Electronic Supplementary Material (ESI) for Nanoscale Horizons.

Revealing Improved Stability of Amorphous Boron-Nitride upon Carbon Doping

Onurcan Kaya,[†,‡] Luigi Colombo,[¶] Aleandro Antidormi,[†] Mario Lanza,[§] and Stephan Roche[*,†,‖]

†*Catalan Institute of Nanoscience and Nanotechnology (ICN2), CSIC and BIST, Campus UAB, Bellaterra, 08193, Barcelona, Spain*

‡*School of Engineering, RMIT University, Melbourne, Victoria, 3001, Australia*

¶*Department of Materials Science and Engineering, The University of Texas at Dallas, Richardson, TX, 75080, United States*

§*Department of Material Science and Engineering, King Abdullah University of Science and Technology, Thuwal 23955, Saudi Arabia*

‖*ICREA Institucio Catalana de Recerca i Estudis Avancats, 08010 Barcelona, Spain*

E-mail: stephan.roche@icn2.cat

Methods

Training of the GAP potential

Training procedure The GAP potential is trained and validated using two distinct datasets, called "training" and "validation" databases, respectively.[1–4] Each of the datasets contains a sufficiently large set of atomic energies, forces and stresses derived by DFT calculations and ab-initio MD simulations. The training set is quite large and includes both results from DFT and ab-initio MD simulations to include the instrinsic randomness which enhance the performance of ML process.[5,6]

The training of the potential is realized through a multi-step procedure in which the training database is gradually enlarged. Initially, a) a training dataset of energies, forces and stresses is built using atomistic structures generated using ab-initio Molecular Dynamics and b) a first GAP potential is trained; c) using the previously trained potential several new atomistic structures are generated using classical MD and their energies and forces are consequently evaluated using DFT calculations. d) A new GAP potential is subsequently trained on the new larger dataset and validated. These steps are cyclically repeated several times until sufficient accuracy in the potential predictions is reached.

The training and validation sets were initially populated with energies and forces extracted on a set of atomistic srtuctures of C-doped amorphous boron nitride with a variable concentration of carbon atoms, Boron Nitride in its amorphous and crystalline phases and carbon polymorphs. Specifically, several atomistic samples were generated corresponding to different phases of C-doped boron nitride and simple carbon at different densities and temperatures: crystalline h-BN, amorphous boron nitride with variable Carbon content (from 0% to 80%), amorphous carbon, graphene, graphite and diamond. The configurations of amorphous samples are characterized by a wide range of densities, from 1.0g/cm^3 to 3.5g/cm^3. Furthermore, several samples were generated corresponding to distinct molecular configurations involving B, N and C chemical species (C_2, N_2, CN_2, C_2N_2, C_3N_4, B_4C, B_4C_2, B_4C_3). Finally, the

isolated atoms of C, B and N were also added to the training database.

The final training dataset contains a number of different atomistic configurations as large as 2300 yielding a number of local environments as large as 84000. Final validation set has 2000 different structures.

DFT calculations The first-principles calculations have been performed using the DFT code as implemented in the Quantum ESPRESSO package[7–9]with PBE exchange-correlation functional[10] and Projector augmented wave (PAW) pseudopotentials. We used energy cut-offs of 75 Ry and 600 Ry for the wavefunction and electronic density, respectively. Periodic boundary conditions are generally applied to the cell with the Brillouin zone sampled at the Γ point. When systems have been relaxed, relaxation steps were performed until all the atomic forces were smaller than 0.001 a.u. In all the calculations, we used the electronic free energy as the energy to be predicted, whose derivatives yield the atomic forces according to the Hellman-Feynman theorem.

The ab-initio molecular dynamics calculations were performed using the Car-Parrinello CP package implemented in Quantum ESPRESSO under a different set of parameters. In particular, UltraSoft Pseudo potentials have been used with a LDA as exchange-correlation functional;[11] a kinetic energy cutoff for wavefunctions and energy have been set to 150 Ry and 20 Ry cutoff, respectively.

Training the GAP potential The training database was employed to build a GAP potential using the quippy code (www.libatoms.org) with the parameters as summarized in table 1: the columns refer to the 2-body, 3-body and the Smooth Overlap of Atomic Positions (SOAP)[2] descriptors, respectively. The latter has been recently introduced to model the many-body interactions and has proved useful in the generation of GAP models for tungsten, amorphous Carbon and Silicon.[12–15] For the meaning of the reported values and the mathematical framework of the GAP potential we refer to Ref.[16] Moreover, regularization parameters for expected errors on energies and forces were set to 0.002 eV and 0.2 eV/ Å,

respectively. Finally, sparsification with the CUR method[17] has been adopted for the SOAP kernel while sparsification on a uniform grid has been the choice for the 2-body and 3-body terms.

Table 1: Parameters used to train the GAP potential for C-doped a-BN.

	2-body	3-body	SOAP
δ (eV)	2.0	0.1	0.1
r_{cut} (Å)	3.7	3.0	3.7
r_{Δ} (Å)			0.5
σ_{at} (Å)			0.5
n_{max}, l_{max}			8
ζ			4
Sparsification	Uniform	Uniform	CUR
N_t (a C-BN bulk)		150	2000
N_t (crystalline samples)		50	500
N_t (total)	15	200	2500

Validation of the GAP potential In order to assess the quality of the trained GAP for future predictions on C-doped BN samples, we compared the energies per atom and forces obtained with the GAP with the exact values from DFT calculations. Fig. 1 shows the scatterplots of the mentioned values from DFT and from the GAP potentials as computed on the training and validation datasets. A significantly small value for the Root Mean Squared Error (RMSE) is found in all the training data, slightly increasing on the validation set (since the latter has not been used to train the potential). The overall agreement is consequently good and no trace of overfitting is observed.

Fig. 2 shows the statistical distribution of the errors on the energies (absolute value) and forces (left and right top panels). The corresponding cumulative functions are also shown in the corresponding bottom panels.

As a final validation step of the GAP potential, we assessed the capability of the GAP potential to allow for the generation of atomistic samples with morphological and structural properties similar to those obtained via ab-initio MD. To this aim we performed a systematic comparison of the structural properties of samples generated with the two methods. In order

to perform a consistent comparison, we built systems with 200 atoms and a C-content of 5%. The samples are analyzed at two different temperatures, 300K (in the solid phase) and 5000K (in the liquid phase) and several densities. A standard quench-from-melt protocol has been employed to generate the samples, as described below: the systems were first equilibrated in the liquid phase at 5000K for then cooled at 300K with a cooling rate of 100K/ps.

Extremely good agreement between the DFT and GAP curves with respect both to the position of the peaks and their relative amplitude, for the considered temperatures and densities. In particular, all the samples show a complete lack of structural order for distances larger than 4 Å. This confirms that our GAP potential is suitable and accurate to perform a systematic analysis of BN in amorphous form upon Carbon doping.

Sample generation

The procedure we followed to generate the samples of amorphous boron nitride with varying carbon content consists of a sequence of different computational steps.[18] First, we started from an atomic configuration of Boron, Nitrogen and Carbon atoms randomly located in the cell volume; then, we equilibrated the system at high temperature (5000K) for 10 ps (timestep 0.25 fs) using a Nosé-Hover thermostat in order for the system to be found in a liquid melt state.In all the samples considered in this work, the mass density is chosen to be close to the experimentally observed value,[19] ~ 2.1 g/cm^3. We subsequently reduced the temperature to 300K in the constant-volume, constant-temperature (NVT) ensemble with a cooling rate of 100K/ps. The latter parameter has been chosen in order to obtain realistic amorphous samples yet maintaining the computational cost at a reasonable level. The main structural features and results of this work have proved not to be affected by the accessible range of cooling rate.

Finally, the structures thus obtained are then relaxed at 300K for 3 ps and eventually equilibrated for 3 ps at the same temperature in the NPT ensemble to let the system relax the size of the cell. Periodic boundary conditions have been applied to the cell and the

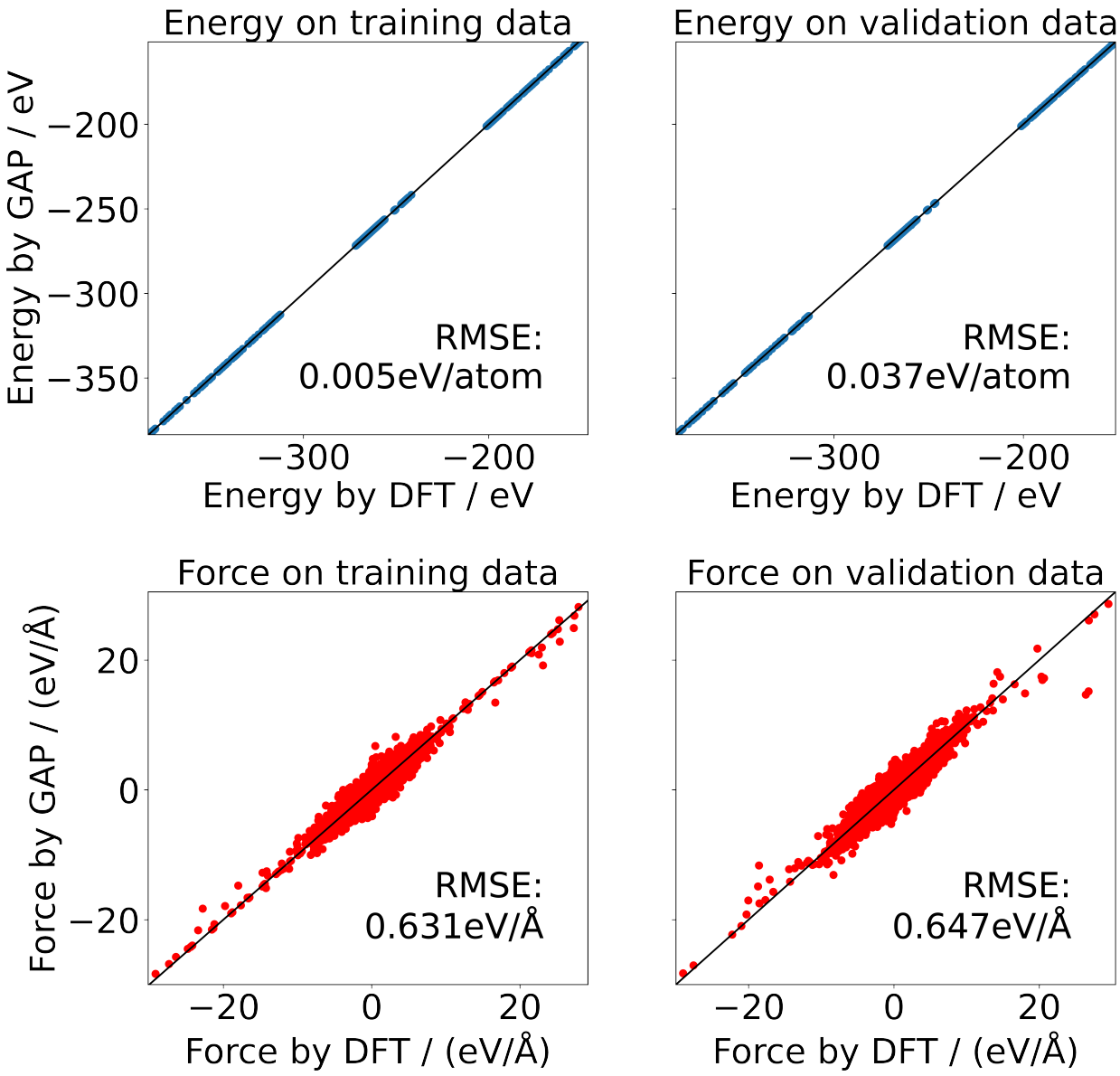

Figure 1: Scatter plots of the energies and forces obtained using GAP as a function of the corresponding DFT values. Values are computed on the training (left) and validation (right) datasets.

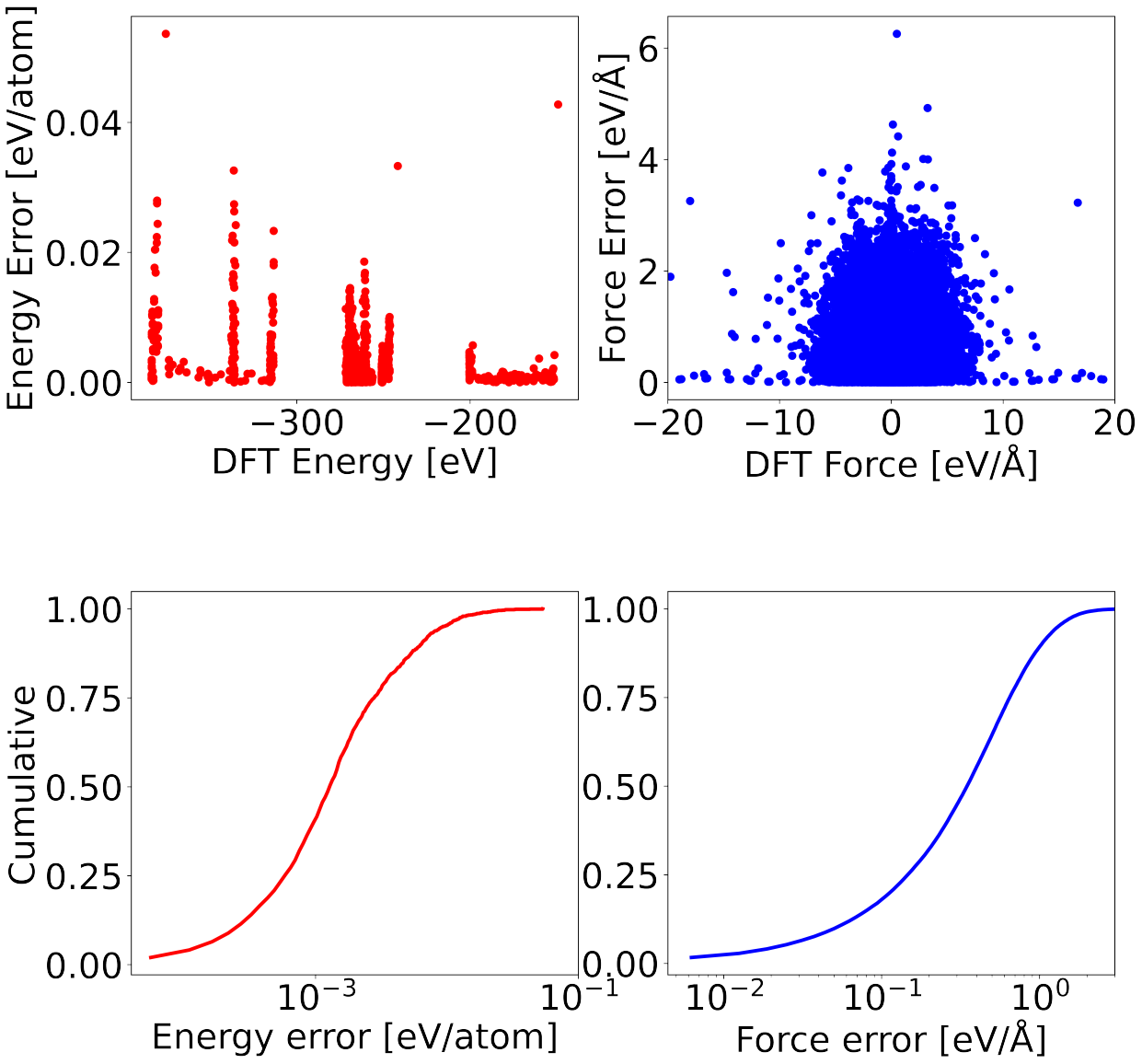

Figure 2: (top panels) Absolute values of the errors of energies and atomic forces as a function of the corresponding DFT values for a set of 300 a-BN structures with varying C-content. (bottom panels) Cumulative distribution functions of the absolute values of errors of energies and forces as extracted from the same set of generated atomistic samples.

evolution of the system is performed by ageing the atomic trajectories with the velocity-Verlet algorithm.

Bonding Character

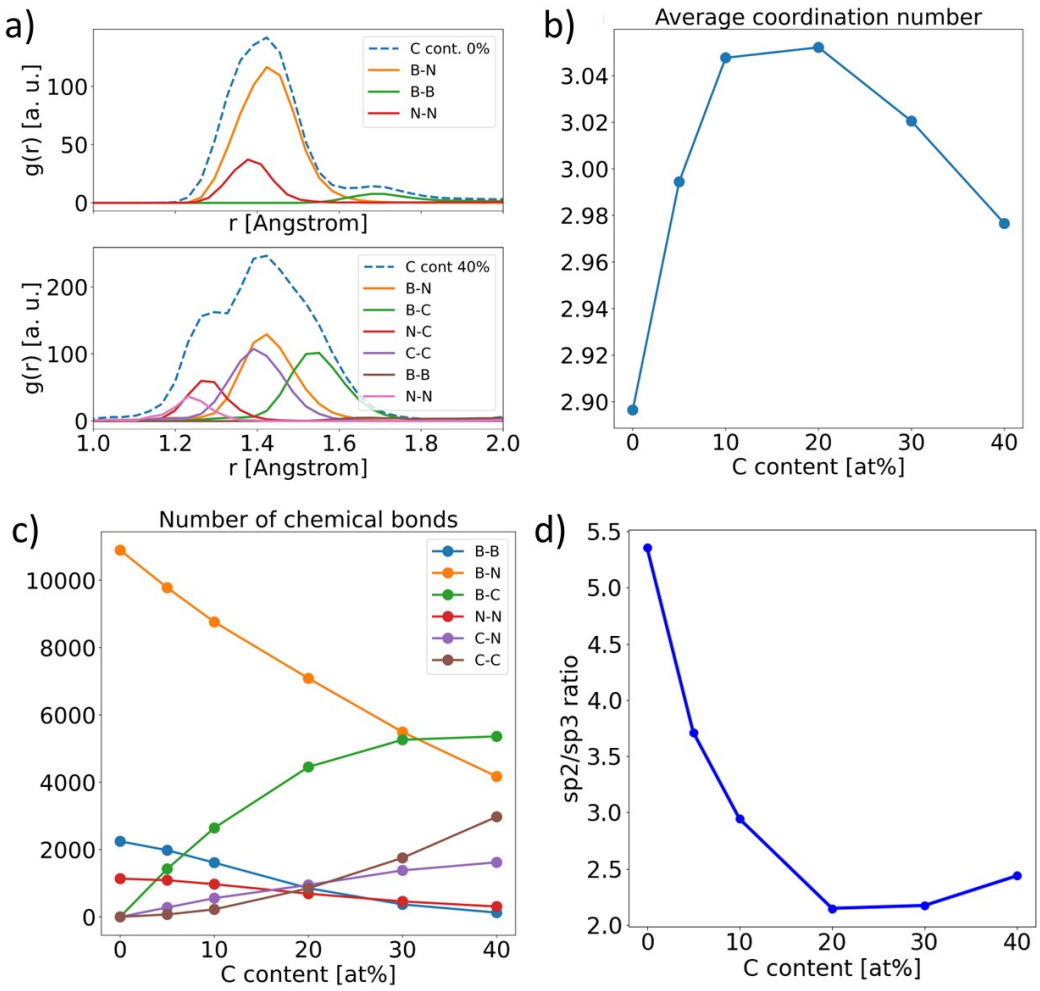

Figure 3: (a) RDF for specific atomic bonding; (b) average coordination number versus carbon-content, (c) evolution of the number of chemical bonds with C concentration; (d) evolution of the ratio number of sp^2-hybridized bonds with sp^3-hybridized bonds with C concentration.

Diffusivity

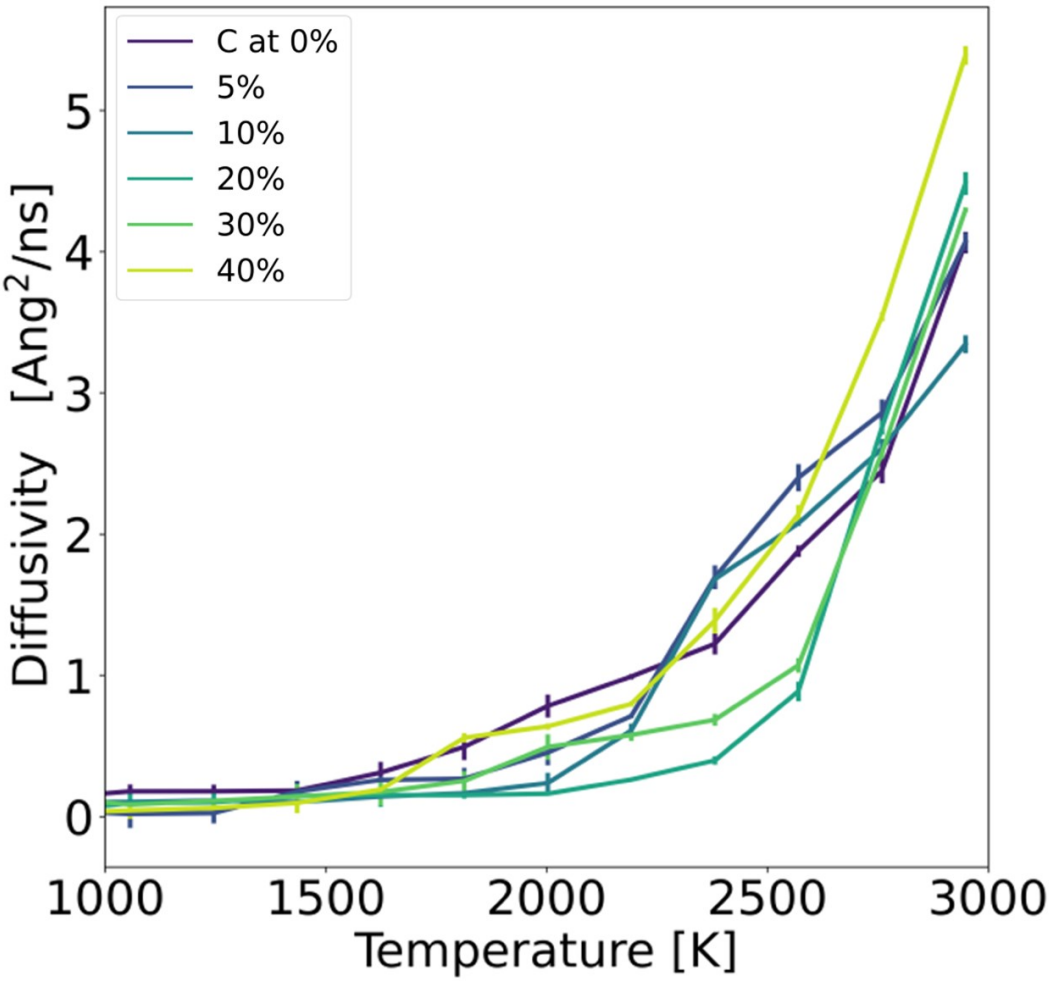

Figure 4: Diffusivity of atoms as a function of temperature for different samples of α-BN:C. The values have been averaged over 5 samples for each given C-content. Vertical bars show the statistical error.

References

(1) Bartók, A. P.; Payne, M. C.; Kondor, R.; Csányi, G. Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. *Physical review letters* **2010**, *104*, 136403.

(2) Bartók, A. P.; Kondor, R.; Csányi, G. On representing chemical environments. *Physical Review B* **2013**, *87*, 184115.

(3) Bartók, A. P.; Csányi, G. G aussian approximation potentials: A brief tutorial introduction. *International Journal of Quantum Chemistry* **2015**, *115*, 1051–1057.

(4) Deringer, V. L.; Bartók, A. P.; Bernstein, N.; Wilkins, D. M.; Ceriotti, M.; Csányi, G. Gaussian process regression for materials and molecules. *Chemical Reviews* **2021**, *121*, 10073–10141.

(5) Çağla Odabaşı,; Yıldırım, R. Performance analysis of perovskite solar cells in 2013–2018 using machine-learning tools. *Nano Energy* **2019**, *56*, 770–791.

(6) Dongale, T. D.; Sutar, S. S.; Dange, Y. D.; Khot, A. C.; Kundale, S. S.; Patil, S. R.; Patil, S. V.; Patil, A. A.; Khot, S. S.; Patil, P. J.; Bae, J.; Kamat, R. K.; Kim, T. G. Machine learning-assisted design guidelines and performance prediction of CMOS-compatible metal oxide-based resistive switching memory devices. *Applied Materials Today* **2022**, *29*, 101650.

(7) Giannozzi, P.; Baroni, S.; Bonini, N.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Chiarotti, G. L.; Cococcioni, M.; Dabo, I., et al. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *Journal of physics: Condensed matter* **2009**, *21*, 395502.

(8) Giannozzi, P.; Andreussi, O.; Brumme, T.; Bunau, O.; Nardelli, M. B.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Cococcioni, M., et al. Advanced capabilities for

materials modelling with Quantum ESPRESSO. *Journal of physics: Condensed matter* **2017**, *29*, 465901.

(9) Giannozzi, P.; Baseggio, O.; Bonfà, P.; Brunato, D.; Car, R.; Carnimeo, I.; Cavazzoni, C.; De Gironcoli, S.; Delugas, P.; Ferrari Ruffino, F., et al. Quantum ESPRESSO toward the exascale. *The Journal of Chemical Physics* **2020**, *152*, 154105.

(10) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Physical review letters* **1996**, *77*, 3865.

(11) Kohn, W.; Sham, L. J. Self-consistent equations including exchange and correlation effects. *Physical review* **1965**, *140*, A1133.

(12) Szlachta, W. J.; Bartók, A. P.; Csányi, G. Accuracy and transferability of Gaussian approximation potential models for tungsten. *Physical Review B* **2014**, *90*, 104108.

(13) De, S.; Bartók, A. P.; Csányi, G.; Ceriotti, M. Comparing molecules and solids across structural and alchemical space. *Physical Chemistry Chemical Physics* **2016**, *18*, 13754–13769.

(14) Deringer, V. L.; Bernstein, N.; Bartók, A. P.; Cliffe, M. J.; Kerber, R. N.; Marbella, L. E.; Grey, C. P.; Elliott, S. R.; Csányi, G. Realistic atomistic structure of amorphous silicon from machine-learning-driven molecular dynamics. *The journal of physical chemistry letters* **2018**, *9*, 2879–2885.

(15) Deringer, V. L.; Csányi, G. Machine learning based interatomic potential for amorphous carbon. *Physical Review B* **2017**, *95*, 094203.

(16) Deringer, V. L.; Csányi, G. Machine learning based interatomic potential for amorphous carbon. *Phys. Rev. B* **2017**, *95*, 094203.

(17) Mahoney, M. W.; Drineas, P. CUR matrix decompositions for improved data analysis. *Proceedings of the National Academy of Sciences* **2009**, *106*, 697–702.

(18) Antidormi, A.; Colombo, L.; Roche, S. Thermal transport in amorphous graphene with varying structural quality. *2D Materials* **2020**,

(19) Gautam, C.; Tiwary, C. S.; Jose, S.; Brunetto, G.; Ozden, S.; Vinod, S.; Raghavan, P.; Biradar, S.; Galvao, D. S.; Ajayan, P. M. Synthesis of Low-Density, Carbon-Doped, Porous Hexagonal Boron Nitride Solids. *ACS Nano* **2015**, *9*, 12088–12095, PMID: 26580810.