Exploring Phononic Properties of Two-Dimensional Materials using Machine Learning Interatomic Potentials

Bohayra Mortazavi[a,b], Ivan S Novikov[c,d], Evgeny V Podryabinkin[c], Stephan Roche*[e,f], Timon Rabczuk[g], Alexander V Shapeev**[c] and Xiaoying Zhuang[g]

[a]*Institute of Continuum Mechanics, Leibniz Universität Hannover, Appelstraße 11, 30157 Hannover, Germany.*
[b]*Cluster of Excellence PhoenixD (Photonics, Optics, and Engineering–Innovation Across Disciplines), Gottfried Wilhelm Leibniz Universität Hannover, Hannover, Germany.*
[c]*Skolkovo Institute of Science and Technology, Skolkovo Innovation Center, Nobel St. 3, Moscow 143026, Russia.*
[d]*Institute of Materials Science, University of Stuttgart, Pfaffenwaldring 55, 70569 Stuttgart, Germany*
[e]*Catalan Institute of Nanoscience and Nanotechnology (ICN2), CSIC and BIST, Campus UAB, Bellaterra, 08193 Barcelona, Spain.*
[f] *ICREA Institució Catalana de Recerca i Estudis Avancats, 08010 Barcelona, Spain*
[g]*College of Civil Engineering, Department of Geotechnical Engineering, Tongji University, Shanghai, China.*

Abstract

Usually phononic properties are mostly studied using density functional perturbation theory (DFPT) simulations. Although DFPT simulations offer accurate estimations of phononic properties, but for low-symmetrical and nanoporous structures, the computational cost becomes quickly very demanding. Besides these, due to computational setups nonphysical negative branches may appear in phonon dispersions, which impede the assessment of phononic properties and dynamical stability. Here, we compute phonon dispersion relations and examine the dynamical stability of a large ensemble of novel materials and compositions. We propose a fast and convenient alternative to DFPT simulations to evaluate the phononic properties of low-symmetrical and porous structures, via machine-learning interatomic potentials passively trained over computationally inexpensive ab-initio molecular dynamics trajectories. Results for diverse two-dimensional (2D) nanomaterials confirm that the proposed approach can reproduce fundamental phononic properties in close agreements with those by DFPT approach. The proposed approach offers a stable, efficient and convenient solution for the examination of dynamical stability and exploring the phononic properties of low-symmetry and porous 2D materials.

Corresponding authors: *stephan.roche@icn2.cat; **a.shapeev@skoltech.ru

1. Introduction

Phonon dispersion relations (PDRs) are key components to study the lattice dynamics and atomic vibrations in a crystal[1,2]. They also provide useful information regarding the transport properties such as the thermal conductivity. In particular, PDRs are extensively employed to examine the dynamical stability of various compositions and structures. To acquire PDRs, the most popular theoretical approach is to conduct density functional perturbation theory (DFPT) simulations. In order to assess the thermal properties, DFPT calculations are commonly executed using supercell structures. DFPT is a computationally efficient approach for the majority of highly symmetrical lattices. In recent years, two-dimensional (2D) materials[3,4] are gaining remarkable attentions because of their unique properties, suitable to address critical challenges in various advanced technologies like nanoelectronics. Since the isolation of graphene[1,2], 2D materials family has been continuously and quickly extending. This large family of new materials includes highly symmetric lattices, such as the graphene and 2H transition metal dichalcogenides[5,6], as well as many structures with low-symmetrical or nanoporous lattices like graphdiyne[7] and 1T' transition metal dichalcogenides[8]. To date, high-symmetrical and isotropic lattices have received higher concern. However recently, anisotropic and nanoporous 2D lattices are attracting considerable attention, because they offer novel possibilities to design angle-dependent devices and more efficient energy storage/conversion systems [9].

In order to theoretically predict novel 2D systems, it is thus essential to carefully examine the dynamical stability and explore the vibrational properties on the basis of PDRs. For low-symmetrical and nanoporous 2D structures, currently employed DFPT simulations manifest limited flexibilities due to severe computational issues. In these cases, to bypass the computational limitations, one usually treats smaller supercells and lower k-point girds and plane-wave cutoff energy within the DFPT simulations. But such simplifications may actually result in poorer accuracy of the acquired PDRs. Moreover, in some cases negative branches may appear in the PDRs and thus impeding the examination of dynamical stability or thermal properties. Astonishing recent advances in the field of machine-learning, have offered novel possibilities to address various challenges in materials science[10–13]. In this regard, machine-learning interatomic potentials (MLIP) [14] have shown outstanding efficiency in many applications of computational material science, such as predicting novel materials [15,16] lattice dynamics[17] and estimating the thermal conductivity[18], etc. The main advantage of MLIPs is that they enable the efficient use of classical molecular dynamics simulations to evaluate the forces and energies with the density functional theory (DFT) level of accuracy. The employment of MLIPs is particularly promising to study the large systems, for those DFT-based methods become infeasible due to renowned computational limitations.

In this work we propose a computationally efficient and accurate methodology to acquire PDRs and explore other critical phononic properties on the basis of passively trained moment tensor potentials (MTPs)[19]. The proposed approach only requires inexpensive and short ab-initio molecular dynamics trajectories, without need to active learning or additional DFT calculations. We employed the proposed approach to conveniently compute the phononic

properties of a wide-variety of 2D materials, for which close agreements with DFPT results were observed. We discuss that the proposed approach on the basis of passively-trained MTPs can be conveniently employed to explore the phononic properties of low-symmetrical and nanoporous 2D materials, with enhanced stability and computational efficiency.

2. Computational methods

First-principles DFT calculations in this work were carried out using the *Vienna Ab-initio Simulation Package* (VASP)[20–22]. Generalized gradient approximation (GGA) within Perdew–Burke–Ernzerhof (PBE)[23] method was employed in the calculations. We assumed plane-wave cutoff energies of 600 eV and 400 eV for carbon-based and rest of the studied systems, respectively. For geometry optimization, the convergence criterion for the energy and forces were set to 10^{-5} eV and 0.001 eV/Å, respectively. DFPT simulations were performed over supercell samples using a 3×3×1 Monkhorst-Pack[24] k-point grid. The plane-wave cutoff energy of DFPT simulations was set as the default value by VASP. PHONOPY code[25] was utilized to create the optimal sets of atomic position for DFPT calculations and also to acquire phonon dispersions and group velocities with DFPT results as inputs. Ab-initio molecular dynamics (AIMD) simulations were performed with a time step of 1 fs suing a 3×3×1 k-point gird.

A class of machine-learning interatomic potentials—moment tensor potentials[19] were used to describe interatomic interactions. Similar to classical potentials, MTPs include parameters that are found by solving of a minimization problem. In this work AIMD simulations were used to create the training sets. MTP was first proposed for single-component systems[19] and has been recently generalized for the multiple component systems[15,26]. This potential is local, i.e., the total energy *E* of the system containing *N* atoms is partitioned into contributions *V* of neighborhoods u_i of each i-th atom: $E \equiv E^{MTP} = \sum_{i=1}^{N} V(u_i)$. We refer to the j-th atom as neighbor of the i-th (central) atom if the distance between them is less than a predefined cut-off distance R_{cut}. The Neighborhood is then expressed as a tuple, $u_i = \left(\{r_{i1}, z_i, z_1\} \dots, \{r_{ij}, z_i, z_j\} \dots, \left\{r_{iN_{neigh}}, z_i, z_{N_{neigh}}\right\}\right)$, where r_{ij} is the relative atomic position (interatomic vector), z_i and z_j are the types of the central and neighboring atoms, respectively and N_{neigh} is the number of atoms in neighborhood. Each contribution to the total energy has the following form: $V(u_i) = \sum_\alpha \xi_\alpha B_\alpha(u_i)$, where ξ_α are the free parameters of the potential to be optimized, B_α are the potential basis functions. We construct the basis functions as all possible contractions of the moment tensor descriptor:

$$M_{\mu,v}(r_i) = \sum_{j=1}^{N_{ng}} f_\mu\left(|r_{ij}|, z_i, z_j\right) r_{ij}^{\otimes v} \qquad (1)$$

yielding a scalar (see Ref.[19] for details). The first factor $f_\mu(|r_{ij}|, z_i, z_j)$ in the aforementioned summation is the radial part depending only on the distance between atoms i and j and their types. We expand the radial part through a set of radial basis functions $\varphi_\beta(|r_{ij}|)$ multiplied by a factor $\left(R_{cut-}|r_{ij}|\right)^2$ for smoothing near the distances close to cut-off radius.

$$f_\mu(|r_{ij}|, z_i, z_j) = c^{(\beta)}_{\mu,z_i,z_j}\varphi_\beta(|r_{ij}|)(R_{cut}-|r_{ij}|)^2, \quad (2)$$

Where $c^{(\beta)}_{\mu,z_i,z_j}\varphi_\beta$ are the are the radial coefficients. We denote "⊗" by outer product and refer to the second factor in Eq. 2 as the angular part which describes polyatomic interactions. In order to optimize ξ_α and $c^{(\beta)}_{\mu,z_i,z_j}\varphi_\beta$ parameters of a MTP, one needs to solve the following minimization problem (training of MTP):

$$\sum_{k=1}^{K}\left[w_e\left(E_k^{AIMD}-E_k^{MTP}\right)^2 + w_f\sum_{i}^{N}\left|f_{k,i}^{AIMD}-E_{k,i}^{MTP}\right|^2\right] \rightarrow min, \quad (3)$$

where E_k^{AIMD} and $f_{k,i}^{AIMD}$ are the energy and atomic forces in the training set, respectively, K is the number of the configurations in the training set, w_e and w_f are non-negative weights that express the importance of energies and forces in the optimization problem.

Here we used PHONOPY code[25] to evaluate the phononic properties, in which MTP replaces VASP in the force calculation step. To facilitate future studies, in the data availability section, we present a direct guide to create the training set (VASP AIMD inputs), as well as the MTP training procedure and the integration code to calculate forces using the MTP over input structures by PHONOPY code and PHONOPY inputs. Particularly all considered examples in this work are included. Such information will facilitate the learning and practical use of MTP for accurate evaluation of the phononic properties in structural systems of versatile complexity.

3. Results and discussions

The main objective of this study is to develop MTPs to replace DFT simulations in the force-constant calculations for the evaluation of phononic properties. Although our work is focused on the 2D materials, the proposed approach can be equally employed for 3D or 1D lattices. First, we study on the monoelemental 2D structures. Since the carbon shows the unique ability to form diverse stable 2D atomic lattices, we mostly consider carbon-based structures. In Fig. 1, the studied monoelemental 2D lattices are illustrated. Apart from the graphene (Gr), we consider penta-graphene[27], various haeckelite[28] (Haeck) lattices, phagraphene [29] (Pha-Gr) and different graphyne [30] (GY) structures. We note that haeckelite and phagraphene lattices include non-hexagonal carbon rings and can be thus considered as defective graphene lattices, but with high degree of periodicities and close densities to that of the pristine graphene. On the other side, graphyne lattices are highly porous full carbon materials, which recently gained remarkable attentions because of experimental advances and their bright application prospect for the energy storage/conversion systems. Besides full-carbon structures, we also consider the single-layer black phosphorene (P) which shows a rather complicated buckled lattice.

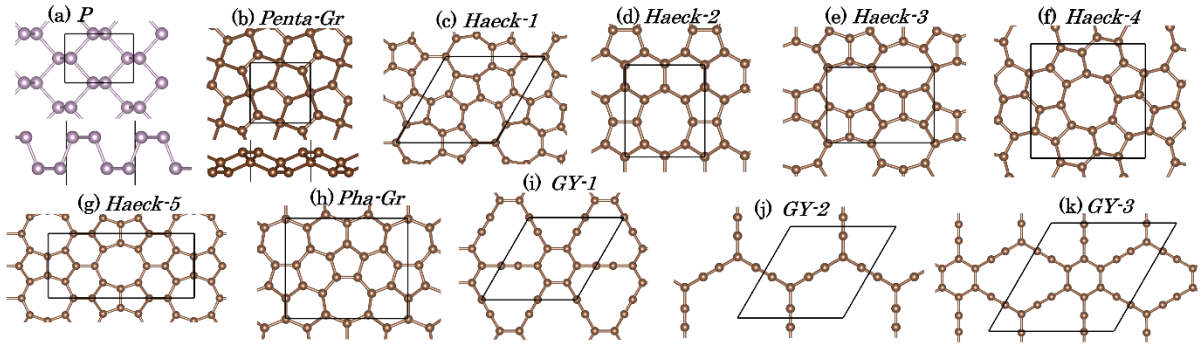

Fig. 1, Top views of atomic structures of considered monoelemental systems. Side views are shown only for non-planar lattices. Black lines illustrate the primitive unitcell.

Since the phononic properties are usually evaluated by applying the small displacements in supercell lattices, we created the training data sets by conducting the AIMD simulations at the very low temperature of 50 K, with an overall simulation time of less than 1 ps (1000 simulation time steps). Since the MTPs are trained with a cutoff distance of 4 or 5 Å, the AIMD simulations were conducted over supercells with periodic box sizes over 10 Å. Larger supercells can be useful to describe the local large deflections better. Nonetheless, supercells with the minimum sizes closer to 10 Å are more computationally efficient because the costs of AIMD simulations increase exponentially with the number of atoms. To facilitate future studies, all created training sets in this work are provided in the data availability section. Using the AIMD results, we trained MTPs with 901 parameters for the monoelemental systems. The cutoff distance for graphene, other carbon allotropes and phosphorene were assumed to be 3, 4 and 5 Å, respectively. In Fig. 2 we compare the PDRs predicted by MTPs with those by the DFT. Remarkably close agreement between the MTP and DFT results are obtained. This is a highly promising outcome since the developed MTPs are first attempts without any long sequence of optimization. Noticeably for the penta-graphene, haeckelite lattices and phosphorene, acoustic and optical branches are reproduced with a high level of accuracy. The same observation is also valid for graphyne lattices. Notably, in some cases (Haeck-5, P, GY-2 and GY-3) slight negative branches appear in the PDRs calculated with DFT, which are not observable in the MTP-based results. These negative branches in the DFT results originate from computational artifacts, and are not representative of dynamical instabilities[31] and might be removed via more precise setups for the calculations, such as increasing the supercell size, increasing plane-wave cutoff energy and improving the resolution of k-point girds. These modifications can be affordably examined for highly symmetrical lattices like graphene, but for less symmetrical and nanoporous lattices may substantially increase the computational costs and lead to complexity of the problem.

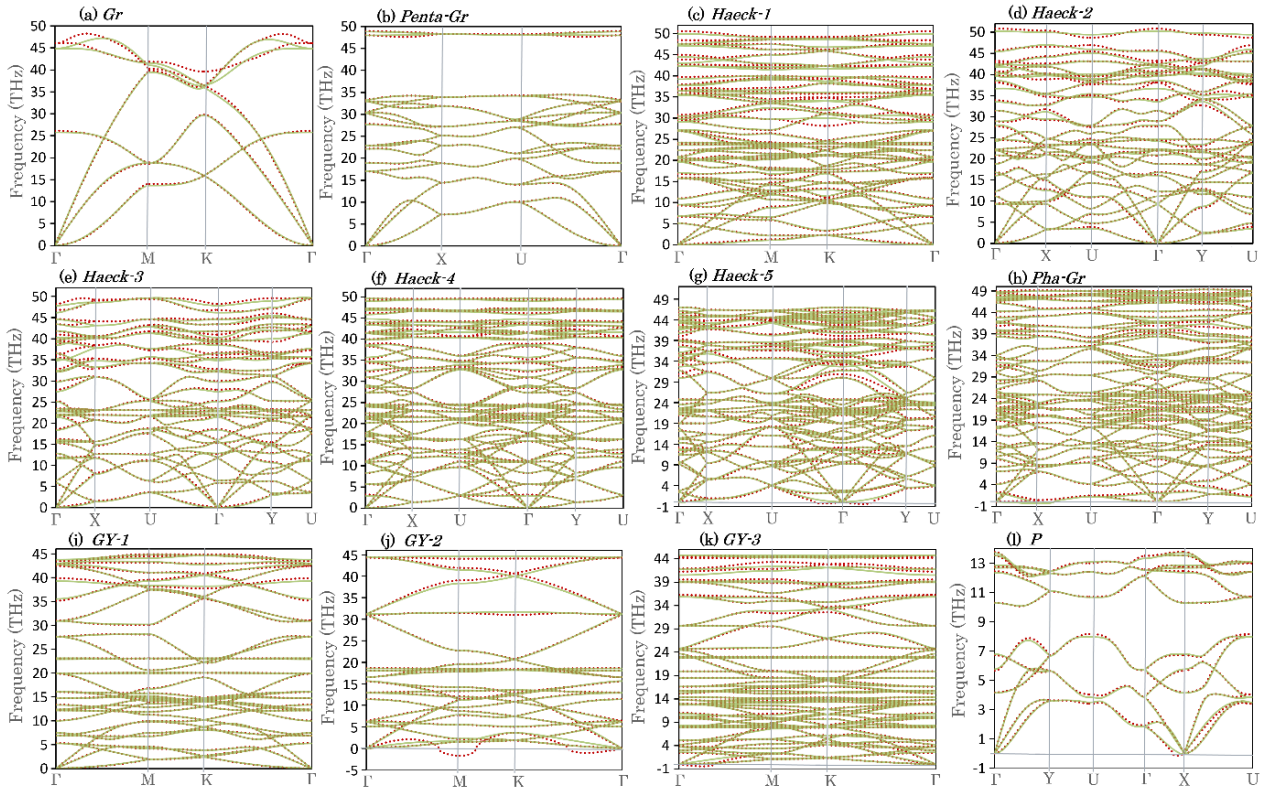

Fig. 2, Phonon dispersion relations of monoelemental 2D lattices acquired by the DFPT (red-dotted lines) method and first-attempt MTPs (continuous green lines). The atomic lattices are show in Fig. 1.

Despite of close agreements between the predicted PDRs for the most of monoelemental structures, two exceptions exist in which MTP based results show some inaccuracies. Among all studied samples, it is clear that high frequency optical modes are well reproduced by the MTP. Unexpectedly for the highest symmetrical lattice of graphene, optical modes with highest frequencies are not well reproduced by the MTP (Fig. 2a). On the other side for the phagraphene (Fig. 2h), a slight soft mode appears in the MTP result which is absent in the DFPT result. These disagreements between the DFPT and MTP results reveal that some degree of extrapolation may have occurred and accordingly suggest that for these samples the created passive training sets may have not been as accurate as those for other samples. We remind that when conducting the AIMD simulations at a low temperature of 50 K for a very short period, different configurations remain highly correlated and such that some critical configurations may never get explored in the training set. To address this issue, a simple effective solution is to expand the training set by including additional uncorrelated AIMD trajectories at higher temperatures. To this aim, an additional 1 ps AIMD trajectories were included by conducting the simulations at 200, 300, 500, 700 and 900 K. Then the new training sets were created by adding equal-length trajectories from different temperatures. In Fig. 3, we include the predicted PDRs for single-layer graphene and phagraphene by the improved training sets and compare the results with the original results. It is clear that by incorporating higher temperature trajectories, the negative branch in the phagraphene's PDR vanishes, while the close agreement with DFPT results at higher frequencies is kept intact. Similarly, for the case of graphene, the accuracy of high-frequency optical modes is

considerably improved without affecting the lower frequency optical and acoustic modes. The acquired results reveal that the incorporation of high-temperature trajectories can further enhance the accuracy and stability of MTPs.

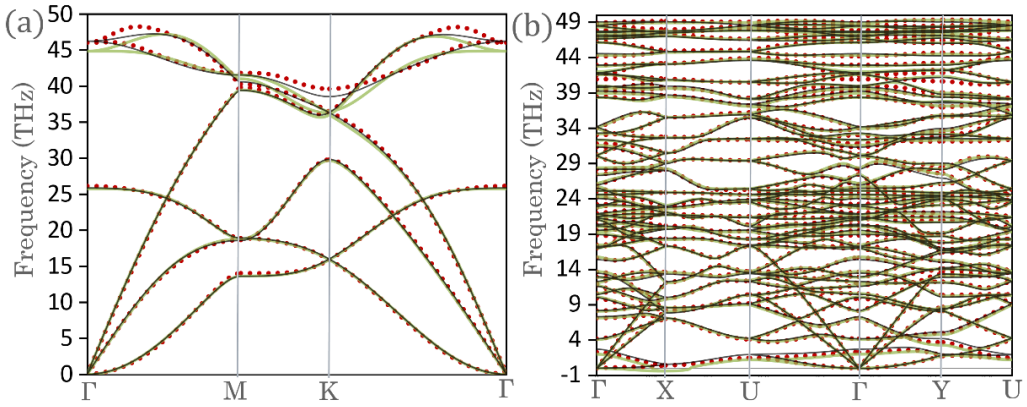

Fig. 3, Phonon dispersion relations of (a) graphene and (b) phagraphene acquired by the DFPT (red-dotted lines) and MTPs passively trained over AIMD trajectories at 50 K (continuous green lines) and 50 to 900 K (continuous green lines).

We next consider the binary 2D systems, which are shown in Fig. 4. We note that various carbon-nitride 2D systems, like CN[31], C_2N[32], C_3N_4[33] and C_3N[34] are among the most attractive 2D semiconductors that have been experimentally fabricated and exhibit promising performances for a wide-range of applications. BC_3[35] has been also experimentally fabricated, C_4N and C_7N_6, C_9N_4 and $C_{10}N_3$ have been theoretically predicted by Li et al. [36] and Mortazavi et al. [37], respectively. We also consider carbon-free binary structures of hexagonal boron-nitride (BN), 2H transition metal dichalcogenides (MX_2, M= Mo, W and X=S, Se, Te), SiP_2 and As_2Se_3 monolayers. Similar to the case of monoelemental 2D lattices, we took MTPs with 1009 parameters and trained them on 1ps long AIMD trajectories at the temperature of 50 K.

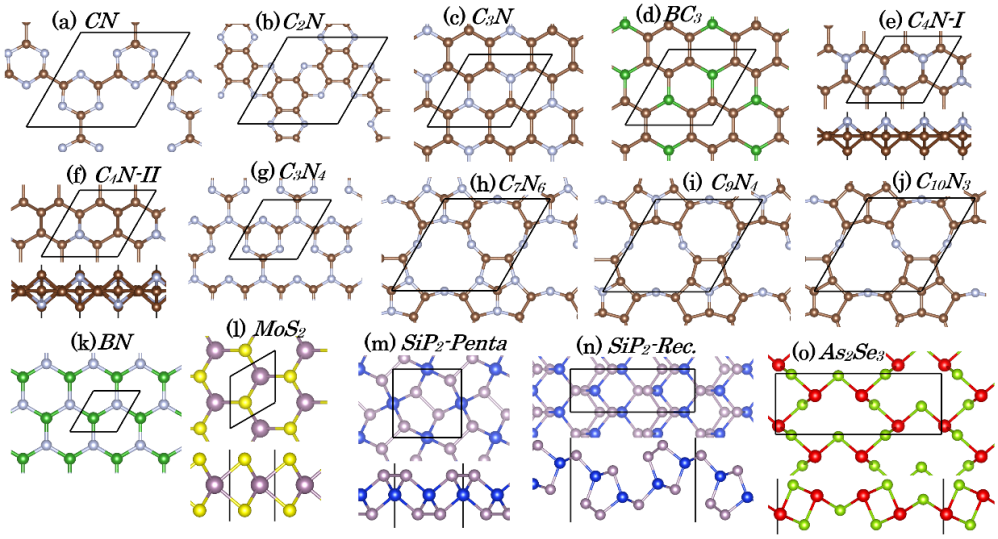

Fig. 4, Top views of atomic structures of binary 2D systems. Side views are shown only for non-planar lattices. Black lines illustrate the primitive unitcell.

The predicted PDRs by MTPs are compared with those by the DFPT method for the binary 2D lattices in Fig. 5. As it is clear, passively-trained MTPs can very accurately reproduce the PDRs for the considered binary lattices. It is noticeable that PDRs on the basis of MTPs are free of slight negative branches that have occurred around the Γ points in the DFTP results for some samples. The C_3N_4 monolayer with a completely flat structure shows conspicuous negative branches, which are well reproduced by the MTP. These results are highly promising taking into account the complexity of considered structures, especially rectangular SiP_2 and As_2Se_3 exhibit intricate buckled lattices. It should also be noted that while the acoustic modes are very closely reproduced for all considered samples, slight deviations are observable for some of the optical modes in few samples, particularly for the C_2N and BC_3 monolayers. As discussed for the case of graphene, to better reproduce the optical modes, the training set should be improved by including the AIMD trajectories at higher temperatures.

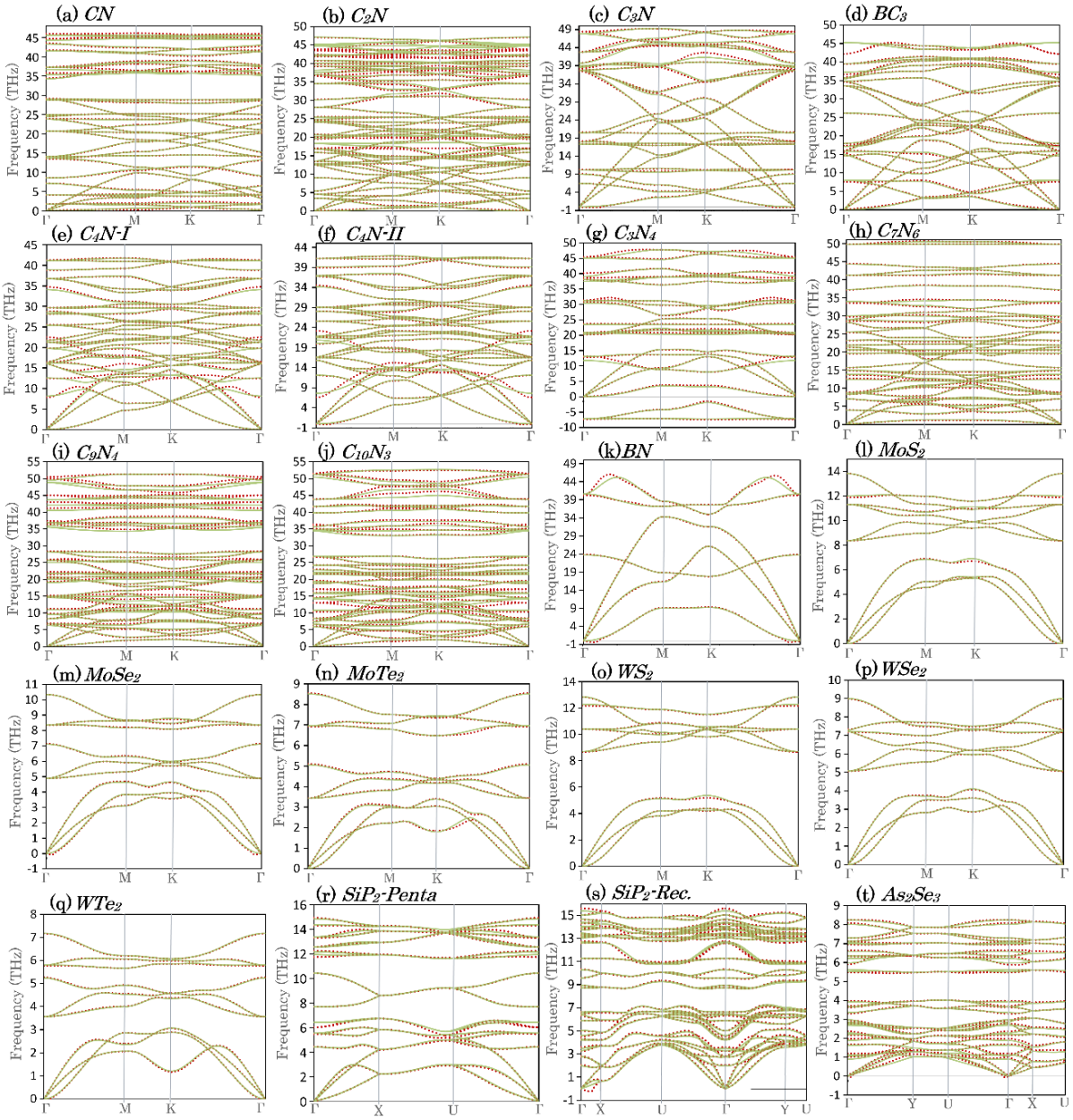

Fig. 5, Phonon dispersion relations of binary 2D lattices acquired by the DFPT (red-dotted lines) and first-attempt MTPs (continuous green lines). The corresponding atomic lattices are show in Fig. 4.

We next explore the accuracy of the PDRs predictions by MTP method for ternary 2D lattices. In Fig. 6 we compare the PDRs for six different ternary lattices by the DFTP and MTP methods. We note that $BC_{10}N_2$ is recently predicted by Tromer et al.[38], BC_6N shows hexagonal and rectangular atomic lattices, $BrCuTe_2$ and $ICuTe_2$ are also predicted according to their bulk counterparts and BC_6N_6 is a novel ternary lattice that we examine its stability. It is noticeable that negative branches in the DFTP results are also well reproduced by the passively trained MTPs. The results consistent with our previous observations, highlight that the accuracy of MTPs are insensitive to the number of elements' types and/or complexity of the structure. Despite we limited our AIMD simulations for less than 1 ps at 50 K, our results shown up to this stage reveal that for the most of cases no additional AIMD trajectories are required to assess the phonon-dispersions. Nonetheless, if required incorporation of additional AIMD trajectories at high temperatures in the training set are expected to improve the accuracy and stability as well.

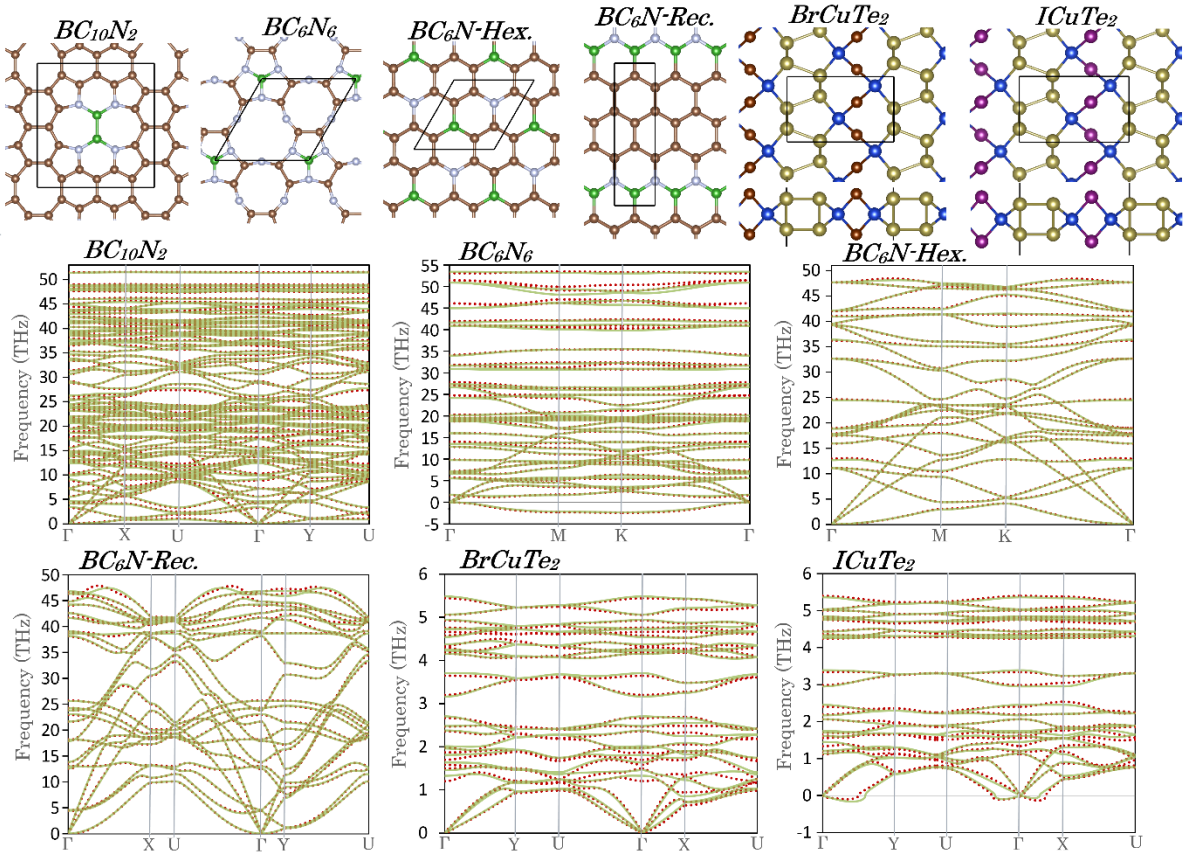

Fig. 6, Phonon dispersion relations for six different ternary 2D lattices acquired by the DFPT (red-dotted lines) and first-attempt MTPs (continuous green lines).

Besides the phonon dispersions, the phonons group velocity is another important phononic property, which can provide useful information concerning the lattice thermal conductivity. In Fig. 7 we compare the phonon group velocities predicted by MTP method with those obtained within the DFT using the PHONOPY code[25]. Remarkably, passively trained MTPs can very closely reproduce the phonon group velocities for the studied samples. For some of the samples like graphene, penta-graphene and -SiP_2, CN, BC_6N_6 and transition metal

dichalcogenides (MX_2, M= Mo, W and X=S, Se, Te) excellent agreements between the MTP method and DFT based estimations are observable. It is clear that in the worst cases, like $BrCuTe_2$, BC_3 or C_2N the group velocities are slightly up- or down-shifted and the results are free of substantial inaccuracies.

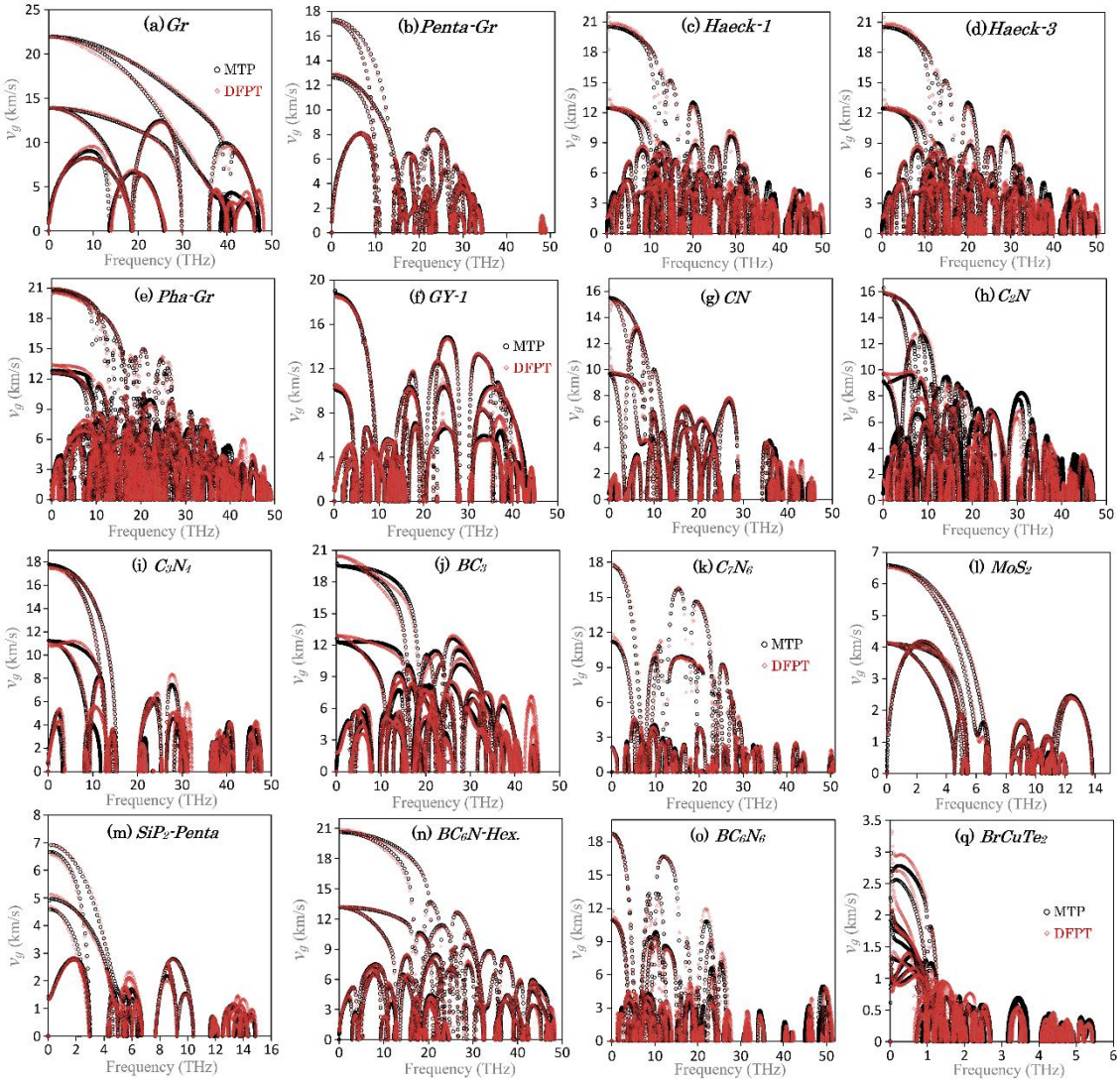

Fig. 7, Phonon group velocity (v_g) predicted by MTP and DFT methods.

Last but not least we examine the agreement between the MTP- and DFT-based approaches to calculate the free energy, heat capacity, and entropy from their statistical thermodynamic expressions using the PHONOPY code[25]. We compared the MTP- and DFT-based estimations for the aforementioned thermal properties in Fig. 8. As it is clear, the observable agreements are impressive, which accordingly clear any concern to extend the training set in order to improve the accuracy.

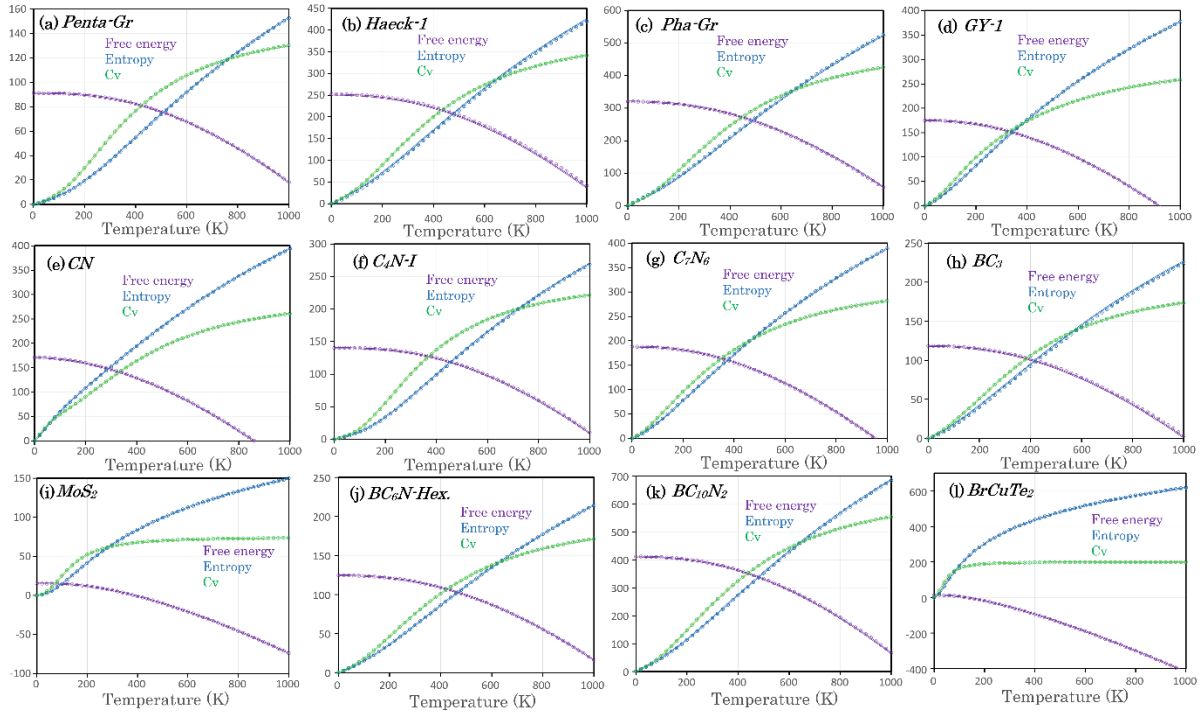

Fig. 8, Free energy, heat capacity, and entropy with the units of kJ/mol, J/K/mol, and J/K/mol, respectively, on the basis of MTP (continuous lines) and DFT (circles) methods acquired using the PHONOPY code[25].

Our results for a wide variety of 2D structures examined in this work highlight the accuracy and stability of the MTP-based method to explore the phononic properties of complex 2D lattices. The required AIMD simulations to assess the phononic properties are computationally less expensive than the those conducted to examine the thermal stability, which usually require more than 10 ps long AIMD simulations. Nonetheless, we believe that for highly symmetrical structures like graphene, transition metal dichalcogenides with 2H lattice and penta-graphene, the standard DFPT method is by a large extent more computationally efficient. Accordingly, the computational efficiency of MTP enhances as the symmetry decreases, like that for haeckelites and C_7N_6, or for materials with large primitive unitcells, such as metal- or conductive-organic frameworks and graphyne/graphdiyne lattices. Moreover, acquired results highlight that MTP-based approach generally offers a simple and convenient solution. We note that in this approach changing the supercell size for the evaluation of PDRs can be achieved with negligible computational costs. Moreover, in this approach the effects of plane-wave cutoff energy or k-point grids in the force constant calculations basically vanish, though they should be carefully examined within the AIMD step. Additionally, slight changes in the structure and lattice parameters can be easily examined with the MTP-based approach, without the need for retraining of existing MTP. One of the most prominent advantages of the MTP-based method is that it usually yields stable results without nonphysical imaginary branches in the acquired PDRs. This is an important feature, knowing that slight negative branches in PDRs may results in the instabilities for evaluating other thermal properties, particularly when calculating the lattice thermal conductivity. Although the majority of our results obtained on the basis of short AIMD trajectories at low temperature reveal outstanding accuracy, it is nonetheless highly recommended to include

AIMD trajectories at higher temperatures to ensure improved stability and accuracy as well. Worthy to remind that since AIMD simulations are available by a vast number of free and open-source first-principles packages, the proposed methodology by this study can be efficiently employed to conveniently examine the phononic properties of complex structures.

4. Conclusion

Our extensive results for a wide variety of 2D materials, highlight that machine-learning interatomic potentials trained over short ab-initio molecular dynamics trajectories are able to reproduce the phononic properties in strikingly close agreements with those by DFT based results. The proposed methodology could therefore play a pivotal role to conveniently explore the phononic properties of a large variety of low-symmetrical and porous nanomembranes, with a high level of accuracy and reproducibility. In recent years astonishing advances have been achieved on the synthesis of low-symmetrical and highly porous atomic lattices, like metal- or conductive-organic frameworks and graphdiyne nanomembranes. It is clear that phononic properties of these novel systems can be now effectively explored using the proposed approach by this work, which would otherwise require overdemanding computational resources with DFT based methods. To facilitate the practical application, we include the full details of the proposed approach in the data availability section.

Acknowledgment

B.M. and X.Z. appreciate the funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy within the Cluster of Excellence PhoenixD (EXC 2122, Project ID 390833453). E.V.P, I.S.N., and A.V.S. were supported by the Russian Science Foundation (Grant No 18-13-00479).

Data availability

The following data are available to download: (1) a brief guide for the installation of MLIP package, (2) all energy minimized lattices, (3) passive training approach and related commands, (4) VASP input script for the AIMD simulations, (5) VASP output file (vasprun.xml) for the DFPT calculations of all considered 38 examples, (6) created training sets and trained MTPs (p.mtp) for the all considered 38 examples, (7) the C++ code to calculate the force constants using the PHONOPY and MTP for the input geometries and force calculator, respectively and (8) PHONOPY related scripts to adjust the outputs, from: GITLAB: and Mendeley:

References

1. Dove, M. T. Introduction to the theory of lattice dynamics. *École thématique la Société Française la Neutron.* (2011). doi:10.1051/sfn/201112007
2. Dove, M. T. Introduction to lattice dynamics. *Introd. to lattice Dyn.* (1993). doi:10.1119/1.17708
3. Novoselov, K. S. *et al.* Electric field effect in atomically thin carbon films. *Science* **306**, 666–9 (2004).

4. Geim, A. K. & Novoselov, K. S. The rise of graphene. *Nat. Mater.* **6**, 183–191 (2007).
5. Wang, Y. & Ding, Y. Strain-induced self-doping in silicene and germanene from first-principles. *Solid State Commun.* **155**, 6–11 (2013).
6. Radisavljevic, B., Radenovic, A., Brivio, J., Giacometti, V. & Kis, A. Single-layer MoS2 transistors. *Nat. Nanotechnol.* **6**, 147–50 (2011).
7. Baughman, R. H., Eckhardt, H. & Kertesz, M. Structure-property predictions for new planar forms of carbon: Layered phases containing sp^{2} and sp atoms. *J. Chem. Phys.* **87**, 6687 (1987).
8. Jariwala, D., Sangwan, V. K., Lauhon, L. J., Marks, T. J. & Hersam, M. C. Emerging device applications for semiconducting two-dimensional transition metal dichalcogenides. *ACS Nano* **8**, 1102–1120 (2014).
9. Mortazavi, B., Shahrokhi, M., Cuniberti, G. & Zhuang, X. Two-Dimensional SiP, SiAs, GeP and GeAs as Promising Candidates for Photocatalytic Applications. *Coatings* **9**, 522 (2019).
10. Sun, B. & Barnard, A. S. Visualising multi-dimensional structure/property relationships with machine learning. *J. Phys. Mater.* **2**, 34003 (2019).
11. Oda, H., Kiyohara, S. & Mizoguchi, T. Machine learning for structure determination and investigating the structure-property relationships of interfaces. *J. Phys. Mater.* **2**, 34005 (2019).
12. Schleder, G. R., Padilha, A. C. M., Acosta, C. M., Costa, M. & Fazzio, A. From DFT to machine learning: recent approaches to materials science–a review. *J. Phys. Mater.* **2**, 32001 (2019).
13. Schmidt, J., Marques, M. R. G., Botti, S. & Marques, M. A. L. Recent advances and applications of machine learning in solid-state materials science. *npj Comput. Mater.* **5**, 83 (2019).
14. Podryabinkin, E. V & Shapeev, A. V. Active learning of linearly parametrized interatomic potentials. *Comput. Mater. Sci.* **140**, 171–180 (2017).
15. Gubaev, K., Podryabinkin, E. V., Hart, G. L. W. & Shapeev, A. V. Accelerating high-throughput searches for new alloys with active learning of interatomic potentials. *Comput. Mater. Sci.* **156**, 148–156 (2019).
16. Podryabinkin, E. V., Tikhonov, E. V., Shapeev, A. V. & Oganov, A. R. Accelerating crystal structure prediction by machine-learning interatomic potentials with active learning. *Phys. Rev. B* **99**, 064114 (2019).
17. Ladygin, V. V, Korotaev, P. Y., Yanilkin, A. V & Shapeev, A. V. Lattice dynamics simulation using machine learning interatomic potentials. *Comput. Mater. Sci.* **172**, 109333 (2020).
18. Korotaev, P., Novoselov, I., Yanilkin, A. & Shapeev, A. Accessing thermal conductivity of complex compounds by machine learning interatomic potentials. *Phys. Rev. B* **100**, 144308 (2019).
19. Shapeev, A. V. Moment tensor potentials: A class of systematically improvable interatomic potentials. *Multiscale Model. Simul.* **14**, 1153–1173 (2016).
20. Kresse, G. & Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **6**, 15–50 (1996).
21. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).
22. Kresse, G. From ultrasoft pseudopotentials to the projector augmented-wave

method. *Phys. Rev. B* **59**, 1758–1775 (1999).
23. Perdew, J., Burke, K. & Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
24. Monkhorst, H. & Pack, J. Special points for Brillouin zone integrations. *Phys. Rev. B* **13**, 5188–5192 (1976).
25. Togo, A. & Tanaka, I. First principles phonon calculations in materials science. *Scr. Mater.* **108**, 1–5 (2015).
26. Gubaev, K., Podryabinkin, E. V. & Shapeev, A. V. Machine learning of molecular properties: Locality and active learning. *J. Chem. Phys.* **148**, 241727 (2018).
27. Zhang, S. *et al.* Penta-graphene: A new carbon allotrope. *Proc. Natl. Acad. Sci. U. S. A.* (2015). doi:10.1073/pnas.1416591112
28. Terrones, H. *et al.* New Metallic Allotropes of Planar and Tubular Carbon. *Phys. Rev. Lett.* **84**, 1716–1719 (2000).
29. Wang, Z. *et al.* Phagraphene: A Low-Energy Graphene Allotrope Composed of 5-6-7 Carbon Rings with Distorted Dirac Cones. *Nano Lett.* **15**, 6182–6186 (2015).
30. Baughman, R. H., Eckhardt, H., Kertesz, M., Baughman, R. H. & Eckhardt, H. Structureproperty predictions for new planar forms of carbon: Layered phases containing sp2 and sp atoms Structure-property predictions for new planar forms of carbon : Layered phases containing Sp2 and sp atoms. *J. Chem. Phys.* **11**, 6687–6699 (1987).
31. Popov, V. N. & Lambin, P. Theoretical Raman fingerprints of $\ensuremath{\alpha}$-, $\ensuremath{\beta}$-, and $\ensuremath{\gamma}$-graphyne. *Phys. Rev. B* **88**, 75427 (2013).
32. Mahmood, J. *et al.* Nitrogenated holey two-dimensional structures. *Nat. Commun.* **6**, 1–7 (2015).
33. Algara-Siller, G. *et al.* Triazine-based graphitic carbon nitride: A two-dimensional semiconductor. *Angew. Chemie - Int. Ed.* **53**, 7450–7455 (2014).
34. Mahmood, J. *et al.* Two-dimensional polyaniline (C3N) from carbonized organic single crystals in solid state. *Proc. Natl. Acad. Sci.* **113**, 7414–7419 (2016).
35. Tanaka, H. *et al.* Novel macroscopic BC3honeycomb sheet. *Solid State Commun.* (2005). doi:10.1016/j.ssc.2005.06.025
36. Li, L. *et al.* Carbon-Rich Carbon Nitride Monolayers with Dirac Cones: Dumbbell C4N. *Carbon N. Y.* **118**, 285–290 (2017).
37. Mortazavi, B., Shahrokhi, M., Shapeev, A. V, Rabczuk, T. & Zhuang, X. Prediction of C7N6 and C9N4: stable and strong porous carbon-nitride nanosheets with attractive electronic and optical properties. *J. Mater. Chem. C* **7**, 10908–10917 (2019).
38. Tromer, R. M., Felix, I. M., Freitas, A., Azevedo, S. & Pereira, L. F. C. Diboron-porphyrin monolayer: A new 2D semiconductor. *Comput. Mater. Sci.* **172**, 109338 (2019).