one can still notice visible deviations for some AMC structures. With the inclusion of ADF in the input, the predicted el-DoS can be further improved. From the MSE loss, the improvement is estimated ∼10%, which further supports that PCF contains sufficient information to qualitatively reconstruct electronic properties (Section III.2 of the SI).

For the electric conductivity σ, we take the frequency limit $\omega \to 0$ for the real part of optical conductivity $\sigma(\omega)$ calculated using DFT. In this more challenging case, we find the accuracy of convolutional network can only be mildly improved via enlarging the network depth. To this end, an alternative machine learning architecture, namely the Extreme Gradient Boosting (XGBoost)[41] algorithm, is implemented and reaches the highest accuracy among all the settings tested (Section III.3 of the SI). XGBoost is more suitable and easier to tune for scalar predictions in complex tasks, hence it is integrated as a part of SPRamNet. The predicted electrical conductivity is shown in Fig. 4(c), where both training and testing results are displayed for comparison. The relative testing error shown in Fig. 4(d) is mostly less than 25%, while the root mean square error (RMSE) of the testing result is $13.2\sigma_0$ for the diagonal conductivity. The results suggest that with a suitable machine learning architecture, PCF can still be used as an effective descriptor for electrical conductivity.

Although the performance of SPRamNet is slightly worse for electronic properties compared to thermal properties, the achievement is a bigger surprise. We also note that predicting electronic properties such as the electronic density of states (el-DoS) remains a significant challenge in the literature. Compared to previous works focusing on el-DoS prediction for crystals using graph network embedding,[42,43] our SPRamNet demonstrates superior performance using a much more simplified global descriptor PCF, particularly in accurately capturing the key peaks in the el-DoS. These peaks are critical for extracting material behavior, such as thermoelectric and optical properties.

PCF carries structural information on amorphous material by encoding the atomic pairs at various distances. The descriptor itself is a two-body metric, but many-body information is indirectly accounted. This hints at a potential reduction in the effective dimensionality of the problem, where two-body interactions encapsulate much of the behavior that would traditionally be attributed to higher-

order interactions. Moreover, the success of SPRamNet suggests that simple descriptors may contain rich information, which was not fully explored in traditional physics studies. To test the limit of a minimal descriptor, we analyze the effective range of $g(r)$. The results reveal that all significant contributions are confined to features in $g(r)$ with radial distances $r < 1.8$ Å (Section III.4 of the SI). This indicates that a very small range of $g(r)$ could already reflect the most relevant structural information. In Section IV of the SI, for the interest of relevant readers, we show tests with an E(3)-equivariant network adopted from the e3nn package[44]. As expected, SPRamNet demonstrates superior efficiency and accuracy compared to the prevailing GNN approaches (Section IV of the SI).

To summarize, we propose SPRamNet, a machine learning workflow to reconstruct both electronic and phonon properties of AMC by using structure correlation functions as global descriptors. The SPRamNet framework establishes the complex SPR of AMC, and demonstrates that AMC's macroscopic properties can be captured through the simplest pairwise correlations. Leveraging the power of machine learning, our results indicate that even as a global descriptor representing the average effects of local environment, PCF is still generally linked to electronic, phonon, and potentially other measurable properties even for many-body interactions. SPRamNet could be generalized to suit other glassy systems beyond two-dimension and single-element carbon, such as amorphous alloy and semiconducting compounds (see Section V of the SI for an example). In existing machine learning studies of disordered systems, local structure features are often considered as the main input and their aggregations are used to characterize the degree of disorder.[45,46] In contrast, this study promotes the use of simple yet physically meaningful descriptors for machine learning studies of complex amorphous materials, which offers several other advantages, including facile training and high interpretability. In the future, it is desirable to apply the idea of SPRamNet to the rapidly developing foundation materials models, which is likely to further improve their performance on non-crystalline materials. Finally, because SPRamNet only takes PCF as the input, it can take the advantage of many mature experimental techniques for amorphous materials. At the current stage, experimentally measuring atomic coordinates of amorphous materials is still one of the greatest challenges in the field. Local structure features would require the information of atomic coordinates, but for a globally averaged PCF one can circumvent the issue by employing diffraction techniques such as X-ray and neutron scattering, which are much more available in experimental labs.

Acknowledgements. This work is supported by the National Key R&D Program of China under Grant No. 2021YFA1400500, the Strategic Priority Research Program of the Chinese Academy of Sciences under Grant No. XDB33000000, the National Natural Science Foundation of China under Grant No. 12334003, and the Beijing Municipal Natural Science Foundation under Grant Nos. JQ22001 and QY23014. The authors are grateful for computational resources provided by the High Performance Computing Platform of Peking University.

References

[1] Chen C and Ong S P 2022 *Nat. Comput. Sci.* **2** 718
[2] Deng B, Zhong P, Jun K, Riebesell J, Han K, Bartel C J, and Ceder G 2023 *Nat. Mach. Intell.* **5** 1031
[3] Batatia I, Benner P, Chiang Y, Elena A M, Kovács D P, Riebesell J, Advincula X R, Asta M, Avaylon M, Baldwin W J *et al.* 2023 arXiv:2401.00096 [physics.chem-ph]
[4] Yang H, Hu C, Zhou Y, Liu X, Shi Y, Li J, Li G, Chen Z, Chen S, Zeni C *et al.* 2024 arXiv:2405.04967 [cond-mat.mtrl-sci]
[5] Anderson P W 1995 *Science* **267** 1615
[6] Huang P Y, Kurasch S, Alden J S, Shekhawat A, Alemi A A, McEuen P L, Sethna J P, Kaiser U, and Muller D A 2013 *Science* **342** 224
[7] Hong S, Lee C S, Lee M H *et al.* 2020 *Nature* **582** 511
[8] Toh C T, Zhang H J, Lin J H *et al.* 2020 *Nature* **577** 199
[9] Joo W J, Lee J H, Jang Y *et al.* 2017 *Sci. Adv.* **3** e1601821
[10] Tian H, Ma Y, Li Z, Cheng M, Ning S, Han E, Xu M, Zhang P F, Zhao K, Li R *et al.* 2023 *Nature* **615** 56
[11] Bai X *et al.* 2024 *Nature* (in press)
[12] Felix L C, Tromer R M, Autreto P A S *et al.* 2020 *J. Phys. Chem. C* **124** 14855
[13] Kim I S, Shim C E, Kim S W *et al.* 2023 *Adv. Mater.* **35** 2204912
[14] Van Tuan D, Kumar A, Roche S *et al.* 2012 *Phys. Rev. B* **86** 121408
[15] Antidormi A, Colombo L, and Roche S 2022 *Nano Mater. Sci.* **4** 10
[16] Cheng M, Chen H, and Chen J 2024 *Chin. Phys. B* **33** 107201
[17] Antidormi A, Colombo L, and Roche S 2020 *2D Mater.* **8** 015028
[18] Fan Z Y, Garcia J H, Cummings A W, Barrios-Vargas J E, Panhans M, Harju A, Ortmann F, and Roche S 2021 *Phys. Rep.* **903** 1
[19] Wright A C 2014 *Int. J. Appl. Glass Sci.* **5** 31
[20] Kotakoski J, Krasheninnikov A V, Kaiser U, and Meyer J C 2011 *Phys. Rev. Lett.* **106** 105505
[21] Gilmer J, Schoenholz S S, Riley P F *et al.* 2017 *Proc. ICML* **70** 1263
[22] Reiser P, Neubert M, Eberhard A *et al.* 2022 *Commun. Mater.* **3** 93
[23] Xie T and Grossman J C 2018 *Phys. Rev. Lett.* **120** 145301
[24] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, and Dubourg V 2011 *J. Mach. Learn. Res.* **12** 2825
[25] Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L *et al.* 2019 *Adv. Neural Inf. Process. Syst.* 8026
[26] Vashishta P, Kalia R K, Nakano A, and Rino J P 2007 *J. Appl. Phys.* **101** 103515
[27] Anderson D A and Spear W E 1977 *Philos. Mag.* **35** 1
[28] Bazant M Z and Kaxiras E 1996 *Phys. Rev. Lett.* **77** 4370
[29] Bazant M Z, Kaxiras E, and Justo J F 1997 *Phys. Rev. B* **56** 8542
[30] Luo J P, Zhou C Y, Cheng Y J, and Liu L J 2020 *J. Cryst. Growth* **546** 125785
[31] Lucas G, Bertolus M, and Pizzagalli L 2009 *J. Phys.: Condens. Matter* **22** 035802
[32] Batatia I, Kovács D P, Simm G N C, Ortner C, and Csányi G 2022 *Adv. Neural Inf. Process. Syst.* **35** 11423
[33] von Alfthan S, Kuronen A, and Kaski K 2003 *Phys. Rev. B* **68** 073203

[34] Wooten F, Winer K, and Weaire D 1985 *Phys. Rev. Lett.* **54** 1392

[35] Kumar A, Wilson M, and Thorpe M F 2012 *J. Phys.: Condens. Matter* **24** 485003

[36] Tersoff J 1988 *Phys. Rev. Lett.* **61** 2879

[37] Dinnebier R E and Billinge S J 2008 *Powder Diffraction: Theory and Practice* (RSC)

[38] LeCun Y, Bottou L, Bengio Y, and Haffner P 1998 *Proc. IEEE* **86** 2278

[39] He K, Zhang X, Ren S, and Sun J 2016 *Proc. CVPR* **2016** 770

[40] Chandler D 1987 *Introduction to Modern Statistical Mechanics* (Oxford Univ. Press)

[41] Chen T and Guestrin C 2016 *Proc. KDD* 785

[42] Louis S Y, Zhao Y, Nasiri A, Wang X, Song Y, Liu F, and Hu J 2020 *Phys. Chem. Chem. Phys.* **22** 18141

[43] Kong S, Ricci F, Guevarra D, Neaton J B, Gomes C P, and Gregoire J M 2022 *Nat. Commun.* **13** 949

[44] Geiger M and Smidt T 2022 arXiv:2207.09453 [cs.LG]

[45] Chapman J, Hsu T, Chen X, Heo T W, and Wood B C 2023 *Nat. Commun.* **14** 4030

[46] Aykol M, Merchant A, Batzner S, Wei J N, and Cubuk E D 2023 arXiv:2310.01117 [cond-mat.mtrl-sci]