ChIMES Carbon 2.0: A Transferable Machine-Learned Interatomic Model Harnessing Multifidelity Training Data

Rebecca K. Lindsey,[*,†,‡] Sorin Bastea,[¶] Sebastien Hamel,[¶] Yanjun Lyu,[‡] Nir Goldman,[¶,§] and Vincenzo Lordi[¶]

†*Department of Chemical Engineering, University of Michigan, Ann Arbor, Michigan 48109, United States*

‡*Department of Materials Science and Engineering, University of Michigan, Ann Arbor, Michigan 48109, United States*

¶*Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, California 94550, United States*

§*Department of Chemical Engineering, University of California, Davis, California 95616, United States*

E-mail: rklinds@umich.edu

Abstract

We present a new parameterization of the ChIMES physics informed machine-learned interatomic model for simulating carbon under conditions ranging from 300 K and 0 GPa to 10,000 K and 100 GPa, along with a new multi-fidelity active learning strategy. The resulting model shows significant improvement in accuracy and temperature/pressure transferability relative to the original ChIMES carbon model developed in 2017, and can serve as a foundation for future transfer-learned ChIMES parameter

sets. Model applications to carbon melting point prediction, shockwave-driven conversion of graphite to diamond, and thermal conversion of nanodiamond to graphitic nanoonion are provided. Ultimately, we find our new model to be robust, accurate, and well-suited for modeling evolution in carbon systems under extreme conditions.

1 Introduction

The nature of carbon under extreme conditions (i.e., $>$ 1000 K and 10 GPa) remains a topic of significant scientific interest due to relevance across a broad range of research domains including planetary, fusion, materials manufacturing, and detonation sciences.[1–5] At the same time, elucidating response and transformation in carbon remains a significant experimental challenge arising from limitations in drivers capable of imposing such conditions, and probes capable of characterizing ensuing evolution in low-Z materials like carbon.[1,6,7] In particular, static or quasi-static compression methods (e.g., using a diamond anvil cell) can generate conditions up to 100s of GPa and 1000s of K over long timescales and thereby allow precise characterization, but use of diamond as the pressurizing material when measuring properties of a carbon sample leads to difficulties, particularly near the material melt line.[8] While dynamic compression (e.g., shock) methods circumvent this challenge, the extreme conditions they generate are very short lived (e.g., ≈ 1 μs), significantly complicating characterization.[9] The complimentary capabilities afforded by simulation have led it to become a mainstay in extreme conditions research, e.g., providing an atomically-resolved picture of system structure, and enabling assignment of properties particularly challenging to determine in analogous experiments.

First principles (FP) approaches have long been the preferred method for computational studies under these conditions since they preempt the need to identify suitable approximate (e.g., classical mechanics) interatomic descriptions and because timescales for material evolution under such high T and P are quite amenable to those accessible through FP simulation.[10] This approach has been used to probe shock-induced transition from graphite to

diamond,[11] investigate the possibility of a liquid liquid carbon phase transition,[12] characterize amorphous carbon properties and kinetics,[13,14] and predict phase boundaries under TPa pressures.[15] However the associated computational expense precludes probing phenomena with characteristic spatial scales exceeding more than a few hundred atoms. Early efforts to overcome these limitations manifested in computationally efficient "bond-order" approaches including LCBOP[16,17] and REBO,[18] which have since been deployed in comprehensive studies of thermodynamic properties of carbon at extreme conditions.[17,19,20] However, the rigid underlying functional form of these models hinders accurate description of the disordered molten carbon phase.

Recently, machine-learning (ML) -based interatomic potentials (IAP) have emerged as a powerful tool for bridging the FP/classical interatomic model accuracy/efficiency gap, thereby providing an opportunity to revisit the problem of carbon under extreme conditions. In contrast to molecular mechanics methods, which require assumptions of underlying physics by way of the expressions used to describe interatomic interactions, ML-IAPs employ a highly flexible underlying form (e.g., neural network, Gaussian process, among other approaches). Provided appropriate volumes of relevant FP training data, these models are able to "learn" topography of the FP method's underlying potential energy surface, significantly minimizing foreknowledge necessary to model a given system. In the space of carbon under extreme conditions, particularly notable examples of these models include GAP,[21,22] which was the first among the high-accuracy general purpose ML-IAPs for carbon, quadratic SNAP,[23,24] which has since been used to conduct billion atom simulations predicting shock synthesis of BC8 carbon[25] (i.e., under TPa conditions), and ACE,[26] which has been shown to be both efficient and effective for modeling mechanical properties of carbon. Nevertheless, there remains a need for ML-IAPs capable of modeling shockwave-driven phenomena, i.e., applicable from nominally ambient conditions to intermediately extreme conditions of 1000s of K and 10s to 100s of GPa – regimes relevant to planetary science, shock- and detonation-induced chemistry, and advanced materials manufacturing.

In 2017, we developed the ChIMES ML-IAP to enable accurate modeling of condensed-phase chemistry under extreme conditions[27] and published its first application to molten carbon.[28] Though the model yielded significant improvement upon classical bond-order methods, the domain for model applicability was confined to 5000 – 6000 K and 2.25 – 3.00 g cm^{-3}. Since then, the ChIMES model and underlying framework has been significantly enhanced[27,29–31] to improve model accuracy, robustness, efficiency and automation/reproducibility in training, bringing it in line with the capabilities and efficiency of GAP, SNAP, and ACE. At the same time, the unique underlying form of ChIMES, discussed in greater detail below, provides an inherent capability for explicit chemical extensibility without the need for retraining – that is, a sufficiently accurate ChIMES-C model can, in principle, serve as *fixed* foundation upon which multi-atom-type models containing carbon can be built, which could greatly simplify future model development efforts.

Therefore, in this manuscript we describe a new ChIMES carbon model intended for use primarily in the range 1000 K and 10 GPa to 10,000 K and 100 GPa that is also accurate under nominally ambient conditions (i.e., 300 K, 0 GPa). In depth discussion of our fitting procedures and improvements relative to our early model are provided, and a new strategy for multifidelity learning is presented. The remaining manuscript is structured as follows: the ChIMES model is presented in detail, followed by discussion of training data generation and fitting protocols. Model training and validation results are then presented, followed by selected "proof-of-concept" applications that include predicting carbon melting points as well as modeling shockwave-driven synthesis of diamond from a graphitic precursor and carbon nanoonion synthesis from nanodiamond. Where possible, methodologies and results are compared against our original ChIMES-C model. We conclude with discussion of best practices, and implications for future research. We note that parameters for the presently developed model are available in our public repository.[32]

2 Methods

2.1 The ChIMES Interatomic Model

ChIMES is a unique, physics informed ML-IAP that describes system energy through a generalized explicit many-body cluster expansion, i.e.:

$$E_{n_\mathrm{B}} = \sum_{i_1}^{n_\mathrm{a}} {}^{1}E_{i_1} + \sum_{i_1>i_2}^{n_\mathrm{a}} {}^{2}E_{i_1 i_2} + \sum_{i_1>i_2\ldots\ i_{n_\mathrm{B}-1}>i_{n_\mathrm{B}}}^{n_\mathrm{a}} {}^{n_\mathrm{B}}E_{i_1 i_2 \ldots i_{n_\mathrm{B}}}, \tag{1}$$

where E_{n_B} is the total ChIMES system energy, n_B is the maximum bodiedness, ${}^{n}E_{i_1 i_2\ldots\ i_n}$ is the n-body ChIMES energy for a given set of n atoms with indices $\boldsymbol{i} = \{i_1, i_2, \ \ldots\ , i_n\}$, and n_a is the total number of atoms in the system.

In contrast to the many other ML-IAPs that employ a cluster-expansion formalism for system energy based on *atom*-centered clusters, ChIMES uses a *cluster*-centered descriptor - that is, ChIMES views n-body clusters as fully-connected graphs of constituent atoms rather than as a graph for which edges are only connected to a central node. This subtle but important distinction makes ChIMES particularly well-suited for modeling molecular, covalently bonded, and/or reacting systems,[30,31,33–38] and gives ChIMES models the unique property of being explicitly chemically extensible, since energy of a given cluster depends only on the types of atoms represented within that specific cluster rather than on the types of all possible atoms described by the model.

In the ChIMES framework, single-body energies are constant values and n-body energies are constructed from the product of Chebyshev polynomials for the pairs that comprise that cluster. Thus, a 2-body interaction would involve a single pair, ij, while a three-body interaction would involve 3 pairs, ij, ik, jk, a 4-body interaction would involve $\binom{4}{2}$ pairs, and so on. For a 2-body interaction, we have:

$${}^{2}E_{ij} = f_\mathrm{p}^{e_i e_j}\left(r_{ij}\right) + f_\mathrm{s}^{e_i e_j}\left(r_{ij}\right) \sum_{\alpha=1}^{\mathcal{O}_\mathrm{2B}} c_\alpha^{e_i e_j} T_\alpha\left(s_{ij}^{e_i e_j}\right), \tag{2}$$

where r_{ij} is the distance between a pair of atoms ij of respective element type e_i and e_j, s_{ij} is the transformed pair distance that lies within the $[-1, 1]$ domain over which Chebyshev

polynomials are defined, and $\mathcal{O}_{\mathrm{2B}}$ is the 2-body polynomial order. The functions f_{p} and f_{s} are penalty and smooth cutoff functions defined below.

The transformation from r_{ij} to s_{ij} can generally be defined as:

$$s_{ij}^{e_ie_j} = (x_{ij}^{e_ie_j} - x_{\mathrm{avg}}^{e_ie_j})/x_{\mathrm{diff}}^{e_ie_j} \tag{3}$$

where

$$x_{\mathrm{avg}}^{e_ie_j} = 0.5(x_{\mathrm{c,out}}^{e_ie_j} + x_{\mathrm{c,in}}^{e_ie_j}), \tag{4}$$

$$x_{\mathrm{diff}}^{e_ie_j} = 0.5|x_{\mathrm{c,out}}^{e_ie_j} - x_{\mathrm{c,in}}^{e_ie_j}|. \tag{5}$$

Here, we define $x_{ij}^{e_ie_j} = \exp(-r_{ij}/\lambda^{e_ie_j})$. This choice of Morse-like transformation function encodes that interatomic interaction strength generally decreases with increasing interatomic distance, which implies that the potential energy surface topography is more featured at short range. In essence, this transformation stretches the short-ranged r portion of the $[r_{\mathrm{cut,in}}, r_{\mathrm{cut,out}}]$ domain while compressing the longer-ranged r portion, sensitizing our Chebyshev basis in the former region. Practically, this affords higher accuracy models with lower overall polynomial order.[39]

For higher bodied interactions, cluster energies are taken as the product of interactions for constituent atom pairs, where for a 3-body interaction, we have:

$${}^3E_{ijk} = f_{\mathrm{s}}^{e_ie_j}\left(r_{ij}\right) f_{\mathrm{s}}^{e_ie_k}\left(r_{ik}\right) f_{\mathrm{s}}^{e_je_k}\left(r_{jk}\right) \sum_{\alpha=0}^{\mathcal{O}_{\mathrm{3B}}} \sum_{\substack{\beta=0 \\ *}} \sum_{\gamma=0} c_{\alpha,\beta,\gamma}^{\boldsymbol{E}} T_\alpha\left(s_{ij}^{e_ie_j}\right) T_\beta\left(s_{ik}^{e_ie_k}\right) T_\gamma\left(s_{jk}^{e_je_k}\right), \tag{6}$$

where $\boldsymbol{E}$ gives the atom pair element types, $\{e_ie_j, e_ie_k, e_je_k\}$. As shown elsewhere,[29,31] higher bodied interactions follow a similar convention, with a larger number of constituent pairs (e.g., 6 pairs for a 4-body) interaction.

In our 2017 ChIMES-C model, we used a cubic smooth cutoff function:

$$f_{\mathrm{s}}^{e_ie_j}(r_{ij}) = \left(1 - \frac{r_{ij}}{r_{\mathrm{c,out}}^{e_ie_j}}\right)^3 \tag{7}$$

However, when applied to models with $>$ 3-body interactions, this form was found to modify the polynomial series too aggressively (e.g., for a 4-body interaction, this form would result in the sum over Chebyshev polynomials being multiplied by the product of 6 numbers < 1).

Hence, we have since transitioned to use of a "Tersoff" style[40] smooth cutoff function that leaves a significant portion of the interaction unmodified, i.e.:

$$f_{\mathrm{s}}^{e_i e_j}(r_{ij}) = \begin{cases} 0, & \text{if } r_{ij} > r_{\mathrm{c,out}}^{e_i e_j} \\ 1, & \text{if } r_{ij} < d_{\mathrm{t}} \\ \frac{1}{2} + \frac{1}{2}\sin\left(\pi\left[\frac{r_{ij}-d_{\mathrm{t}}}{r_{\mathrm{c,out}}^{e_i e_j}-d_{\mathrm{t}}}\right] + \frac{\pi}{2}\right), & \text{otherwise} \end{cases} \tag{8}$$

where d_t is a hyperparameter controlling point at which this function activates. Finally, the penalty function is given by:

$$f_{\mathrm{p}}^{e_i e_j}\left(r_{ij}\right) = \begin{cases} A_{\mathrm{p}}^{e_i e_j}\left(r_{\mathrm{c,in}}^{e_i e_j} + d_{\mathrm{p}}^{e_i e_j} - r_{ij}\right)^3, & \text{if } r_{ij} < r_{\mathrm{c,in}}^{e_i e_j} + d_{\mathrm{p}} \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

where $d_{\mathrm{p}}^{e_i e_j}$ and $A_{\mathrm{p}}^{e_i e_j}$ are hyperparameters controlling the point at which the penalty function activates, and the penalty function strength, taken as 0.02 Å and 1×10^6 kcal mol^{-1}, respectively.

2.2 ChIMES Model Fitting

There are several interrelated aspects to fitting ML-IAPs. A target application space must be clearly defined, a "ground truth" data labeling method must be specified, a training set and target properties within that set must be generated, "conventional" hyperparameters (i.e., controlling model architecture) must be specified, other hyperparameters (e.g., weights to be applied during fitting) must be identified, and training protocols (e.g., single-step, iterative, or active learning) must be selected. An overview of these choices for our 2017 and present ChIMES-C model (henceforth referred to as the 2024 model) is provided in Table 1.

Both our 2017 and present ChIMES models treat spin-restricted Kohn-Sham density functional theory (DFT) as the ground truth. In both cases, initial training sets were generated using VASP[41–44] with the Perdew–Burke–Ernzerhof generalized gradient approximation functional,[45,46] projector-augmented wave pseudopotentials[47,48] (PAW), and the DFT-D2 method[49] for description of dispersion interactions. This method was selected because it has

Table 1: ChIMES hyperparameters used for generation of the 2017 and present (2024) ChIMES-C model. Weight units are the inverse of the property they are multiplied by, i.e., Å mol kcal^{-1}, mol kcal^{-1}, and Å^3 mol kcal^{-1} for w_{F}, w_{E}, and w_σ, respectively; all cutoffs and λ are given in Å.

hyperparameter	2017 Model	2024 Model
$\mathcal{O}_{2b}$	12	20
$\mathcal{O}_{3b}$	5	10
$\mathcal{O}_{4b}$	0	4
$r_{c,in}$	1.0	0.98
$r_{c,out,2b}$	3.15	5.0
$r_{c,out,3b}$	3.15	5.0
$r_{c,out,4b}$	–	4.5
λ	1.25	1.40
Cutoff	cubic	Tersoff
d_t	–	0.75
w_{F}	1.0	1.0
w_{E}	–	100.0
w_σ	–	0.1
$n_{f,train}$	26	1116
n_{param}	42	442

been previously shown well-suited for modeling C, H, O, and N-containing materials under the target conditions and because resulting data will be consistent with our previously generated datasets, facilitating ongoing transfer learning efforts described in the introduction. However, future work will explore use of more modern dispersion-correction methods. All DFT calculations were finite temperature, where electronic eigenstates were occupied according to Fermi-Dirac distribution with the electronic temperature set equivalent to the target ionic temperature. The plane-wave-basis set energy cutoff was set to 1000 eV; henceforth, these protocols are simply referred to as PBE-D2.

The 2017 model was developed specifically to reproduce DFT-predicted carbon structure and dynamic properties at 5000 K and 12 GPa (i.e., in the triple point vicinity); the training set for the model comprised 26 evenly spaced frames from a 5 ps 256 carbon atom simulation at the target T and P. In contrast, the 2024 model's objective is to yield accurate prediction of structure, dynamics, and equation of state in diamond, graphite, and molten carbon over conditions ranging from 300 K and 0 GPa to 10,000 K and 100 GPa. Hence, the *initial* training set contains 25 evenly spaced frames over 5 ps simulations for *each* of these phases at the range of conditions detailed in Table 2. Additionally, 5 cold-compression configurations

Table 2: Overview of initial training set used for the 2024 model. Bolded lines indicate state points that were used for parallel active learning. "MD" indicates single-phase simulations at constant temperature and density, while "Cold Curve" indicates structure optimizations performed at 0 K, with external pressure applied.

State Point Index	n_F	Type	Phase	ρ (g cm^{-3})	P (GPa)	T (K)
0	25	MD	diamond	3.56	7.14	300
1	**25**	**MD**	**diamond**	**3.67**	**40.57**	**3000**
2	25	MD	graphite	2.39	0.47	300
3	25	MD	graphite	2.56	8.28	1500
4	**25**	**MD**	**graphite**	**2.67**	**17.74**	**3000**
5	**25**	**MD**	**liquid**	**0.50**	**-0.29**	**1000**
6	**25**	**MD**	**liquid**	**1.00**	**-0.84**	**2000**
7	**25**	**MD**	**liquid**	**2.00**	**6.17**	**6000**
8	25	MD	liquid	2.00	8.16	7000
9	25	MD	liquid	2.50	16.83	6000
10	**25**	**MD**	**liquid**	**3.00**	**55.73**	**8000**
11	25	MD	liquid	3.60	95.04	6000
12	5	Cold Curve	diamond	–	0 to 40	0
13	5	Cold Curve	graphite	–	0 to 40	0

were included for diamond and graphite, for P ranging from 0 to 40 GPa.

Conventional hyperparameters for the original ChIMES model were selected based on a comprehensive hyperparameter sensitivity study.[50] Since then, we have established physically-motivated guidelines for selecting these parameters.[39] For example, $r_{\mathrm{c,in}}^{e_i e_j}$ are taken to be slightly less than the minimum distance sampled in the DFT-MD simulations used to generate the initial training/validation sets (by typically 0.02 Å); d_p is then set to 0.02 Å, which ensures that the penalty function will never be sampled for an interatomic distance observed in a DFT-MD simulation. $\lambda^{e_i e_j}$ is set to the location of the first peak in the radial pair distribution function for the e_i,e_j pair and therefore can be thought of as a characteristic bonding distance. Outer cutoffs are generally selected such that $r_{\mathrm{c,out,2B}}^{e_i e_j} \geq r_{\mathrm{c,out,3B}}^{e_i e_j} \geq r_{\mathrm{c,out,4B}}^{e_i e_j} \geq \ldots$ and $r_{\mathrm{c,out,2B}}^{e_i e_j}$ is typically selected to be long enough to capture the longest 3$^{\mathrm{rd}}$ non-bonded solvation shell. We note that, in contrast to our 2017 model, outer cutoffs for the 2024 model (see Table 1) were selected to enable approximate recovery of the graphite inter-layer spacing, but due to a desire for computational efficiency enabling large-scale (e.g.., million atom) simulations under high pressure conditions, our selected cutoffs preclude fine recovery of the low-lying dispersion forces that modulate energetics associated with small changes in inter-

layer spacing. To overcome this limitation, future work will explore training models directly to PBE and overlaying a suitable dispersion correction during the ChIMES simulation.

Selection of bodiedness and corresponding polynomial order still requires some trial and error - for this work, we evaluated fitting RMSE based on our *initial* training set, varying polynomial orders and outer cutoffs to identify a combination affording relative computational efficiency and accuracy. Notably, this endeavor provides only a rough estimate of appropriate polynomial orders, since necessary complexity is directly connected with training set completeness. In particular, if a sparse initial training set is produced, hold-out cross validation may recommend a model that is insufficiently complex to recover key features in the ground truth PES. Conversely, if a higher complexity model is used, some form of iterative refinement or active learning will likely be needed to generate a robust model. However this new data may sample configurations that cannot be adequately described using model hyperparameters originally identified based on the initial sparse training set. Therefore, we find it best to initially err on the side of higher model complexity - once a final training set is identified, the model can always be refit at lower complexity as needed.

In general, ChIMES models can be fit by force-, stress-, and/or energy- matching, i.e. to an objective function of the form:

$$F_{\mathrm{obj}} = \frac{1}{n_{\mathrm{f}}\left(3n_{\mathrm{a}}+10\right)} \sum_{i=1}^{n_{\mathrm{f}}} \left[\sum_{j=1}^{n_{\mathrm{a}}} \sum_{k=1}^{3} w_{\mathrm{F}_{ijk}}^{2} \left(\Delta F_{ijk}\right)^{2} + w_{\mathrm{E}_{i}}^{2} \left(\Delta E_{i}\right)^{2} + \sum_{j=1}^{9} w_{\sigma_{ij}}^{2} \left(\Delta \sigma_{ij}\right)^{2} \right], \tag{10}$$

where $\Delta X = X^{\mathrm{DFT}} - X^{\mathrm{ChIMES}\{c\}}$. F_{obj} and $\{c\}$ are the weighted root-mean-squared error and model coefficients, respectively. The number of frames and atoms are given by n_{f} and n_{a}, respectively, and the factor of 10 in the denominator arises from inclusion of a single per-configuration energy and stress tensor, E_i and σ_{ijk}. F_{ijk} indicates the k^{th} Cartesian component of the force on atom j in configuration i while σ_{ij} indicates the j component of the stress tensor for configuration i. Note that units of kcal mol^{-1} Å^{-1}, and kcal mol^{-1}, and kcal mol^{-3} Å^{-1} were used for forces and energies, respectively. The superscripts "ChIMES" and "DFT" indicate forces/energies predicted from the present force-matched model and the DFT molecular dynamics (DFT-MD) training trajectory, respectively. Our 2017 model was

only fit to forces, without any explicit weighting, whereas the present model included forces, energies, and stresses, weighted by an initial factor of 1.0, 100.0, and 0.1, respectively to afford a balance between magnitude and quantity of each type of data point.

Since ChIMES is entirely linear in its fitted parameters, the model optimization problem can be recast as the following over-determined matrix equation:

$$\boldsymbol{w}\boldsymbol{M}\boldsymbol{c} = \boldsymbol{w}\boldsymbol{X}_{\mathrm{DFT}}, \tag{11}$$

where $\boldsymbol{X}_{\mathrm{DFT}}$ is the vector of F_{ijk}^{DFT}, E_i^{DFT}, and $\sigma_{ij}^{\mathrm{DFT}}$ values, $\boldsymbol{w}$ is a diagonal matrix of weights to be applied to the elements of $\boldsymbol{X}_{\mathrm{DFT}}$ and rows of $\boldsymbol{M}$, and the elements of design matrix $\boldsymbol{M}$ are given by:

$$M_{ab} = \frac{\partial X_{a,\mathrm{ChIMES}\{c\}}}{\partial c_b}. \tag{12}$$

In the above, a represents a combined index over force and energy components, and b is the index over permutationally invariant model parameters. Parameters for our 2017 model were generated using regularized singular value decomposition (SVD) through the SciPy package,[51] whereas the present model employed a locally developed distributed code[34] that implements the LARS algorithm[52,53] for LASSO[54] for large problems, capable of performing feature (parameter) selection (see Ref. 29 for implementation details). Practically, models generated via LASSO can have fewer parameters than implied by specified polynomial orders – 10% fewer for the 2024 model.

Our 2017 model was fit in a single step – that is, a fixed training set was generated, from which the final model was produced via steps 1 and 2 using the ChIMES-LSQ code[55] as shown in Figure 1. In contrast, for the present model, we have no *a priori* indication that our initial training set was sufficient to constrain our model; hence, the 2024 model was generated through parallel active learning using the basic ChIMES strategy. We note that this process was conducted autonomously through the ChIMES Active Learning Driver.[56] As shown in Fig 1, this entails generating an initial ChIMES model (steps 1 and 2), simulating with that model (step 3), deciding whether the model is suitable as is, or if it requires further

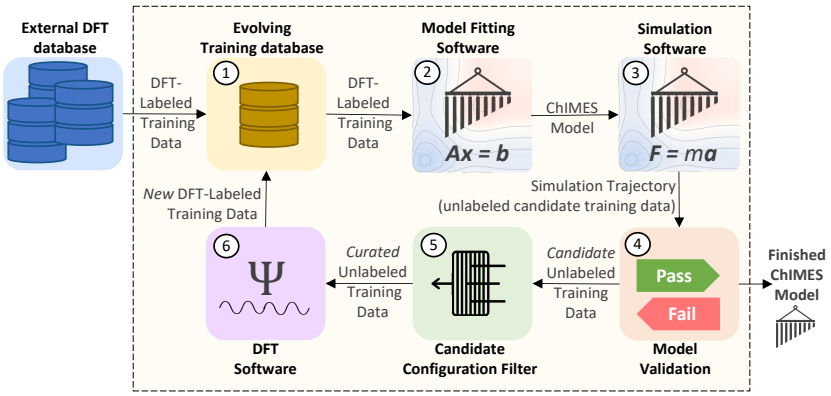

Figure 1: Workflow for generating a ChIMES model via active learning. This process is automated through the ChIMES Active Learning Driver.[29]

refinement (step 4). If additional refinement is needed, candidate configurations from the ChIMES simulations are selected to add to the training set (step 5), DFT forces, energies, and/or stresses for the selected configurations are assigned (step 6), and the process is repeated until user-specified performance criteria are met. We note that step 4 relies on user inspection since acceptable tolerances can vary significantly between different applications. In this case, our criteria were stable simulations (i.e., conserving the appropriate quantity) and yielding good agreement with DFT in predicted structure *via* the radial pair distribution function, equation of state, and dynamic properties including vibrational power spectra and diffusion coefficients.

For the present ChIMES-C model, parallel active learning was conducted using the basic ChIMES strategy, meaning at step 3, simulations were launched for several different state points (see bolded data in Table 2), from which candidate configurations were harvested for step 5. Configurations were selected using two criteria: (1) up to 20 evenly spaced configurations were taken from each state point's trajectory, for frames in which one or more $r_{\mathrm{c,in}}^{e_i e_j} < r_{ij} \leq r_{\mathrm{c,in}}^{e_i e_j} + d_p$ to inform close-contacts that are generally severely under-sampled in DFT-generated configurations and contribute most significantly to model stability issues, and (2) an additional set of up to 20 evenly spaced configurations were taken from each state point's trajectory, for frames in which all $r_{ij} > r_{\mathrm{c,in}}^{e_i e_j} + d_p$, which contain model-generated

predictions for timescales overlapping with DFT (i.e., that are useful for teaching the fitting framework whether the ChIMES model and DFT are in disagreement), and predictions for longer timescales than were accessed in the DFT training set. During each active learning cycle, force, stress, and energy weights were reduced by a factor equal to the current learning cycle number divided by the total number of learning cycles attempted. As discussed in ref. 31, this helps ensure that during later cycles (1) the initial FP-generated training configuration are not forgotten by the framework and (2) model predictions converge.

2.3 Multifidelity Learning Strategy Details

A portion of the initial training configurations were taken from previously run DFT-MD simulations. To improve the training set's coverage of physicochemical space in a computationally efficient manner, semi-empirical density-functional-based tight binding[57] (DFTB-) MD simulations were used to supplement the training data with additional temperature/pressure/phase points. All initial training configurations were then relabeled with DFTB to ensure consistent information is fed into the fitting framework, and models were generated through a 5-iteration run of the Active Learning Driver, using DFTB as the labeling method. The resulting training set was then relabeled once more via DFT with PBE-D2 and run for an additional iteration using PBE-D2 as the ground truth. DFTB is orders of magnitude faster than DFT even when run on a single processor, hence this strategy uses relatively minimal computational resources during Active Learning Driver runs, consequently reducing time spent queuing during the data labeling phase of each learning cycle and helping improve job synchronicity when using parallel active learning. Notably, the success of this approach also hints at viability of a multifidelity learning approach, wherein *all* simulations for training data generation (i.e., that entail tens to hundreds of thousands of calculations) can be performed at the DFTB-level, with DFT used only to relabel the orders-of-magnitude smaller subset of configurations that are actually used for model training. This could make ML-IAP model generation more tractable for users with relatively modest computing re-

sources. We note that all DFTB calculations were run with self-consistent charges using the DFTB+ code[58] and the Mio-1-1 parameter set[59] with dispersion described through the Universal Force Field[60] (UFF), rather than through a parameter set more optimized for describing carbon materials (e.g., pbc[61,62]) to evaluate robustness of this fitting strategy.

3 Results

3.1 Model validation

Parity plots for the 2024 model are provided in Figure 2. Our model is in excellent agreement with DFT, with root-mean-squared-errors (RMSE) for force, energy and stress of 15.5 kcal mol^{-1} Å^{-1}, 1.5 kcal mol^{-1} atom^{-1}, and 3.3 GPa, respectively (i.e., reduced RMSE, RMSE / $< |x| >$ where x is force, stress, or energy of 0.25, 0.01, and 0.41, respectively). This is a significant improvement upon the 2017 model, which exhibited (reduced) force RMSE of 26.4 (0.439), at only a single state point. A more direct comparison of the two models is given in Figure 3, which provides the radial pair distribution function (RDF) and vibrational power spectrum for both models relative to DFT at the state point for which the 2017 model was fit, 5000 K and 2.43 g cm^{-3}. While both models yield good predictions of these properties, the 2024 model is clearly in better agreement, most notably for the vibrational power spectrum, despite this state point not being included in its training set. Predicted pressure at this state point is also improved by the 2024 model, yielding a value of 10.5, compared to the 2017 and DFT values of 19.7 and 13.8 GPa, respectively. Finally, 0 K lattice parameters and corresponding energies per atom predicted by each model and DFT are given in Table 3, along with finite temperature ($\approx$ 300 K) experimentally derived values. Though neither the 2017 nor 2024 models were trained on or intended for use in describing BC8 carbon, we include it in the present validation to evaluate model transferability. Both models exhibit good recovery of diamond, graphite, and BC8 a parameters, but the 3.15 Å cutoff used by the 2017 model precluded ability to predict the graphite c parameter - in

contrast, we find that the 2024 model yields a reasonable prediction relative to DFT, despite use of a short cutoff relative to the long-range dispersion forces governing interlayer spacing, and that our value is in good agreement with experiment. We also find that, unlike the 2017 model, our 2024 parameterization provides relative energies that are in quantitative agreement with DFT.

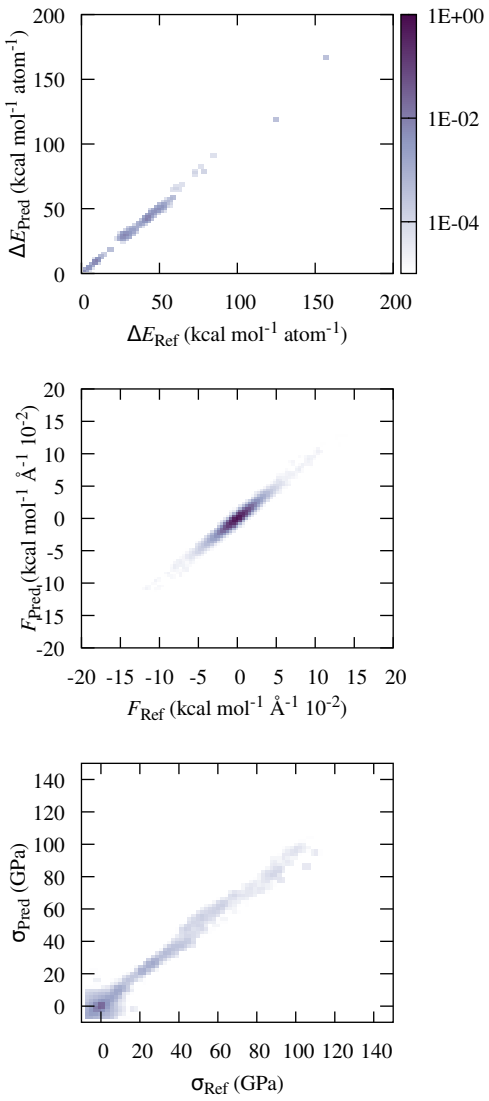

Figure 2: Parity plots predicted per-atom forces, full-system energies, and full system stress tensor components predicted by DFT (Ref) and the 2024 ChIMES model (Pred), for configurations at all T/P considered for fitting. The color bar gives point density.

Figure 4 provides a more comprehensive overview of model performance, i.e. for all state

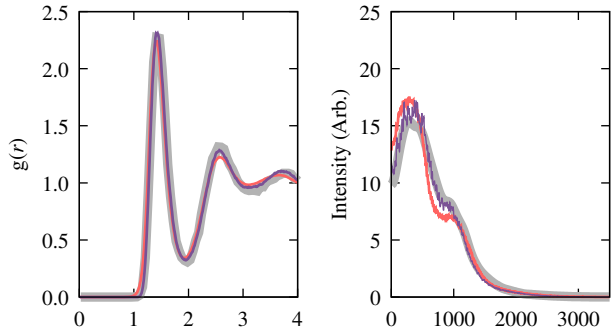

Figure 3: Comparison of the radial pair distribution function (left) and vibrational power spectrum (right) predicted by DFT (grey), and the 2017 and 2024 carbon models (orange and purple, respectively), at 5000 K and 2.43 g cm^{-3}.

Table 3: Energies relative to graphite and lattice parameters (a and c) of graphite, diamond, and BC8 predicted by experiment,[63,64] DFT, and ChIMES models. DFT energies were taken from ref. 65. Experimental measurements were at finite temperature ($\approx$ 300 K) whereas DFT and ChIMES values were computed at 0 K.

		a	c	eV/atom
graphite	Exp	2.462	6.707	
	DFT	2.47	6.391	0
	2017 ChIMES	2.50	–	0
	2024 ChIMES	2.464	6.744	0
diamond	Exp	3.57		
	DFT	3.565		–0.012 - 0.012
	2017 ChIMES	3.565		1.27
	2024 ChIMES	3.564		0.04
BC8	DFT[61]	4.50		0.72
	2017 ChIMES	4.40		1.81
	2024 ChIMES	4.39		0.77

points described in Table 2.2 for which temperature is > 300 K. Overall, we find excellent agreement with DFT, though deviations can be seen in the 1000 K and 2000 K state points. We note that these deviations are driven by the far-from-equilibrium nature of the DFT-MD simulations for these two cases. In particular, these two low-density state points entail highly reactive carbyne-like species that undergo significant changes over the course of the simulations. While useful for model training purposes, time evolved structure of this system over the short timescales used for DFT-MD simulations (and all comparisons made against them) is highly dependent on initialization (e.g., structure and/or velocity distribution).

Comparison of each pressure and fluid-phase self-diffusion coefficients are given in Figure 5. We find excellent agreement for predicted pressures. Diffusion coefficients are also in

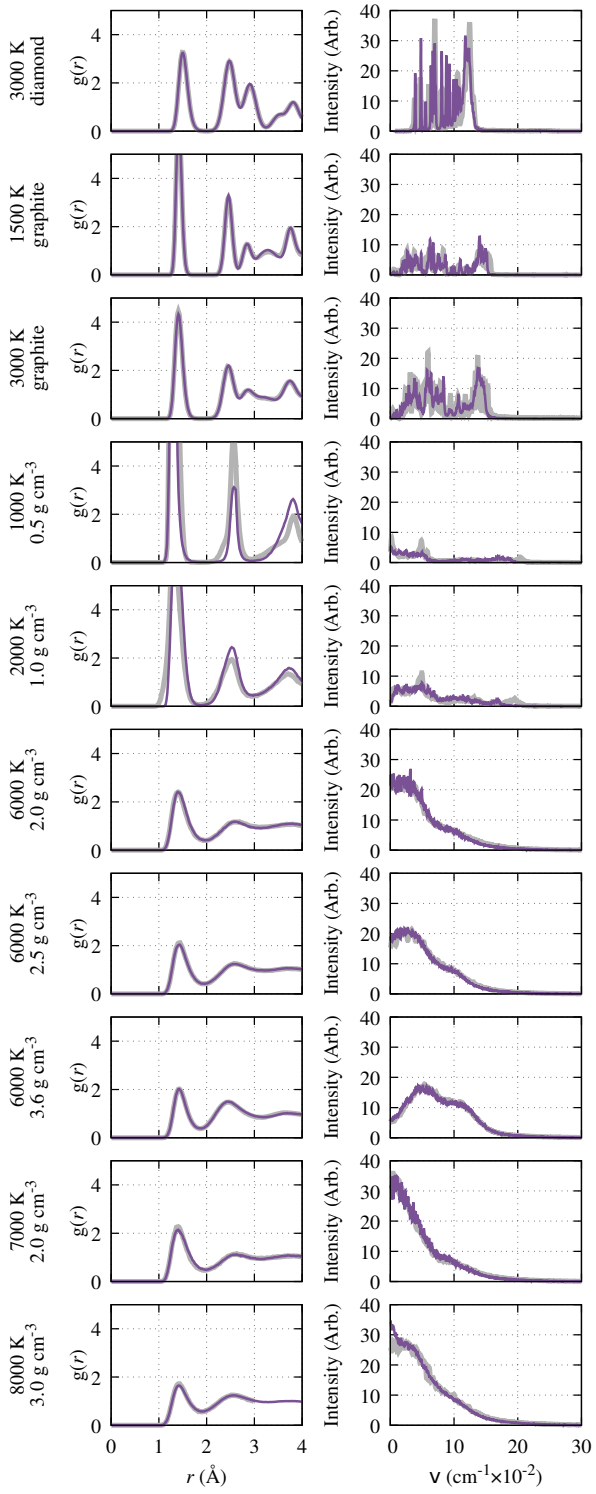

Figure 4: Radial pair distribution functions (left column) and corresponding vibrational power spectra (right column) predicted by DFT (grey) and the 2024 ChIMES model (purple) for diamond, graphite, and liquid carbon at various state points.

good agreement with the exception of a single outlier for 6000 K and 3.6 g cm^{-3}; however given the short timescales of these simulations (5 ps), and the small system sizes ($\approx$ 200 atoms), this discrepancy is likely due to finite size effects. Figure 5 also provides a comparison of cold compression curves for diamond and graphite relative to the present 2024 model, both of which are generally well recovered, with the notable exception of graphite at 5 GPa. We suspect our choice of short (4.5 to 5 Å) cutoff is the source for this discrepancy, but note that we still recover the expected monotonic decrease in cell volume with increasing pressure. Moreover, the model is intended for use at finite temperatures, where thermal energy is sufficient to rival the relatively weak dispersion interactions that modulate interlayer spacing at 0 K. This is seen, for example, in the 1500 K graphite data in Figures 4 and 5 where excellent agreement with DFT is observed. Performance of the 2024 model was also evaluated in terms of the the BC8 cold curve. The model was found to yield cell volumes within 5 % of those predicted by DFT[66] up to 30 GPa, which is remarkable given that no information on this phase of carbon was present within the training set. However, the model was found to exhibit anomalous decreasing pressure with decreasing cell volume above approximately 38 GPa.

3.2 Sample Applications

We provide three sample simulations to demonstrate efficacy of our model for problems not explicitly accounted for in the training set yet crucial for accurately modeling materials synthesis and transformation processes in carbon under extreme conditions: diamond melting point prediction, shock synthesis of diamond from a graphite precursor, and high temperature transformation of nanodiamond to a graphitic nanoonion structure. Notably, each of these applications entail some form of phase transformation.

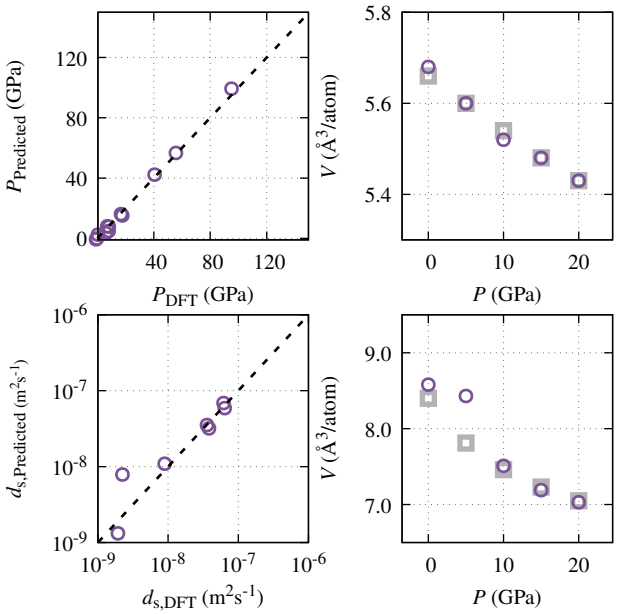

Figure 5: Comparison of pressure, self-diffusion coefficient, and cold compression curves predicted (diamond - top, graphite - bottom) by DFT and the 2024 model. In cold compression plots, DFT and ChIMES are given by grey squares and purple circles, respectively.

3.2.1 Diamond Melting Point Prediction

Experimentally, the two standard approaches for equation of state measurement under extreme conditions entail use of either a diamond anvil cell (DAC) (i.e., where diamonds are used as the pressure-coupling materials), or shock-based methods (i.e., using either a projectile or laser to drive a sample to high pressure). Temperature measurement in these experiments notoriously difficult, especially in carbon materials.[8,9,67] In particular, shock experiments are characterized by exceedingly short (i.e., sub μs) timescales that preclude reliable temperature measurement, whereas in longer timescale DAC experiments, prolonged exposure to high temperature can cause degradation of apparatus materials. DAC challenges are further exacerbated by the fact that the sample and pressurizing material are the same. Due to these difficulties, experimental values for the carbon melt line vary by up to 1500 K from experiment to experiment.[67,68,68] At the same time, finite size effects preclude reliable estimation from FP simulation at the present range of conditions of interest.

ML-IAPs like the presently developed 2024 model provide an exciting opportunity to

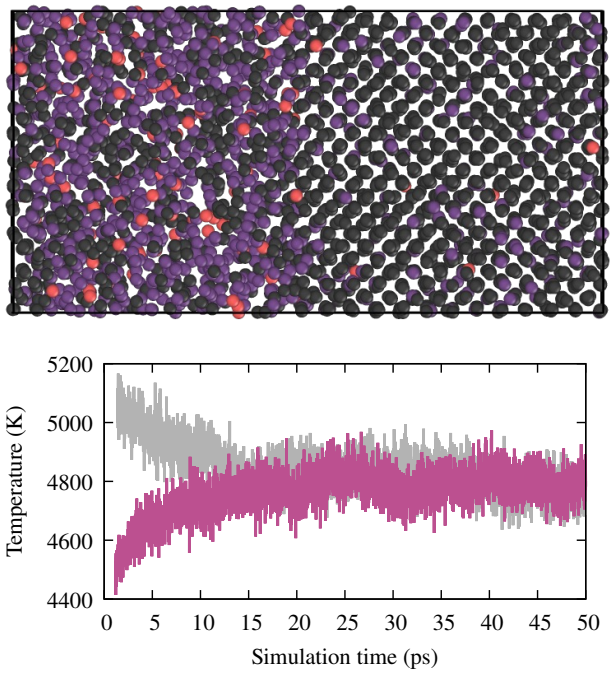

Figure 6: Temperature trajectory of isenthalpic simulations at 30 GPa for diamond melting point prediction (bottom) and representative snapshot showing the two phase structure. Black, purple, and orange atoms are 4-, 3-, and 2-coordinated, respectively.

bridge this capability gap. Hence, we demonstrate use of the 2024 model to predict the melting point at 30 GPa. Isenthalpic ensemble simulations, which are characterized by constant N, P, and H, were for the employed for present predictions.[69] Simulations were run using the LAMMPS MD engine,[70] with ChIMES interactions determined via the ChIMES Calculator library[32] and a 0.5 fs timestep. As shown in Figure 6, two 50 ps two-phase simulations were launched at the target pressure, in which liquid and diamond carbon phases were placed in direct contact. The $\approx$ 5000 atom simulations were initially thermalized at temperatures above and below the expected melt line. In simulations above the melting point, diamond near the interface begins to melt, absorbing heat and driving down the simulation temperature. Conversely, in simulations initialized below the melt line, freezing of liquid near the interface releases heat, driving up the simulation temperature. The two simulations of coexisting liquid-solid phases converge to the same average temperature, which can be taken as the material melting point at the simulation pressure.

Our simulations yield a predicted melting temperature of 4800 ± 60 K at 30 GPa, which

is remarkably consistent with the experimentally determined value of 5000 K,[68] particularly given the comparatively large disagreement between experimental measurements at other state points where multiple measurements are available,[67,71] and that none of our model training data are in this T/P vicinity. Our prediction is also in good agreement with available FP predictions determined using the explicit two-phase bracketing method[15] (5080 K), though those calculations assume zero electronic temperature. We note that the FP values also have a large uncertainty (i.e, up to 1000 K in the T/P vicinity of interest) due to the confluence of use of a bracketing strategy and the computational expense of FP approaches. Notably, other FP predictions using thermodynamic integration methods[72] yielded a melting temperature of $\approx$ 5500 K, while thermodynamic integration using the LCBOPII model yielded a value of $\approx$ 4950 K. Experimental measurements and simulations using bracketing and thermodynamic integration preclude examination of the coexisting solid and liquid phases. The *NPH* simulations enabled by our model overcome this limitation. We find that, as expected, carbon in the diamond phase is chiefly comprised of tetrahedrally (4-fold) coordinated atoms whereas the coexisting liquid phase at this state point exhibits a mixture of tetrahedrally and 3-fold coordinated carbon atoms.

Ultimately, these results suggest that the 2024 model is suitable for carbon phase boundary prediction. Ongoing work is expanding these predictions to include graphite/liquid, graphite/diamond, and diamond/liquid phase boundaries over the range of conditions from 2.5 to 100 GPa, to provide experimentally inaccessible information on interfacial structure and evolution at the phase boundaries, which is critical for understanding phase transition mechanisms governing high-temperature synthesis of carbon-based materials.[36,37,73–76] We expect that the 2024 model will also be useful for the study of amorphous carbon materials, whose properties are dictated by a combination of thermodynamic conditions and kinetic effects on time and length scales which are difficult to study experimentally or fully capture using FP simulations alone.[13,14]

3.2.2 Shockwave-Driven Graphite to Diamond Trasnformation

Under shock compression, graphite is believed to be capable of undergoing a martensitic phase transition to cubic diamond, through a hexagonal diamond intermediate phase.[11,77–81] This shock-induced phase transition was previously explored via FP shock simulations.[11] Since then, DFTB,[79] ML-IAP[80] and molecular mechanics[81] simulations have explored the graphite to diamond phase transformation mechanism via adiabatic compression at constant strain rate, constant volume compression, and by observing evolution in graphite seeded with diamond-like nuclei. In all cases, formation of a hexagonal intermediate phase was identified.

Here, we deploy the 2024 model in simulations following the shockwave-driven approach described in ref. 11 for comparison. Simulations were run using the LAMMPS MD engine,[70] with ChIMES interactions determined via the ChIMES Calculator library.[32] The Multi-Scale Shock Technique[82–84] (MSST) was used to constrain the system to thermodynamic states for a steady planar shock wave within continuum theory (i.e., simulating steady shock waves by constraining the stress and energy of a MD simulation to the Rayleigh line and the Hugoniot energy relations). This approach enables direct prediction of the shock-compressed state provided an initial state and shock velocity, avoiding the need for the much larger system sizes required in direct-shock simulations, or interpolation schemes required for the bracketing simulation approach. Simulations were run for a 768 atom system using a 0.5 fs timestep, cell-mass-like parameter $Q = 1.00 \times 10^{-13}$ kg^2 m^{-4}, macroscopic explicit viscosity parameter $m = 1.2$ kg m^{-1} s^{-1}, and temperature scaling of 0.1. Simulations were run at a shock speed of 12 km s^{-1}, i.e. the same value used in Ref. 11.

As shown in Figure 7, we find qualitative agreement with the results presented in 11, i.e., graphite to cubic diamond transition through a hexagonal diamond intermediate phase. Temperature, pressures, and density for the resulting cubic diamond phase are given in Table 4. Overall, we find values consistent with 11, noting that, in contrast to 11, our simulations used starting configuration initialized from thermalized graphite simulations (i.e., allowing for inter-layer sliding). Between the Born-Oppenheimer (BO) and Car-Parinello

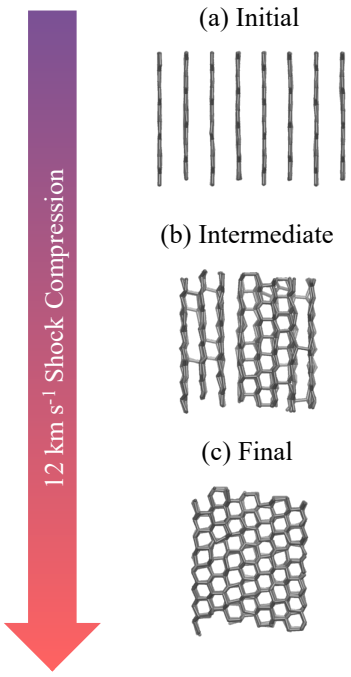

Figure 7: Representative snapshots along an MSST shock-compression simulation showing martensitic transformation from graphite (a) to cubic diamond (c) that proceeds through formation of a short-lived intermediate hexagonal diamond phase (b).

results, our predictions are in better agreement with the former, consistent with the fact that BO-MD was used for training data generation. Our predicted temperature is significantly higher (i.e., in better agreement with the empirical Tersoff potential and equation of state prediction); this discrepancy could be due to either the fact that the present simulations neglect nuclear vibration effects, or because the calculations in ref. 11 were performed at zero electron temperature. Nevertheless, these results show that the present 2024 model is robust for application in non-equilibrium simulations, and extensible to pressures and densities near but outside the initial target regime (e.g., 140 vs. 100 GPa).

3.2.3 Thermally-Driven Nanodiamond to Graphitic Nanoonion Transformation

Experiments dating back to the 1950s have demonstrated that nanodiamond can be rapidly synthesized through detonation of carbon-rich materials.[73] Since then, detonation synthesis of a diversity of interesting and technologically-relevant nanocarbon materials have been realized.[73,74,85–87] However, the mechanism and governing kinetics by which these materials

Table 4: Comparison of properties predicted for cubic diamond synthesized by 12 km s^{-1} shock compression of diamond, predicted by Born-Oppenheimer (BO) and Car-Parinello (CP) FP-MD MSST, classical MD using a Tersoff potential, and equation of state-based prediction (EOS), and the 2024 ChIMES model. P_{cmpr} is the pressure in the compression direction while P_{total} is the overall pressure.

	BO	CP	Tersoff	EOS	ChIMES
T (K)	4084	3351	4981	5300	5000_{100}
P_{cmpr} (GPa)	140	136	130	150	140_3
P_{total} (GPa)	95	83	–	–	96_1
ρ (g cm^{-3})	3.9	3.8	3.6	4.2	4.000_3

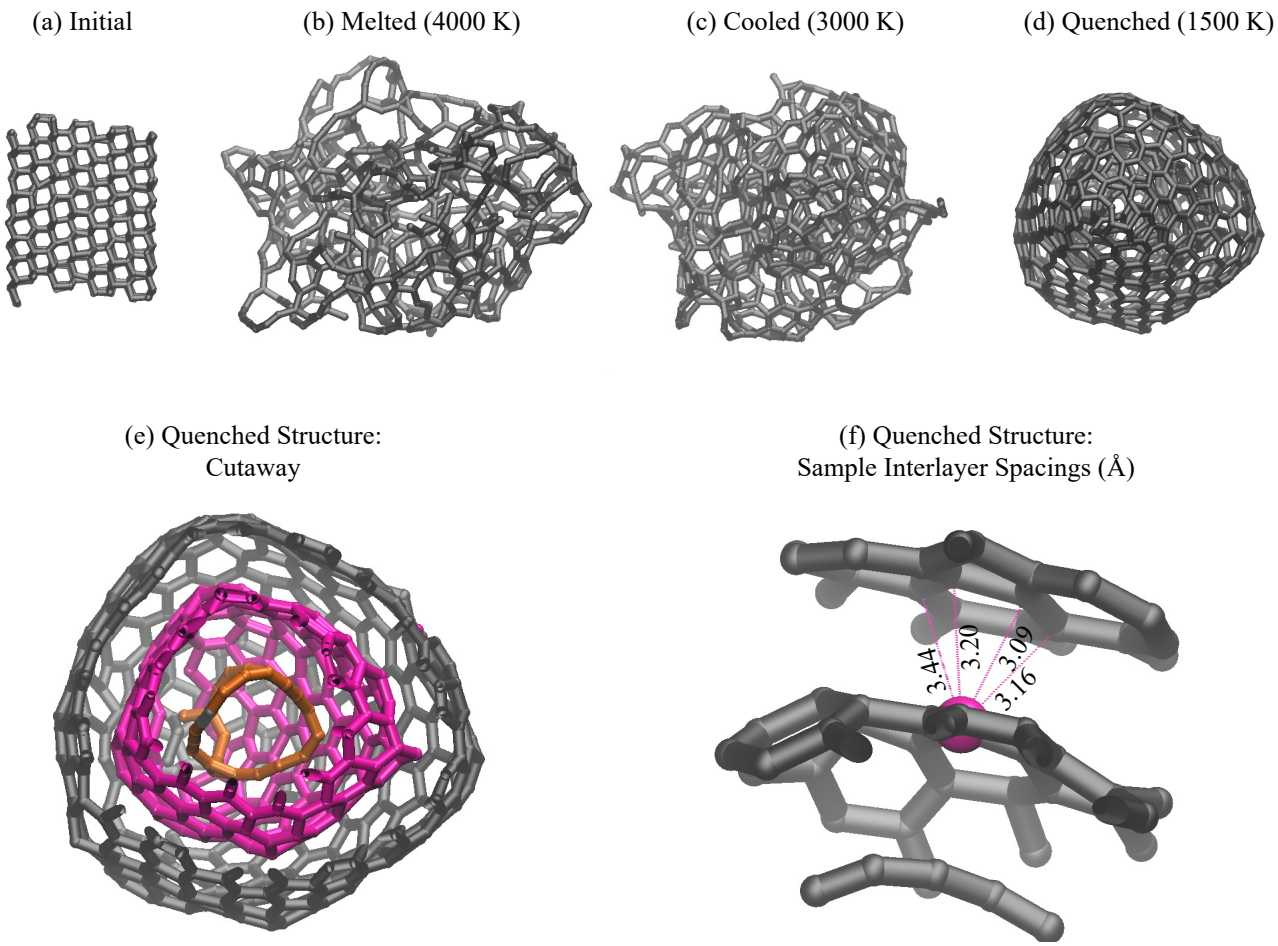

Figure 8: Representative configurations along the thermally driven nanodiamond to graphitic nanoonion transformation pathway are given in a–b. Internal structure of the resulting nanoonion is highlighted in e, with each layer rendered in a different color. A closer look at the layered structure is given in f, with sample interlayer spacing between the two graphitic layers, in terms of distance between a single atom (magenta) of the inner layer and nearby atoms in the outer layer.

evolve remain an open question,[75,88] confining efforts to Edisonian studies. Recently, simulations have been shown capable of providing key missing insights into this phenomenon,[36,37,89]

but necessitate high accuracy interatomic models capable of faithfully describing the complex, reaction-driven phase transformations underlying this process.[27,29,31,33]

To evaluate suitability of the 2024 model for this application space, we probe thermal transformation from a 768 atom cubic nanodiamond to a graphitic ≈ 1 nm nanoonion. The simulations, which were conducted through LAMMPS with the ChIMES Calculator, entailed a series of NVT simulations in vacuum with a 0.5 fs timestep, i.e., melting at 4000 K for 5.5 ps followed by cooling at 3000 K for 15 ps, prior to a final quenching phase at 1500 K for 5.5 ps. Representative snapshots from these simulations are provided in Figure 8.

As shown in Figure 8 panel e, the resulting nanoonion has two distinct layers, with carbyne-like material within its core. This result is particularly interesting within the context of recent experiments probing detonation synthesis of core-shell graphitic nanoonions from 3,4-bis(3-nitrofurazan-4-yl)furoxan (DNTF),[74–76] in which cores appear to have sp character.[90] Notably, prior simulations probing nanoonion formation with the molecular mechanics LCBOPII model found innermost cores still exhibit graphitic character.[91–93]

Figure 8 panel f gives a sense of predicted interlayer spacing, which we find to be less than the 3.35 Å (3.37 Å) expected for planar graphite experimentally (based on the 2024 model), yet consistent with the 3.05 to 3.24 Å, averaging 3.15 Å[75] spacings found for DNTF detonation nanoonions. Overall, these results demonstrate that the present 2024 model is suitable for describing the range of carbon allotropes relevant for T/P from ambient to 10,000 K and 100 GPa, as well as the transformations between them, including those not present in training, (e.g., graphitic nanoonions). Studies of larger nanoparticle sizes, nanoparticle systems under finite pressure, multi-element systems, and amorphous carbon materials are the subject of ongoing work.[36,37]

4 Conclusions

A new, T/P transferable model for carbon targeting application under extreme conditions of up to 10,000 K and 100 GPa has been developed using the physics-informed ChIMES ML-IAP, and a new multifidelity learning strategy has been presented. The resulting 2024 model has been shown to substantially outperform the previous 2017 parameterization and provide accurate modeling of a diversity of carbon allotropes as well as transformations between them. This model is of immediate utility for planetary, astrochemical, detonation, and synthesis sciences, enabling investigation of carbon equation of state and evolution in, e.g., planetary interiors, presolar grains, detonation nanocarbon, and lab-synthesized carbon nanomaterials. Critically, the underlying form of ChIMES also allows for *direct chemical extensibility*, meaning the 2024 ChIMES-C parameters can serve as a fixed foundation upon which models for multielement systems can be generated. Ultimately, this can expedite generation of future models (i.e., by reducing the number of parameters that must be fit) while simultaneously ensuring carbon interactions are well-described and not compromised during the fitting process. Ongoing work is exploring this unique transfer learning strategy.

5 Author Contributions

R.K.L Conceived and directed the study and conducted simulations and analysis. S.H. contributed to generation of the PBE-D2 training set. Y.L. contributed to application simulations. N.G. contributed to development of the ChIMES codebase. V.L., S.B., and S.H. and N.G. contributed to ideation, and all authors contributed to scientific discussions and manuscript development.

6 Competing Interests

The authors declare no competing interests.

7 Data Availability

The datasets used and/or analyzed during the current study available from the corresponding author on reasonable request.

8 Code availability

The underlying code for this study is available on GitHub in the links given by refs 55, 32, and 56.

Acknowledgement

A portion of this work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. This work was partially supported by LLNL LDRD 23-SI-006. LLNL-JRNL-861515.

Supporting Information Available

Parameters for the 2017 and 2024 ChIMES-Carbon model are available free of charge within the ChIMES Calculator repository.[32]

References

(1) Gold, J. S.; Bassett, W. A.; Weathers, M. S.; Bird, J. M. Melting of diamond. *Science* **1984**, *225*, 921–922.

(2) Bundy, F.; Bassett, W.; Weathers, M.; Hemley, R.; Mao, H.; Goncharov, A. The pressure-temperature phase and transformation diagram for carbon; updated through 1994. *Carbon* **1996**, *34*, 141–153.

(3) Ross, M. The ice layer in Uranus and Neptune–diamonds in the sky? *Nature* **1981**, *292*, 435–436.

(4) MacKinnon, A. J.; Meezan, N. B.; Ross, J. S.; Pape, S. L.; Hopkins, L. B.; Divol, L.; Ho, D.; Milovich, J.; Pak, A.; Ralph, J.; Döppner, T.; Patel, P. K.; Thomas, C.; Tommasini, R.; Haan, S.; MacPhee, A. G.; McNaney, J.; Caggiano, J.; Hatarik, R.; Bionta, R.; Ma, T.; Spears, B.; Rygg, J. R.; Benedetti, L. R.; Town, R. P. J.; Bradley, D. K.; Dewald, E. L.; Fittinghoff, D.; Jones, O. S.; Robey, H. R.; Moody, J. D.; Khan, S.; Callahan, D. A.; Hamza, A.; Biener, J.; Celliers, P. M.; Braun, D. G.; Erskine, D. J.; Prisbrey, S. T.; Wallace, R. J.; Kozioziemski, B.; Dylla-Spears, R.; Sater, J.; Collins, G.; Storm, E.; Hsing, W.; Landen, O.; Atherton, J. L.; Lindl, J. D.; Edwards, M. J.; Frenje, J. A.; Gatu-Johnson, M.; Li, C. K.; Petrasso, R.; Rinderknecht, H.; Rosenberg, M.; guin, F. H. S.; Zylstra, A.; Knauer, J. P.; Grim, G.; Guler, N.; Merrill, F.; Olson, R.; Kyrala, G. A.; Kilkenny, J. D.; Nikroo, A.; Moreno, K.; Hoover, D. E.; Wild, C.; Werner, E. High-density carbon ablator experiments on the National Ignition Facility. *Phys. Plasmas* **2014**, *21*, 056318.

(5) Eggert, J.; Hicks, D.; Celliers, P.; Bradley, D.; McWilliams, R.; Jeanloz, R.; Miller, J.; Boehly, T.; Collins, G. Melting temperature of diamond at ultrahigh pressure. *Nat. Phys.* **2010**, *6*, 40–43.

(6) Wang, X.; Scandolo, S.; Car, R. Carbon phase diagram from ab initio molecular dynamics. *Phys. Rev. Lett.* **2005**, *95*, 185701.

(7) Sun, J.; Klug, D. D.; Martoňák, R. Structural transformations in carbon under extreme pressure: beyond diamond. *J. Chem. Phys.* **2009**, *130*, 194512.

(8) Jephcoat, A. P.; Besedin, S. P. Temperature measurement and melting determination in the laser-heated diamond-anvil cell. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* **1996**, *354*, 1333–1360.

(9) Brantley, D. A.; Crum, R. S.; Akin, M. C. Comparing temperature convergence of shocked thin films of tin and iron to a bulk temperature source. *Journal of Applied Physics* **2021**, *129*.

(10) Gygi, F.; Galli, G. Ab initio simulation in extreme conditions. *Materials Today* **2005**, *8*, 26–32.

(11) Mundy, C. J.; Curioni, A.; Goldman, N.; Will Kuo, I.-F.; Reed, E. J.; Fried, L. E.; Ianuzzi, M. Ultrafast transformation of graphite to diamond: An ab initio study of graphite under shock compression. *The Journal of chemical physics* **2008**, *128*, 184701.

(12) Wu, C. J.; Glosli, J. N.; Galli, G.; Ree, F. H. Liquid-liquid phase transition in elemental carbon: A first-principles investigation. *Physical review letters* **2002**, *89*, 135701.

(13) Steele, B. A.; Bastea, S.; Kuo, I.-F. W. Ab initio structural dynamics of pure and nitrogen-containing amorphous carbon. *Scientific Reports* **2023**, *13*, 19657.

(14) Thapa, R.; Ugwumadu, C.; Nepal, K.; Trembly, J.; Drabold, D. Ab initio simulation of amorphous graphite. *Physical Review Letters* **2022**, *128*, 236402.

(15) Correa, A. A.; Bonev, S. A.; Galli, G. Carbon under extreme conditions: Phase boundaries and electronic properties from first-principles theory. *Proceedings of the National Academy of Sciences* **2006**, *103*, 1204–1208.

(16) Los, J.; Fasolino, A. Intrinsic long-range bond-order potential for carbon: performance in Monte Carlo simulations of graphitization. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2003**, *68*, 024107.

(17) Ghiringhelli, L. M.; Valeriani, C.; Los, J.; Meijer, E. J.; Fasolino, A.; Frenkel, D. State-of-the-art models for the phase diagram of carbon and diamond nucleation. *Molecular Physics* **2008**, *106*, 2011–2038.

(18) Brenner, D. W.; Shenderova, O. A.; Harrison, J. A.; Stuart, S. J.; Ni, B.; Sinnott, S. B. A second-generation reactive empirical bond order (REBO) potential energy expression for hydrocarbons. *J. Phys. Condens. Matter* **2002**, *14*, 783.

(19) Ghiringhelli, L. M.; Meijer, E. J. *Computer-Based Modeling of Novel Carbon Systems and Their Properties*; Springer, 2010; pp 1–36.

(20) Ghiringhelli, L. M.; Los, J. H.; Meijer, E. J.; Fasolino, A.; Frenkel, D. Modeling the phase diagram of carbon. *Physical review letters* **2005**, *94*, 145701.

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

(22) Rowe, P.; Deringer, V. L.; Gasparotto, P.; Csányi, G.; Michaelides, A. An accurate and transferable machine learning potential for carbon. *The Journal of Chemical Physics* **2020**, *153*.

(23) Willman, J. T.; Williams, A. S.; Nguyen-Cong, K.; Thompson, A. P.; Wood, M. A.; Belonoshko, A. B.; Oleynik, I. I. Quantum accurate SNAP carbon potential for MD shock simulations. AIP Conference Proceedings. 2020.

(24) Willman, J. T.; Nguyen-Cong, K.; Williams, A. S.; Belonoshko, A. B.; Moore, S. G.; Thompson, A. P.; Wood, M. A.; Oleynik, I. I. Machine learning interatomic potential for simulations of carbon at extreme conditions. *Physical Review B* **2022**, *106*, L180101.

(25) Nguyen-Cong, K.; Willman, J. T.; Moore, S. G.; Belonoshko, A. B.; Gayatri, R.; Weinberg, E.; Wood, M. A.; Thompson, A. P.; Oleynik, I. I. Billion atom molecular dynamics simulations of carbon at extreme conditions and experimental time and length scales. Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis. 2021; pp 1–12.

(26) Qamar, M.; Mrovec, M.; Lysogorskiy, Y.; Bochkarev, A.; Drautz, R. Atomic cluster expansion for quantum-accurate large-scale simulations of carbon. *Journal of Chemical Theory and Computation* **2023**, *19*, 5151–5167.

(27) Lindsey, R. K.; Pham, C. H.; Fried, L. E.; Goldman, N.; Bastea, S. *ChIMES: A Machine-Learned Interatomic Model Targeting Improved Description of Condensed Phase Chemistry in Energetic Materials*; 2020.

(28) Lindsey, R. K.; Fried, L. E.; Goldman, N. ChIMES: A Force Matched Potential with Explicit Three-Body Interactions for Molten Carbon. *J. Chem. Theory Comput.* **2017**, *13*, 6222–6229.

(29) Lindsey, R.; Fried, L. E.; Goldman, N.; Bastea, S. Active Learning for Robust, High-Complexity Reactive Atomistic Simulations. *J. Chem. Phys.* **2020**, *153*, 134117.

(30) Lindsey, R. K.; Huy Pham, C.; Goldman, N.; Bastea, S.; Fried, L. E. Machine-Learning a Solution for Reactive Atomistic Simulations of Energetic Materials. *Propellants, Explosives, Pyrotechnics* **2022**, e202200001.

(31) Lindsey, R. K.; Bastea, S.; Lyu, Y.; Hamel, S.; Goldman, N.; Fried, L. E. Chemical evolution in nitrogen shocked beyond the molecular stability limit. *The Journal of Chemical Physics* **2023**, *159*.

(32) https://github.com/rk-lindsey/chimes_calculator. 2020.

(33) Lindsey, R. K.; Goldman, N.; Fried, L. E.; Bastea, S. Many-body reactive force field development for carbon condensation in C/O systems under extreme conditions. *J. Chem. Phys.* **2020**, *153*, 054103.

(34) Pham, C. H.; Lindsey, R. K.; Fried, L. E.; Goldman, N. Calculation of the detonation state of HN3 with quantum accuracy. *The Journal of Chemical Physics* **2020**, *153*, 224102.

(35) Lindsey, R. K.; Bastea, S.; Goldman, N.; Fried, L. E. Investigating 3, 4-bis (3-nitrofurazan-4-yl) furoxan detonation with a rapidly tuned density functional tight binding model. *The Journal of Chemical Physics* **2021**, *154*, 164115.

(36) Lindsey, R. K.; Goldman, N.; Fried, L. E.; Bastea, S. Chemistry-mediated Ostwald ripening in carbon-rich C/O systems at extreme conditions. *Nature Communications* **2022**, *13*, 1424.

(37) Armstrong, M. R.; Lindsey, R. K.; Goldman, N.; Nielsen, M. H.; Stavrou, E.; Fried, L. E.; Zaug, J. M.; Bastea, S. Ultrafast shock synthesis of nanocarbon from a liquid precursor. *Nature communications* **2020**, *11*, 353.

(38) Pham, C. H.; Lindsey, R. K.; Fried, L. E.; Goldman, N. High-accuracy semiempirical quantum models based on a minimal training set. *The Journal of Physical Chemistry Letters* **2022**, *13*, 2934–2942.

(39) Lindsey, R. K.; Fried, L. E.; Goldman, N. Application of the ChIMES Force Field to Nonreactive Molecular Systems: Water at Ambient Conditions. *J. Chem. Theory Comput.* **2019**, *15*, 436–447.

(40) Tersoff, J. Modeling solid-state chemistry: Interatomic potentials for multicomponent systems. *Physical Review B* **1989**, *39*, 5566.

(41) Kresse, G.; Hafner, J. Ab initio molecular dynamics for liquid metals. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1993**, *47*, 558.

(42) Kresse, G.; Hafner, J. Ab initio molecular-dynamics simulation of the liquid-metal–amorphous-semiconductor transition in germanium. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1994**, *49*, 14251.

(43) Kresse, G.; Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **1996**, *6*, 15–50.

(44) Kresse, G.; Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *54*, 11169.

(45) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865.

(46) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple [Erratum to Phys. Rev. Lett. 77, 3865 (1996)]. *Phys. Rev. Lett.* **1997**, *78*, 1396–1396.

(47) Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1994**, *50*, 17953.

(48) Kresse, G.; Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1999**, *59*, 1758.

(49) Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *Journal of computational chemistry* **2006**, *27*, 1787–1799.

(50) Lindsey, R. K.; Kroonblawd, M. P.; Fried, L. E.; Goldman, N. Force matching approaches to extend density functional theory to large time and length scales. *Computational Approaches for Chemistry Under Extreme Conditions* **2019**, 71–93.

(51) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Millman, K. J.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; Polat, İ.; Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P.; SciPy 1.0 Contributors, SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* **2020**, *17*, 261–272.

(52) Efron, B.; Hastie, T.; Johnstone, I.; Tibshirani, R. Least angle regression. *Ann. Stat.* **2004**, *32*, 407–499.

(53) Friedman, J.; Hastie, T.; Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **2010**, *33*, 1.

(54) Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Statist. Soc. B* **1996**, *58*, 267–288.

(55) https://github.com/rk-lindsey/chimes_lsq. 2022.

(56) https://github.com/rk-lindsey/al_driver. 2022.

(57) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Fraunheim, T.; Suhai, S.; Seifert, G. Self-consistent-charge density-functional tight-binding method for simulations of complex material properties. *Physical Review B* **1998**, *58*, 7260.

(58) Hourahine, B.; Aradi, B.; Blum, V.; Bonafe, F.; Buccheri, A.; Camacho, C.; Cevallos, C.; Deshaye, M.; Dumitrică, T.; Dominguez, A., et al. DFTB+, a software package for efficient approximate density functional theory based atomistic simulations. *J. Chem. Phys.* **2020**, *152*, 124101.

(59) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Physical Review B* **1998**, *58*, 7260.

(60) Rappé, A. K.; Casewit, C. J.; Colwell, K.; Goddard III, W. A.; Skiff, W. M. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *Journal of the American chemical society* **1992**, *114*, 10024–10035.

(61) Goldman, N.; Fried, L. E. Extending the density functional tight binding method to carbon under extreme conditions. *J. Phys. Chem. C* **2011**, *116*, 2198–2204.

(62) Goldman, N.; Goverapet Srinivasan, S.; Hamel, S.; Fried, L. E.; Gaus, M.; Elstner, M. Determination of a density functional tight binding model with an extended basis set and three-body repulsion for carbon under extreme pressures and temperatures. *The Journal of Physical Chemistry C* **2013**, *117*, 7885–7894.

(63) Occelli, F.; Loubeyre, P.; LeToullec, R. Properties of diamond under hydrostatic pressures up to 140 GPa. *Nature materials* **2003**, *2*, 151–154.

(64) Zhao, Y. X.; Spain, I. L. X-ray diffraction data for graphite to 20 GPa. *Physical Review B* **1989**, *40*, 993.

(65) Goldman, N.; Goverapet Srinivasan, S.; Hamel, S.; Fried, L. E.; Gaus, M.; Elstner, M. Determination of a density functional tight binding model with an extended basis set and three-body repulsion for carbon under extreme pressures and temperatures. *The Journal of Physical Chemistry C* **2013**, *117*, 7885–7894.

(66) Xing, M.; Li, X. Structural, mechanical, anisotropic and electronic properties of BC8 carbon under pressure. *Chemical Physics* **2023**, *574*, 112050.

(67) Van Thiel, M.; Ree, F. High-pressure liquid-liquid phase change in carbon. *Physical Review B* **1993**, *48*, 3591.

(68) Bundy, F.; Bassett, W.; Weathers, M.; Hemley, R.; Mao, H.; Goncharov, A. The pressure-temperature phase and transformation diagram for carbon; updated through 1994. *Carbon* **1996**, *34*, 141–153.

(69) Andersen, H. C. Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of chemical physics* **1980**, *72*, 2384–2393.

(70) Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. *J. Phys. Chem. C* **1995**, *117*, 1–19.

(71) Togaya, M. Pressure dependences of the melting temperature of graphite and the electrical resistivity of liquid carbon. *Physical review letters* **1997**, *79*, 2474.

(72) Wang, X.; Scandolo, S.; Car, R. Carbon phase diagram from ab initio molecular dynamics. *Physical review letters* **2005**, *95*, 185701.

(73) Greiner, N. R.; Phillips, D.; Johnson, J.; Volk, F. Diamonds in detonation soot. *Nature* **1988**, *333*, 440.

(74) Bagge-Hansen, M.; Lauderbach, L.; Hodgin, R.; Bastea, S.; Fried, L.; Jones, A.; van Buuren, T.; Hansen, D.; Benterou, J.; May, C.; Graber, T.; Jensen, B. J.; Willey, T. M. Measurement of carbon condensates using small-angle X-ray scattering during detonation of the high explosive hexanitrostilbene. *J. Appl. Phys.* **2019**, *117*, 245902.

(75) Bagge-Hansen, M.; Bastea, S.; Hammons, J. A.; Nielsen, M. H.; Lauderbach, L.; Hodgin, R.; Pagoria, P.; May, C.; Aloni, S.; Jones, A., et al. Detonation synthesis of carbon nano-onions via liquid carbon condensation. *Nature communications* **2019**, *10*, 1–8.

(76) Hammons, J. A.; Nielsen, M. H.; Bagge-Hansen, M.; Bastea, S.; May, C.; Shaw, W. L.; Martin, A.; Li, Y.; Sinclair, N.; Lauderbach, L. M., et al. Submicrosecond aggregation during detonation synthesis of nanodiamond. *The Journal of Physical Chemistry Letters* **2021**, *12*, 5286–5293.

(77) Armstrong, M. R.; Radousky, H. B.; Austin, R. A.; Tschauner, O.; Brown, S.; Gleason, A. E.; Goldman, N.; Granados, E.; Grivickas, P.; Holtgrewe, N., et al. Highly ordered graphite (HOPG) to hexagonal diamond (lonsdaleite) phase transition observed on picosecond time scales using ultrafast x-ray diffraction. *Journal of Applied Physics* **2022**, *132*.

(78) Stavrou, E.; Bagge-Hansen, M.; Hammons, J. A.; Nielsen, M. H.; Steele, B. A.; Xiao, P.; Kroonblawd, M. P.; Nelms, M. D.; Shaw, W. L.; Bassett, W., et al. Detonation-induced transformation of graphite to hexagonal diamond. *Physical Review B* **2020**, *102*, 104116.

(79) Kroonblawd, M. P.; Goldman, N. Mechanochemical formation of heterogeneous diamond structures during rapid uniaxial compression in graphite. *Physical Review B* **2018**, *97*, 184106.

(80) Khaliullin, R. Z.; Eshet, H.; Kühne, T. D.; Behler, J.; Parrinello, M. Nucleation mechanism for the direct graphite-to-diamond phase transition. *Nature materials* **2011**, *10*, 693–697.

(81) Xie, H.; Yin, F.; Yu, T.; Wang, J.-T.; Liang, C. Mechanism for direct graphite-to-diamond phase transition. *Scientific reports* **2014**, *4*, 5930.

(82) Reed, E. J.; Fried, L. E.; Joannopoulos, J. A method for tractable dynamical studies of single and double shock compression. *Physical Review Letters* **2003**, *90*, 235503.

(83) Reed, E.; Fried, L.; Manaa, M.; Joannopoulos, J. *A Multi-Scale Approach to Molecular Dynamics Simulations of Shockwaves*; Elsevier New York, 2005; Chapter 10, pp 297–326.

(84) Reed, E. J.; Fried, L. E.; Henshaw, W. D.; Tarver, C. M. Analysis of simulation technique for steady shock waves in materials with analytical equations of state. *Physical Review E* **2006**, *74*, 056706.

(85) Mochalin, V. N.; Shenderova, O.; Ho, D.; Gogotsi, Y. The properties and applications of nanodiamonds. *Nature nanotechnology* **2012**, *7*, 11.

(86) Titov, V.; Anisichkin, V.; Mal'kov, I. Y. Synthesis of ultradispersed diamond in detonation waves. *Combustion, Explosion and Shock Waves* **1989**, *25*, 372–379.

(87) Satonkina, N. P. The dynamics of carbon nanostructures at detonation of condensed high explosives. *J. Appl. Phys.* **2015**, *118*, 245901.

(88) Anisichkin, V. On the mechanism of the detonation of organic high explosives. *Russian Journal of Physical Chemistry B* **2016**, *10*, 451–455.

(89) Kroonblawd, M. P.; Lindsey, R. K.; Goldman, N. Synthesis of functionalized nitrogen-containing polycyclic aromatic hydrocarbons and other prebiotic compounds in impacting glycine solutions. *Chemical science* **2019**, *10*, 6091–6098.

(90) Willey, T.; Hammons, J.; Nielsen, M.; Ebrahim, A.; Taylor, G.; Paredes Mellone, O.; Lauderbach, L.; Hodgin, R.; Sinclair, N.; Shuman, A., et al. Monitoring Formation of Detonation Nanodiamond and Other Novel Carbon Nanostructures Using Advanced X-ray Scattering and Spectroscopy Techniques. *Bulletin of the American Physical Society* **2023**,

(91) Pineau, N.; Soulard, L.; Los, J.; Fasolino, A. Theoretical study of the nucleation/growth process of carbon clusters under pressure. *The Journal of chemical physics* **2008**, *129*.

(92) Pineau, N.; Chevrot, G.; Bourasseau, E.; Maillet, J.-B.; Los, J.; Fasolino, A. Phase diagram and thermodynamic properties of nanocarbons in detonation conditions from atomistic simulations using the LCBOPII potential. APS Shock Compression of Condensed Matter Meeting Abstracts. 2009; pp CKD–20.

(93) Chevrot, G.; Sollier, A.; Pineau, N. Molecular dynamics and kinetic study of carbon coagulation in the release wave of detonation products. *The Journal of Chemical Physics* **2012**, *136*.

TOC Graphic