PHYSICAL REVIEW B **106**, L180101 (2022)

Letter

Machine learning interatomic potential for simulations of carbon at extreme conditions

Jonathan T. Willman,[1] Kien Nguyen-Cong,[1] Ashley S. Williams,[1] Anatoly B. Belonoshko,[2] Stan G. Moore,[3]
Aidan P. Thompson,[3] Mitchell A. Wood,[3] and Ivan I. Oleynik[1,*]
[1]*Department of Physics, University of South Florida, Tampa, Florida 33620, USA*
[2]*Department of Physics, University of Royal Institute of Technology, 106691 Stockholm, Sweden*
[3]*Sandia National Laboratories, Albuquerque, New Mexico 87185, USA*

(Received 25 April 2022; revised 22 September 2022; accepted 31 October 2022; published 30 November 2022)

A spectral neighbor analysis (SNAP) machine learning interatomic potential (MLIP) has been developed for simulations of carbon at extreme pressures (up to 5 TPa) and temperatures (up to 20 000 K). This was achieved using a large database of experimentally relevant quantum molecular dynamics (QMD) data, training the SNAP potential using a robust machine learning methodology, and performing extensive validation against QMD and experimental data. The resultant carbon MLIP demonstrates unprecedented accuracy and transferability in predicting the carbon phase diagram, melting curves of crystalline phases, and the shock Hugoniot, all within 3% of QMD. By achieving quantum accuracy and efficient implementation on leadership-class high-performance computing systems, SNAP advances frontiers of classical MD simulations by enabling atomic-scale insights at experimental time and length scales.

DOI: 10.1103/PhysRevB.106.L180101

Carbon at extreme pressures and temperatures is a topic of great scientific interest for several disciplines including planetary science [1–5] and inertial confinement fusion (ICF) research [6–9]. Methane ice at megabar pressures and temperatures of thousands of kelvins is predicted to convert to solid or liquid carbon in the cores of giant planets [2–4]. A successful suppression of hydrodynamic instabilities seeded by solid and liquid carbon phases appearing upon strong compression of the outer diamond ablation shell of an ICF capsule [10] was the key for achieving a record-breaking, near-threshold fusion energy ignition at the National Ignition Facility [11].

The exploration of carbon's behavior at extreme conditions is challenging for both theory and experiment. Shock and ramp compression experiments using powerful lasers [12], pulsed power [13], and bright x-ray sources [14,15] uncovered complicated mechanisms of inelastic deformations [16–21], anomalous strength of diamond [18–22], unusual melting [7,23–27], and liquid carbon properties [24,25], as well as extreme metastability of diamond well beyond the pressure-temperature range of its thermodynamic stability [28]. Molecular dynamics (MD) simulations can provide a fundamental understanding of these phenomena, but to be of experimental relevance, it must accurately describe interatomic interactions in a system consisting of a large number of atoms.

Previous simulations of carbon at extreme conditions were predominantly performed using quantum molecular dynamics (QMD) based on density functional theory (DFT) [24,29–36]. Due to high computational cost, QMD simulations are limited to several hundred atoms for up to tens of picoseconds, which is insufficient for uncovering nonequilibrium processes at experimental timescales (ns) and length (μm) scales. In principle, these scales can be accessed by classical MD simulations on massively parallel computers [37]. However, empirical interatomic potentials developed for carbon at near ambient conditions [38–42] singularly fail upon extension to high pressures and temperatures [43,44], thus compromising the predictive power of atomistic simulations.

The advent of machine learning interatomic potentials (MLIPs) [45,46] opens up exceptional opportunities for achieving a classical description of chemical bonding with quantum accuracy [47]. Although numerous MLIPs have been recently introduced and successfully applied to modeling properties of materials at ambient conditions [45,46,48–52], including carbon [53–56], their exceptional power in describing diverse and complex atomic environments at megabar pressures and tens of thousand of kelvins has yet to be demonstrated.

This Letter reports a significant advance in the development of a quantum-accurate spectral neighbor analysis potential (SNAP) for simulations of carbon at extreme pressure-temperature (P-T) conditions. This includes construction of an experimentally relevant training database, implementation of a robust SNAP machine learning training methodology, and extensive validation against QMD and experimental data. The end result is a MLIP that delivers extreme accuracy in simulating carbon over a remarkably wide range of pressures (from 0 to 50 Mbar) and temperatures (up to 20 000 K).

In general, MLIPs fingerprint a unique local atomic environment around each atom by a set of descriptors. SNAP's descriptors are bispectrum components $\{\mathbf{B}^i\}$ of the local neighbor density projected onto a basis of hyperspherical harmonics in four dimensions [48,49]. Other successful MLIPs—neural network potentials (NNPs) [45,52], Gaussian

---

[*] oleynik@usf.edu

©2022 American Physical Society

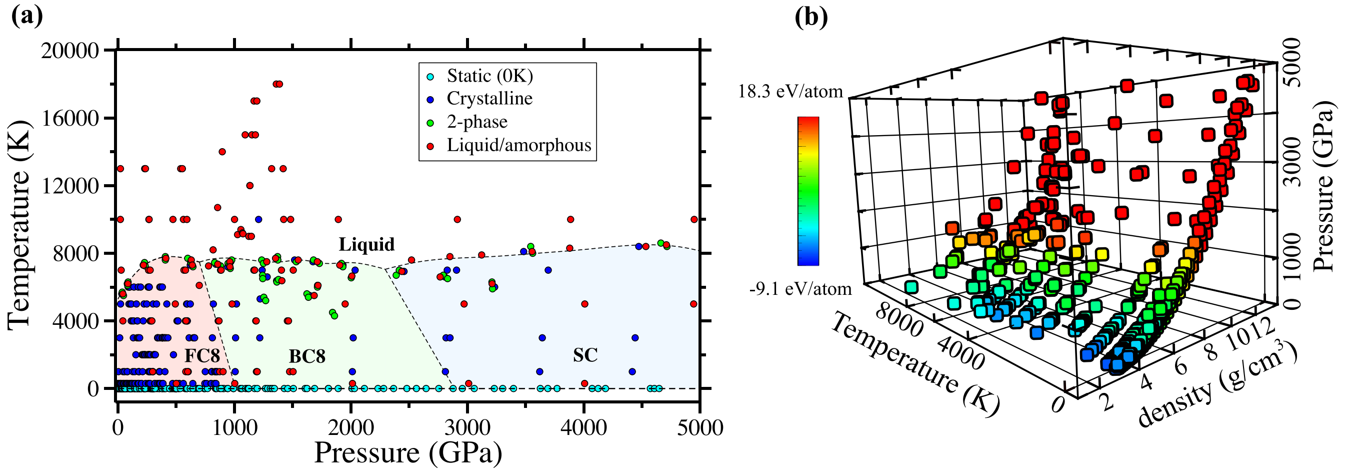

FIG. 1. SNAP training database. (a) Pressure-temperature map of QMD and static DFT simulations included in the database, each represented by a P-T point on carbon phase diagram sampling diamond (FC8), body-centered cubic (BC8), and simple cubic (SC) solid and liquid phases (total number of structures - 636). (b) Pressure-temperature-density-energy/atom distribution.

approximation potential (GAP) [46], and the moment tensor potential (MTP) [50]—employ mathematically different, but physically similar descriptors. All of them, including SNAP, can be mapped into a general atomic cluster expansion (ACE) descriptor framework [51].

Herein, the total potential energy of the system of N atoms is written as a sum of atomic energies E^i, which are quadratic functions of the bispectrum coefficients $\mathbf{B}^i$ [49],

$$E_{\text{tot}}(\{\mathbf{r}^N\}) = \sum_i E^i, \quad E^i = \boldsymbol{\beta}\cdot\mathbf{B}^i + \frac{1}{2}\mathbf{B}^i\cdot\boldsymbol{\alpha}\cdot\mathbf{B}^i. \qquad (1)$$

Machine learning techniques are used to determine the symmetric matrix $\boldsymbol{\alpha}$ and the vector $\boldsymbol{\beta}$, the unknown parameters of SNAP, to reproduce the potential energy, atomic forces, and stress tensor for each structure in the DFT training database.

SNAP displays a good balance between computational cost and accuracy [57]. Both are controlled by SNAP hyperparameters—the cutoff radius r_{cut} and the half-integer angular momentum J. The former specifies the number of neighbor atoms participating in the fingerprinting of the atomic environment around each atom i and the latter refers to the dimensionality of the descriptor space, i.e., the number of bispectrum descriptors $\{\mathbf{B}^i\}$: $(J+1)(J+3/2)(J+2)/3$ for each atom i.

The SNAP development includes (i) construction of a robust training database of first-principles QMD data, (ii) machine learning training, and (iii) extensive validation against QMD and experiment. These three tightly connected steps constitute a single development cycle. Several of such cycles are executed to improve upon deficiencies observed in previous iterations (see the development workflow in Supplemental Fig. S2 [58]). For example, during the validation, two-phase SNAP MD simulations produce melting lines of several high-pressure carbon phases in a substantial disagreement with QMD. The problem has been traced back to SNAP inaccuracies in calculation of enthalpies of solid and liquid phases along the melting line, which, according to the Clausius-Clapeyron relation, define its slope dT/dP. Therefore, separate liquid and solid phases were added to the QMD training database to complement original combined two-phase solid-liquid structures. This update resulted in substantial improvements in SNAP accuracy upon execution of a new training cycle.

One of the highlights of our MLIP development is the dedicated construction of a SNAP training database of experimentally relevant QMD of large systems (by QMD standards)—up to 1024 atoms. These are individual frames (one to three frames per simulation) from QMD production simulations of physical properties of carbon (melting lines, hydrostatic and uniaxial isotherms, shock Hugoniot) performed within a wide range of pressures (from 0 to 5 TPa), temperatures (from 0 to 40 000 K), and densities (from 2.9 to 13.6 g/cm^3). For example, MD frames for each (P, T) point along the melting lines of several crystalline carbon phases were taken from a production two-phase QMD simulation [59]. These complex cells contain both liquid and solid parts separated by a realistic solid-liquid interface, which add an additional complexity to the SNAP database. A series of complimentary QMD simulations of liquid and solid phases were also included in the database. The QMD data are supplemented by static DFT calculations of binding energy curves, point and extended defects, and metastable carbon structures obtained from dedicated crystal structure searching [60]. Figure 1 displays the range of pressures, temperatures, and energies covered by QMD simulations. The SNAP database, consisting of 636 structures, samples 124 907 unique atomic environments. Each structure of N atoms contributes 1 total energy, 6 stress components, and $3N$ atomic forces resulting in a total training complexity (the number of regression equations to fit) of 388 077. Higher temperatures (up to 40 000 K) were employed during SNAP development to make sure it is well behaved at very small interatomic distances occurring in energetic collisions of the atoms with atomic velocities from the tail of the Maxwell distribution even at temperatures below 20 000 K. The SNAP accuracy in this temperature interval between 20 000 and 40 000 K is not sufficient to attempt meaningful MD simulations. Additional information on database composition and its generation is provided in the Supplemental Material [58].

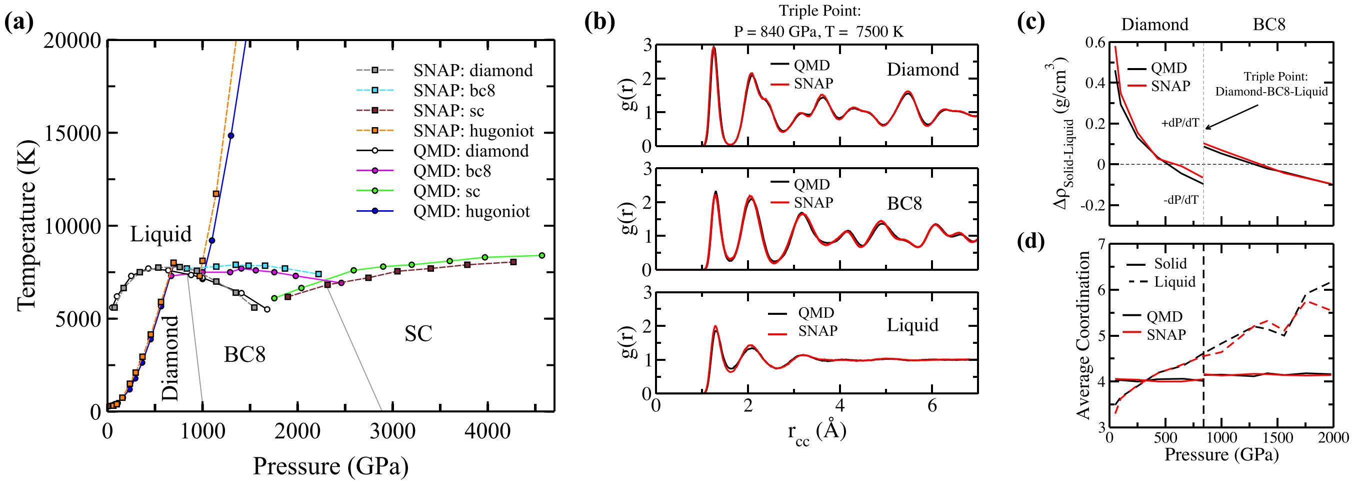

FIG. 2. Validation of SNAP against QMD. (a) Carbon phase diagram, including melting lines of diamond, BC8 and SC crystalline phases of carbon at pressures up to 5000 GPa, and hydrostatic shock Hugoniot. (b) Radial distribution functions $g(r)$ for diamond, BC8, and liquid phase at the diamond-BC8-liquid triple point (840 GPa, 7510 K). (MD trajectories are averaged over a 20 ps time interval.) (c) Density difference $\Delta\rho = \rho_s - \rho_l$ between solid (s) and liquid (l) phases (top panel) and average coordination number of carbon atoms (bottom panel) as a function of pressure along the melting lines of diamond and BC8. $\Delta\rho = 0$ corresponds to melting line maxima: diamond—at 500 GPa; BC8—at 1300 GPa.

Once the database is constructed, the SNAP model for the total potential energy, stress tensor components, and atomic forces is trained using machine learning techniques to determine SNAP parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ through minimization of an objective function—a sum of the normalized squared differences between DFT and SNAP energies, stresses, and atomic forces (Fig. S2) [58]. The training is performed in a series of iterations. For a given set of weights, the SNAP parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are determined through weighted linear regression as implemented in the FitSNAP package [61]. The weights are then optimized to minimize the objective function using a genetic algorithm (GA) within the DAKOTA software package [62]. At every step of GA minimization FitSNAP is called with the current set of weights to determine new $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, which are then fed back to update the objective function being minimized (Fig. S2) [58]. The iterations stop when the GA minimization converges to a final set of weights. In addition to group weights, the SNAP hyperparameters r_{cut} and J are optimized in an offset fashion to find the right balance between SNAP accuracy and computational efficiency (Fig. S2 [58]). The final values for the SNAP hyperparameters are $r_{\text{cut}} = 2.7$ Å and $J = 4$. The resultant quality of SNAP training is discussed in the Supplemental Material [58].

The critical part of SNAP potential development is a thorough validation of SNAP MD results against QMD and experimental data. Only one to three QMD frames per (P, T) state were included in the SNAP training database. Therefore, simulating these states with SNAP using much larger simulation cells and performing thermodynamic averaging of atomic trajectories containing tens of thousands of frames to obtain stresses, densities, internal energies, as well as radial distribution functions is considered as a rigorous validation test against QMD. Further, these validation simulations sample a variety of pressure/temperature points not included in the database but still within the SNAP' s intended interval of pressures (from 50 to 5000 GPa) and temperatures (from 0

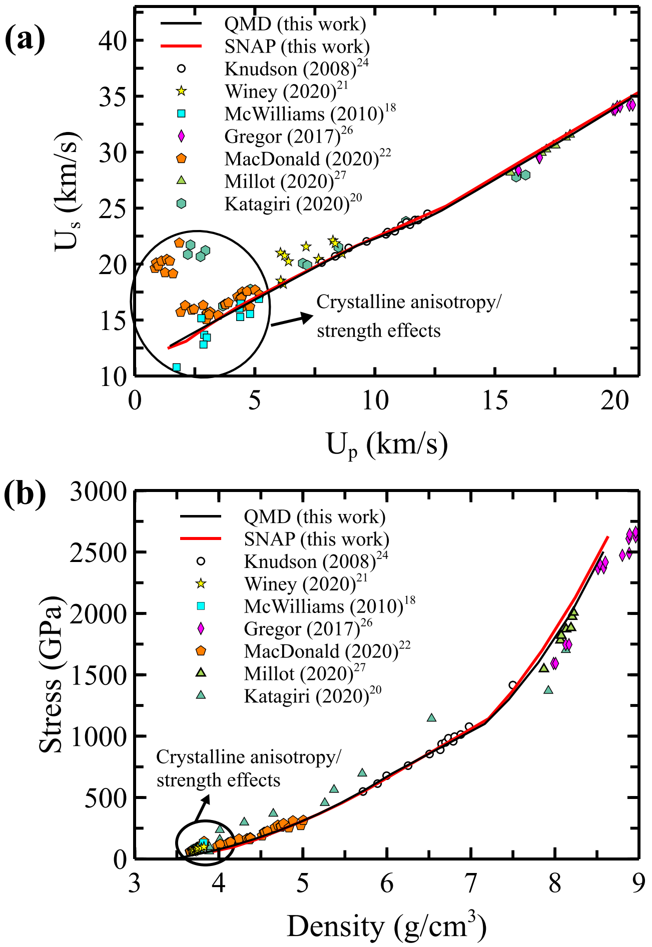

FIG. 3. Carbon shock Hugoniot calculated by QMD and SNAP and compared with experimental data. (a) U_s-U_p Hugoniot, and (b) pressure-density Hugoniots. Points correspond to experimental data.

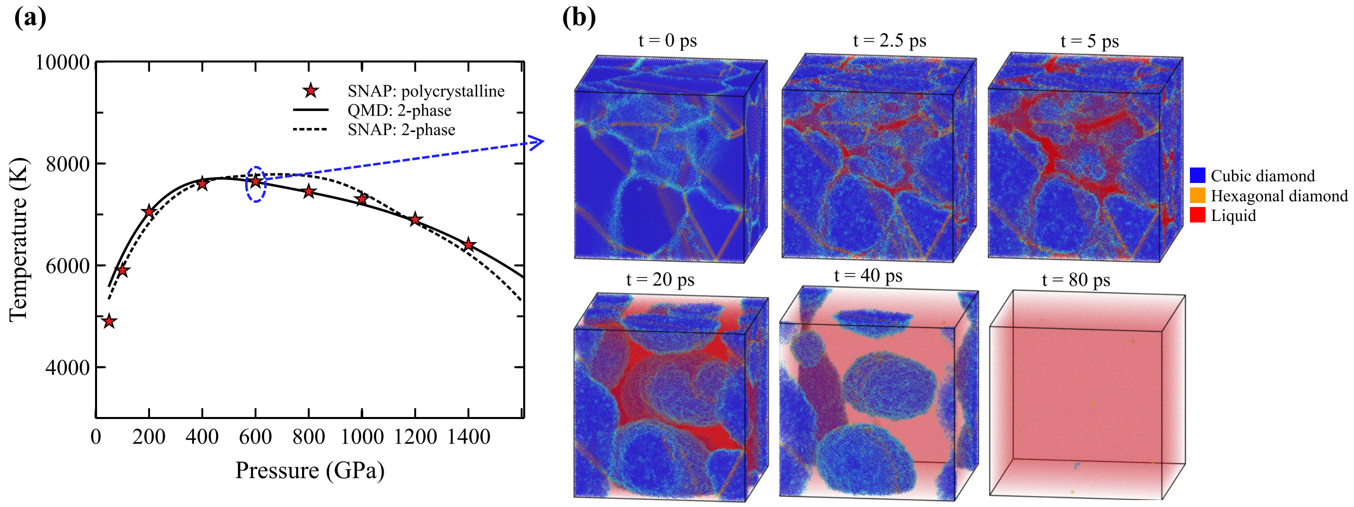

FIG. 4. (a) Comparison of diamond melting curves calculated by three methods: QMD two-phase (solid line), SNAP two-phase (dashed line), and direct melting of 1-million-atom polycrystalline sample using SNAP (stars). (b) Time progression of polycrystalline sample melting at 600 GPa and 7750 K. The sample is initially composed of crystalline diamond regions (blue) separated by grain boundaries (light blue). Liquid regions (red) emerge at the grain boundaries, grow in size, and eventually consume the isolated diamond crystallites.

to 20 000 K). SNAP transferability is demonstrated by good agreement with both QMD and experiment.

The important validation test is concerned with carbon phase diagram, including melting lines of several high-pressure phases at pressures up to 5000 GPa [see Fig. 2(a)]. A series of isobaric-isothermal NPT two-phase MD simulations were run at a given pressure but varying temperature to determine the P-T value of the phase coexistence [59]. SNAP melting lines are in excellent agreement with those from QMD, the average temperature error being ~200 K or 3% in a pressure interval from 0 to 5000 GPa. Figure 2(b) displays SNAP and QMD radial distribution functions $g(r)$ at the diamond-BC8-liquid triple point: They are almost indistinguishable from each other.

A remarkable property of carbon at extreme conditions is the negative slope (dT/dP) of diamond and BC8 melting lines at high pressures [25,30–33,36]. This is because liquid carbon becomes denser than the corresponding solid phase upon increase of pressure above ~500 GPa for diamond and ~1300 GPa for BC8 [see Fig. 2(c)]. This can be traced back to a significant increase of carbon packing in the liquid as carbon atom coordination changes from less than 4 to higher values [see Fig. 2(d)]. SNAP accurately predicts this subtle change in density upon an increase of pressure as well as a corresponding pressure-dependent evolution of the average coordination of carbon atoms [Fig. 2(d)], which is in an excellent agreement with QMD. The latter is a result of a "snap-on" agreement between QMD and SNAP radial distribution functions for both diamond and BC8 over a large range of pressures (Fig. S4 [58]).

Another validation test of great experimental importance is the prediction of the carbon shock Hugoniot, which passes through both solid and liquid parts of the phase diagram [Fig. 2(a)]. The Hugoniot points are calculated in a hydrodynamic approximation by ignoring crystalline anisotropy in a series of MD simulations at a given pressure P but varying temperature T to satisfy the Hugoniot condition of conservation of mass, momentum, and energy: $\frac{1}{2}(P+P_0)(V-V_0)=(E-E_0)$, where P, V, and E are the pressure, volume, and internal energy at a given point on the Hugoniot, and P_0, V_0, and E_0 are those at ambient conditions of 0 GPa and 300 K. The Hugoniot in (P, T) space is shown in Fig. 2(a), in pressure-density (P-ρ) space in Fig. 3(b), and particle velocity (U_p)–shock velocity (U_s) space in Fig. 3(a). The SNAP Hugoniots [red lines in Figs. 2(a) and 3] are in a very good agreement with those from QMD [black lines in Figs 2(a) and 3]. Visible differences in temperature of the P-T Hugoniot at very high temperatures [Fig. 2(a)] are due to electronic entropy effects, which are not captured by the *temperature-independent* description of interatomic interactions [63].

Both SNAP and QMD Hugoniots are also in a very good agreement with multiple experiments in a pressure range from 300 to 1500 GPa [18,20–22,24,26,27]. Some difference between SNAP/QMD and experiment at higher pressures is due to experimental uncertainty in density, which is not measured directly, but rather determined using an impedance matching method [26,27]. Differences at low pressures between experiment [18,21,22] and SNAP/QMD [Fig. 3(a)] are due to crystalline anisotropy and strength effects, which are not well described by a hydrodynamic approximation [21]. To make a proper comparison with experiment in this split elastic-inelastic shock wave regime, explicit large-scale SNAP MD simulations of piston-driven shock waves are required, which will be the focus of future work.

To demonstrate SNAP's ability to attack problems that are impossible to solve with QMD, we simulated melting of polycrystalline diamond at pressures between 50 and 1200 GPa using a 1-million-atom sample (Fig. 4). For each pressure point, a series of NPT simulations is performed to determine the temperature at the onset of melting. In addition to validating SNAP, this simulation also validates the two-phase melting curve calculation. The presence of grain boundaries suppresses superheating characteristic of single crystals: The melting starts at the most weakly bonded defective regions of the sample, followed by the growth of liquid fraction at the expense of crystalline grains, which gradually transform to shrinking round crystallites embedded in the liquid carbon (Fig. 4).

This Letter represents a major step towards solving the extremely challenging but fundamentally important problem of predictive atomic-scale simulations of carbon at extreme pressure-temperature conditions at experimental timescales and length scales. The quantum-accurate SNAP is a MLIP that describes the properties of carbon at extreme pressures up to 5 TPa and temperatures up to 20 000 K, including the phase diagram, melting curves of diamond, BC8, and simple cubic phases of carbon and shock Hugoniots with unprecedented accuracy within 3% of QMD results. Although the accurate description of carbon at low pressures below 50 GPa is outside of the scope of this Letter, the ambient pressure properties of diamond are well described by SNAP, lattice constant, energy, and elastic constants being in excellent agreement (within 1.5%) with those predicted by DFT (Table S2). The ground state properties of graphene calculated by SNAP were less accurate [58]. The development of accurate SNAP for ambient carbon is a subject of our future work. SNAP's linear scaling with number of atoms, and its efficient implementation within the LAMMPS MD simulation package [37] allows billion-atom simulations on leadership-class high-performance computing systems. By advancing the frontiers of classical MD simulations, SNAP provides critical insights by uncovering the fundamental atomic-scale mechanisms of the materials' response which are difficult or even impossible to observe in experiment [64].

The work at USF is supported by DOE/NNSA (Grants No. DE-NA-0003910 and No. DE-NA-0004089) and DOE/FES (Grant No. DE-SC0023508). The work at KTH is supported by Swedish Scientific Council (VR) (Grant No. 2017-03744). Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy' s (DOE) National Nuclear Security Administration under Contract No. DE-NA0003525. The computations were performed using leadership-class HPC systems: OLCF Summit at Oak Ridge National Laboratory (ALCC and INCITE Awards No. MAT198) and TACC Frontera at University of Texas at Austin (LRAC Award No. DMR21006).

---

[1] M. Ross, Nature (London) **292**, 435 (1981).

[2] F. Ancilotto, G. L. Chiarotti, S. Scandolo, and E. Tosatti, Science **275**, 1288 (1997).

[3] L. R. Benedetti, J. H. Nguyen, W. A. Caldwell, H. Liu, M. Kruger, and R. Jeanloz, Science **286**, 100 (1999).

[4] D. Kraus, J. Vorberger, A. Pak, N. J. Hartley, L. B. Fletcher, S. Frydrych, E. Galtier, E. J. Gamboa, D. O. Gericke, S. H. Glenzer, E. Granados, M. J. MacDonald, A. J. MacKinnon, E. E. McBride, I. Nam, P. Neumayer, M. Roth, A. M. Saunders, A. K. Schuster, P. Sun, T. van Driel, T. Döppner, and R. W. Falcone, Nat. Astron. **1**, 606 (2017).

[5] N. Madhusudhan, K. K. M. Lee, and O. Mousis, Astrophys. J. **759**, L40 (2012).

[6] J. S. Ross, D. Ho, J. Milovich, T. Döppner, J. McNaney, A. G. Macphee, A. Hamza, J. Biener, H. F. Robey, E. L. Dewald, R. Tommasini, L. Divol, S. Le Pape, L. B. Hopkins, P. M. Celliers, O. Landen, N. B. Meezan, and A. J. Mackinnon, Phys. Rev. E **91**, 021101 (2015).

[7] M. Millot, P. M. Celliers, P. A. Sterne, L. X. Benedict, A. A. Correa, S. Hamel, S. J. Ali, K. L. Baker, L. F. Berzak Hopkins, J. Biener *et al.*, Phys. Rev. B **97**, 144108 (2018).

[8] L. B. Hopkins, S. LePape, L. Divol, A. Pak, E. Dewald *et al.*, Plasma Phys. Controlled Fusion **61**, 014023 (2019).

[9] A. L. Kritcher, A. B. Zylstra, D. A. Callahan, O. A. Hurricane *et al.*, Phys. Plasmas **28**, 072706 (2021).

[10] D. T. Casey, B. J. MacGowan, J. D. Sater, A. B. Zylstra, O. L. Landen, J. Milovich, O. A. Hurricane, A. L. Kritcher, M. Hohenberger, K. Baker *et al.*, Phys. Rev. Lett. **126**, 025002 (2021).

[11] J. Tollefson, Nature (London) **597**, 163 (2021).

[12] T. S. Duffy and R. F. Smith, Front. Earth Sci. **7**, 23 (2019).

[13] D. B. Sinars, M. A. Sweeney, C. S. Alexander, D. J. Ampleford, T. Ao *et al.*, Phys. Plasmas **27**, 070501 (2020).

[14] M. I. McMahon, in *Synchrotron Light Sources Free. Lasers*, edited by E. J. Jaeschke, S. Khan, J. R. Schneider, and J. B. Hastings (Springer International Publishing, Cham, 2020), pp. 1857–1896.

[15] E. A. Seddon, J. A. Clarke, D. J. Dunning, C. Masciovecchio, C. J. Milne, F. Parmigiani, D. Rugg, J. C. H. Spence, N. R. Thompson, K. Ueda, S. M. Vinko, J. S. Wark, and W. Wurth, Rep. Prog. Phys. **80**, 115901 (2017).

[16] M. N. Pavlovskii, Sov. Phys. Solid State **13**, 741 (1971).

[17] K.-I. Kondo and T. J. Ahrens, Geophys. Res. Lett. **10**, 281 (1983).

[18] R. S. McWilliams, J. H. Eggert, D. G. Hicks, D. K. Bradley, P. M. Celliers, D. K. Spaulding, T. R. Boehly, G. W. Collins, and R. Jeanloz, Phys. Rev. B **81**, 014111 (2010).

[19] J. M. Lang, J. M. Winey, and Y. M. Gupta, Phys. Rev. B **97**, 104106 (2018).

[20] K. Katagiri, N. Ozaki, Y. Umeda, T. Irifune, N. Kamimura, K. Miyanishi, T. Sano, T. Sekine, and R. Kodama, Phys. Rev. Lett. **125**, 185701 (2020).

[21] J. M. Winey, M. D. Knudson, and Y. M. Gupta, Phys. Rev. B **101**, 184105 (2020).

[22] M. J. MacDonald, E. E. McBride, E. Galtier, M. Gauthier, E. Granados, D. Kraus, A. Krygier, A. L. Levitan, A. J. MacKinnon, I. Nam, W. Schumaker, P. Sun, T. B. van Driel, J. Vorberger, Z. Xing, R. P. Drake, S. H. Glenzer, and L. B. Fletcher, Appl. Phys. Lett. **116**, 234104 (2020).

[23] S. Brygoo, E. Henry, P. Loubeyre, J. Eggert, M. Koenig, B. Loupias, A. Benuzzi-Mounaix, and M. Rabec Le Gloahec, Nat. Mater. **6**, 274 (2007).

[24] M. D. Knudson, M. P. Desjarlais, and D. H. Dolan, Science **322**, 1822 (2008).

[25] J. H. Eggert, D. G. Hicks, P. M. Celliers, D. K. Bradley, R. S. McWilliams, R. Jeanloz, J. E. Miller, T. R. Boehly, and G. W. Collins, Nat. Phys. **6**, 40 (2010).

[26] M. C. Gregor, D. E. Fratanduono, C. A. McCoy, D. N. Polsin, A. Sorce, J. R. Rygg, G. W. Collins, T. Braun, P. M. Celliers,

J. H. Eggert, D. D. Meyerhofer, and T. R. Boehly, Phys. Rev. B **95**, 144114 (2017).

[27] M. Millot, P. A. Sterne, J. H. Eggert, S. Hamel, M. C. Marshall, and P. M. Celliers, Phys. Plasmas **27**, 102711 (2020).

[28] A. Lazicki, D. McGonegle, J. R. Rygg, D. G. Braun, D. C. Swift *et al.*, Nature (London) **589**, 532 (2021).

[29] S. Scandolo, G. L. Chiarotti, and E. Tosatti, Phys. Rev. B **53**, 5051 (1996).

[30] M. P. Grumbach and R. M. Martin, Phys. Rev. B **54**, 15730 (1996).

[31] X. Wang, S. Scandolo, and R. Car, Phys. Rev. Lett. **95**, 185701 (2005).

[32] A. A. Correa, S. A. Bonev, and G. Galli, Proc. Natl. Acad. Sci. USA **103**, 1204 (2006).

[33] A. A. Correa, L. X. Benedict, D. A. Young, E. Schwegler, and S. A. Bonev, Phys. Rev. B **78**, 024101 (2008).

[34] J. Sun, D. D. Klug, and R. Martoňák, J. Chem. Phys. **130**, 194512 (2009).

[35] M. Martinez-Canales, C. J. Pickard, and R. J. Needs, Phys. Rev. Lett. **108**, 045704 (2012).

[36] L. X. Benedict, K. P. Driver, S. Hamel, B. Militzer, T. Qi, A. A. Correa, A. Saul, and E. Schwegler, Phys. Rev. B **89**, 224109 (2014).

[37] A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in 't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. J. Stevens, J. Tranchida, C. Trott, and S. J. Plimpton, Comput. Phys. Commun. **271**, 108171 (2022).

[38] D. W. Brenner, O. A. Shenderova, J. A. Harrison, S. J. Stuart, B. Ni, and S. B. Sinnott, J. Phys.: Condens. Matter **14**, 783 (2002).

[39] S. J. Stuart, A. B. Tutein, and J. A. Harrison, J. Chem. Phys. **112**, 6472 (2000).

[40] J. H. Los and A. Fasolino, Phys. Rev. B **68**, 024107 (2003).

[41] L. Pastewka, P. Pou, R. Pérez, P. Gumbsch, and M. Moseler, Phys. Rev. B **78**, 161402(R) (2008).

[42] S. G. Srinivasan, A. C. T. van Duin, and P. Ganesh, J. Phys. Chem. A **119**, 571 (2015).

[43] I. I. Oleynik, A. C. Landerville, S. V. Zybin, M. L. Elert, and C. T. White, Phys. Rev. B **78**, 180101(R) (2008).

[44] R. Perriot, X. Gu, Y. Lin, V. V. Zhakhovsky, and I. I. Oleynik, Phys. Rev. B **88**, 064101 (2013).

[45] J. Behler and M. Parrinello, Phys. Rev. Lett. **98**, 146401 (2007).

[46] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, Phys. Rev. Lett. **104**, 136403 (2010).

[47] P. Friederich, F. Häse, J. Proppe, and A. Aspuru-Guzik, Nat. Mater. **20**, 750 (2021).

[48] A. P. Thompson, L. P. Swiler, C. R. Trott, S. M. Foiles, and G. J. Tucker, J. Comput. Phys. **285**, 316 (2015).

[49] M. A. Wood and A. P. Thompson, J. Chem. Phys. **148**, 241721 (2018).

[50] A. V. Shapeev, Multiscale Model. Simul. **14**, 1153 (2016).

[51] R. Drautz, Phys. Rev. B **99**, 014104 (2019).

[52] L. Zhang, J. Han, H. Wang, R. Car, and E. Weinan, Phys. Rev. Lett. **120**, 143001 (2018).

[53] R. Z. Khaliullin, H. Eshet, T. D. Kühne, J. Behler, and M. Parrinello, Phys. Rev. B **81**, 100103 (2010).

[54] P. Rowe, V. L. Deringer, P. Gasparotto, G. Csányi, and A. Michaelides, J. Chem. Phys. **153**, 034702 (2020).

[55] V. L. Deringer and G. Csányi, Phys. Rev. B **95**, 094203 (2017).

[56] M. Qamar, M. Mrovec, Y. Lysogorskiy, A. Bochkarev, and R. Drautz, arXiv:2210.09161; B. Cheng, S. Hamel, and M. Bethkenhagen, arXiv:2207.02927.

[57] Y. Zuo, C. Chen, X. Li, Z. Deng, Y. Chen, J. Behler, G. Csányi, A. V. Shapeev, A. P. Thompson, M. A. Wood, and S. P. Ong, J. Phys. Chem. A **124**, 731 (2020).

[58] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevB.106.L180101 for details on the (1) SNAP training database; (2) SNAP machine learning development workflow; (3) accuracy assessment of SNAP training including the detailed comparison of SNAP and DFT predictions of diamond and graphene properties at ambient conditions; (4) SNAP validation: radial distribution functions of solid and liquid phases along diamond and BC8; and (5) computational details of polycrystalline melting simulations.

[59] A. B. Belonoshko, Geochim. Cosmochim. Acta **58**, 4039 (1994).

[60] A. S. Williams, K. Nguyen-Cong, J. T. Willman, and I. I. Oleynik, arXiv:2211.09974.

[61] FitSNAP: Software for generating SNAP machine learning interatomic potentials: http://github.com/FitSNAP.

[62] K. Dalbey, M. S. Eldred, G. Geraci, J. D. Jakeman, K. A. Maupin, J. A. Monschke, D. T. Seidl, L. P. Swiler, A. Tran, F. Menhorn, and X. Zeng, *Dakota A Multilevel Parallel Object-Oriented Framework for Design Optimization Parameter Estimation Uncertainty Quantification and Sensitivity Analysis: Version 6.12 Theory Manual* (United States, 2020), http://dx.doi.org/10.2172/1630693.

[63] By construction, SNAP is a temperature-independent potential. Therefore it is trained on zero-electronic-temperature DFT energies, stresses, and atomic forces even for atomic configurations taken from QMD simulations performed at finite ionic (T_i) and electronic (T_e) temperatures ($T_e = T_i$) in the Born-Oppenheimer approximation. SNAP is very accurate at temperatures below 10 000 K due to the vanishing contribution of electron entropy to DFT free energy under such conditions.

[64] K. Nguyen-Cong, J. T. Willman, S. G. Moore, A. B. Belonoshko, R. Gayatri, E. Weinberg, M. A. Wood, A. P. Thompson, and I. I. Oleynik, in *SC '21: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis* (ACM Press, New York, 2021), https://doi.org/10.1145/3458817.3487400.