LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-TR-814690

ChIMES: A Machine-Learned Interatomic Model Targeting Improved Description of Condensed Phase Chemistry in Energetic Materials

R. K. Lindsey, C. H. Pham, L. E. Fried, N. Goldman, S. Bastea

September 18, 2020

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

ChIMES: A Machine-Learned Interatomic Model Targeting Improved Description of Condensed Phase Chemistry in Energetic Materials

Rebecca K. Lindsey,* C. Huy Pham, Laurence E. Fried, Nir Goldman, and Sorin Bastea
Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, California 94550, United States
(Dated: September 16, 2020)

In this report we detail completion of a Physics and Engineering Model Level Two Milestone targeting improved reactive interatomic potentials (IAPs) for energetic materials (EM) through machine learning. The specific goals of this milestone were to develop, validate, and document a new reactive molecular dynamics method for EM, based on machine learning by (1) generating databases of first-principles-derived forces, stresses, and energies for HN_3 and 3,4-bis(3-nitrofurazan-4-yl)furoxan (DNTF) (2) generate atomistic force fields from these databases *via* ML, and (3) benchmark model performance against first principles calculations. These goals were achieved by (1) further developing a machine learned reactive IAP and generation approach (i.e. the Chebyshev Interaction Model for Efficient Simulation or "ChIMES"), for which resulting IAPs can approach the predictive power of quantum-mechanical approaches at a fraction of the computational expense, and (2) applying the ChIMES framework to develop models for HN_3 and DNTF.

We find that for simple energetic materials like HN_3, high accuracy ChIMES models can be obtained through application of a fitting approach that does not use active machine learning. We demonstrate the suitability of ChIMES models for simulations involving EM by using the HN_3 model in multiscale shock technique simulations to predict the HN_3 Chapman-Jouguet detonation state and investigate chemical evolution out to 1 ns following shock compression. This model is then used in larger direct shock (DS) simulations for a preliminary investigation of how bubbles (i.e. voids) influence material response under shock compression. We find that more complex EM (i.e. DNTF) necessitate a more sophisticated fitting approach, and develop a new active learning method and python tool to meet this challenge. We demonstrate that this fitting approach yields ChIMES models that out-perform commonly used standard reactive IAPs as well as semi-empirical quantum methods, and discuss the systematic improvability of these actively learned ChIMES models.

We also describe challenges related to model development for EM such as DNTF, for which few experimental or previous simulation data are available (e.g. which could otherwise inform generation of training data). To overcome this issue, we establish a semi-empirical quantum ChIMES capability which can be used to efficiently map out relevant thermodynamic and configurational space, and generate ChIMES-IAP training data in a multiscale manner. We also show that these semi-empirical quantum ChIMES models can be used to generate predictions for the shock Hugoniot (the Hugoniot is the locus of thermodynamic states found in a shocked material) equation of state, investigate related thermochemistry, and explore carbon condensation following shock compression. This work represents a substantial advance in our atomistic modeling capability for EM that will provide much-needed information on the chemistry of detonation for continued development of continuum models based on the Cheetah thermochemical code.

*To whom correspondence should be addressed:
lindsey11@llnl.gov

I. MILESTONE DEFINITION

- **Title:** Machine Learning for High Explosive Reactions
- **Completion date:** 09/30/2020
- **Description:** The ability to predict high explosive reactions at the molecular level is necessary to improve high explosive equation of state and reactive burn models. We will present, document, and validate a new method for the reactive molecular dynamics of high explosives and other molecular compounds based on machine learning. Results will be used in the future to improve high explosive equation of state and reactive burn models in the Cheetah thermochemical code.
- **Completion criteria:** Databases of forces, stresses, and energies from first principles density functional calculations will be collected for simple compounds to demonstrate the method. Atomic force fields will be generated from the databases using machine learning and regularized regression techniques. The performance of the force fields will be assessed through comparison to density functional calculations. Application to a simplified high explosive (HN_3) will be demonstrated. Areas for improvement will be identified and discussed.
- **Milestone certification method:** A program review is conducted and its results are documented. Professional documentation, such as a report or a set of view-graphs with a written summary, is prepared as a record of milestone completion.

II. INTRODUCTION

LLNL has based advanced continuum models [1] of high explosives on the Cheetah thermochemical code [2], which uses a statistical mechanical model with explicit chemistry and simplified chemical kinetics to predict the equation of state and burn rates of high explosives. In order to improve Cheetah predictions, a more detailed understanding of chemical processes occurring in detonation is required. One route to such information is through reactive molecular dynamics simulations, which have the ability to predict the quantity of chemical species formed during detonation and their rates. Understanding the rate of species formation is particularly important for non-ideal explosives such as TATB (i.e. 2,4,6-triamino-1,3,5- trinitrobenzene), where chemical processes, such as soot formation, occurring on timescales greater than 100 ns influence the hydrodynamic energy delivery during detonation [3]. However, the accuracy of molecular dynamics simulations, depends crucially on the underlying inter-atomic potential (IAP). Traditionally, atomistic simulation models have largely fallen into one of 3 categories: quantum mechanical, semi-empirical quantum mechanical, or molecular mechanical, each affording a unique balance of accuracy, predictive power, and accessible time and length scales. For example, quantum mechanical approaches (e.g. Kohn Sham density functional theory [4], DFT) have been shown to be highly suitable for equation of state and chemistry predictions in materials under extreme temperature and pressure conditions [5–7] (e.g. 1000s of K and 10s to 100s of GPa), however they remain confined to 10^0 nm and 10^1 ps simulation scales due to the high computational cost and poor scaling with the number of electrons (formally N^3). Semi-empirical quantum approaches are able to reach 100s of ps timescales while retaining much of the accuracy of DFT through the use of approximate quantum mechanics (e.g. density functional theory tight binding [8], DFTB, which relies on a pre-tabulated Hamiltonian) and are thus helpful in probing longer timescale chemistry but remain confined to equivalently small system sizes because the semi-empirical approach follows the same size scaling of the DFT method.

In contrast to DFT and DFTB, molecular mechanical (MM), or "force field" simulations use IAPs based on an explicit analytical form. These models exhibit high efficiency (e.g. linear scaling of computational cost with increased number of atoms, and a small single-atom overhead) and are thus heavily utilized for large scale simulations of $> 10^5$ atoms, (e.g. those used to investigate hot spot initiation [9, 10], etc). However, MM models (specifically those targeting the condensed phase chemistry under extreme conditions of interest relevant to EM) have many associated disadvantages including that:

1. They are minimally transferable and both time-consuming and challenging to parameterize.
2. The underlying form is generally not compatible with description of materials under high pressure, such that accuracy in one area (e.g. speciation) comes at the expense of accuracy in another (e.g. chemical kinetics, equation of state, etc.).
3. Models are not systematically improvable.

Nevertheless, these models are continually used due to a lack of alternative options.

To overcome these limitations, we have developed the Chebyshev Interaction Model for Efficient Simulation [11–14] (ChIMES), a machine learned reactive interatomic model and development approach targeting condensed-phase chemistry in molecular organics (e.g. EM) under extreme conditions. Machine Learning (ML) has gained significant attention in the force field development community, largely due to the afforded versatility and potential to significantly decrease the human effort required to generate high fidelity, complex models. In general, ML methods are well suited for problems demanding first principles-level accuracy in conjunction with the computational efficiency of force field-based methods.

This is particularly true for materials under extreme conditions (e.g. 1000s of K and 10s to 100s of GPa) which can be highly reactive and exhibit phase separation over several ns, leading to chemical and structural heterogeneity on several-nm scales (e.g. in the case of reaction-driven carbon condensation in shock-compressed materials [15–17]). In these scenarios, machine learned interatomic potentials (ML-IAPs) can offer an ideal balance between predictive power and computational efficiency, allowing simulations to more closely approach experimental time and length scales than possible with quantum simulations alone. Though many success stories exist surrounding use of ML for atomistic model development (e.g. the Spectral Neighbor Analysis Potential [18] (SNAP), which targets non-molecular materials and the Accurate Neural network engine for Molecular Energies [19] (ANI), which targets gas-phase chemistry and non-molecular materials), our approach is the only method targeting condensed phase chemistry in organic materials (i.e EM).

In this report, we describe the ChIMES approach and its application to two energetic materials. We note that, although ChIMES is in principle applicable to any energetic material, the ChIMES approach becomes increasingly complicated as the number of atom types in the energetic material increases. For this reason, we chose two simplified energetic materials for study in this milestone. Hydrazoic acid (HN_3) is one of the simplest known energetic materials, (i.e. comprised of only two elements) and was the first EM studied with ChIMES. Though of low chemical complexity, this liquid EM undergoes band gap closures during its detonation, which presents significant challenges to IAP model development (i.e. the system transitions from insulating to nearly metallic, and back to insulating electronic configurations during the detonation process).

DNTF was the next EM investigated, serving as a logical stepping stone from HN_3 to TATB. This material is substantially more complex than HN_3 in that it contains it contains 3 atom types (i.e. C, O, and N, the former two of which can form a greater maximum number of bonds than H, giving rise to a far more complex manifold of chemistry). Moreover, DNTF exists as a solid under ambient conditions (i.e. material strength factors into shock response), and it is known to form detonation soot, which involves a complicated and inherently multiscaled reactive process. Nevertheless, DNTF presents a less formidable challenge than TATB, which similarly exists as a solid under ambient conditions and forms detonation soot, but comprises an additional atom type (i.e. it exhibits even more complex chemistry thereby necessitating higher complexity models) and undergoes significantly slower chemistry (i.e. complicating training data generation and model validation). Ultimately, our systematic approach to ChIMES extension ensures we are able to produce robust and reproducible models of high fidelity, offering an optimal balance between accuracy, complexity, and efficiency.

Challenges unique to IAP development for simulation of EM under extreme conditions are presented, along with two new approaches to overcome them (e.g. an active learning framework and a multiscale ChIMES approach). We show that resulting ChIMES EM models can approach quantum accuracy and far exceed the performance of existing models and represents a significant advancement in our multiscale model development capability. ChIMES simulation results are compared to predictions by the Cheetah thermochemical code. We conclude with a discussion of remaining challenges for ML EM model development, and future directions. We note that, for the remainder of this work, IAP is used to refer to an interatomic force field rather than a quantum simulation.

III. CHIMES OVERVIEW

ChIMES [11–14] is a recently developed ML-IAP explicitly describing many-body interaction energies through linear combinations of Chebyshev polynomials:

$$E_{n_\mathrm{B}} = \sum_{i_1}^{n_\mathrm{a}} {}^1E_{i_1} + \sum_{i_1>i_2}^{n_\mathrm{a}} {}^2E_{i_1 i_2} + \sum_{i_1>i_2>i_3}^{n_\mathrm{a}} {}^3E_{i_1 i_2 i_3} + \cdots + \sum_{i_1>i_2\ldots\, i_{n_\mathrm{B}-1}>i_{n_\mathrm{B}}}^{n_\mathrm{a}} {}^{n_\mathrm{B}}E_{i_1 i_2 \ldots i_{n_\mathrm{B}}}, \tag{1}$$

where E_{n_B} is the total ChIMES system energy, n_B is the maximum bodiedness, ${}^nE_{i_1 i_2 \ldots\, i_n}$ is the n-body ChIMES potential 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 the ChIMES framework, single-body energies are constant values and n-body energies are constructed from the product of polynomials of transformed atom pair distances. 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. Taking as an example a 3-body interaction, we define the following: $\boldsymbol{A} = \{i, j, k\}$ is the index over atoms within an interaction cluster, with the corresponding set of pairs given by $\boldsymbol{P} = \{ij, ik, jk\}$, their element pair types by $\boldsymbol{E} = \{e_i e_j, e_i e_k, e_j e_k\}$, and the polynomial orders (we refer to the polynomial degree as its order) for each pair given by $\boldsymbol{O} = \{\alpha, \beta, \gamma\}$. Two mapping functions are used to relate pair indices $\boldsymbol{P}$ to the three aforementioned pair properties: $m_1 = \boldsymbol{P} \rightarrow \boldsymbol{E}$, and $m_2 = \boldsymbol{P} \rightarrow \boldsymbol{O}$, where the index y refers to a particular component of $\boldsymbol{P}$, defining an interaction pair.

Using these definitions, we write the generalized ChIMES energy for a cluster of n atoms as:

$$^{n}E_{\boldsymbol{A}} \propto \sum_{\boldsymbol{O}}^{\mathcal{O}_n^*} c_{\boldsymbol{O}}^{\boldsymbol{E}} \prod_{y \in \boldsymbol{P}} T_{m_2(y)} \left(s_y^{m_1(y)} \right). \tag{2}$$

As is given above, the $\sum_{\boldsymbol{O}}$ notation indicates a multiple sum for which there are $\binom{n}{2}$ distinct indices, $\mathcal{O}_n^*$ is the maximum polynomial order for an n body interactions, and the asterisk indicates a sufficient number of non-zero terms exist that the graph formed by the edges of interacting atoms connects all n atoms, which guarantees a true n-body interaction. $T_{m_2(y)} \left(s_y^{m_1(y)} \right)$ is a Chebyshev polynomial of order $m_2(y)$ that depends on pair distance $s_y^{m_1(y)}$ for pair y of atom types $m_1(y)$ that has been transformed from r_y, to ensure it exists in the $[-1, 1]$ domain over which Chebyshev polynomials are defined, and $f_{\mathrm{s}}^{m_1(y)} \left(r_y \right)$ is a cutoff function that ensures smooth behavior at the outer cutoff. In essence, this generating function (Eq. 2) indicates that ChIMES interactions are taken as the product of interactions for constituent atom pairs, which are constructed from Chebyshev polynomial series. For example, a two-body interaction is simply taken as:

$$^{2}E_{ij} \propto \sum_{\alpha=1}^{\mathcal{O}_{2\mathrm{B}}} c_{\alpha}^{e_i e_j} T_{\alpha} \left(s_{ij}^{e_i e_j} \right). \tag{3}$$

ChIMES parameters are fit by force- (and optionally stress- and energy-) matching to configurations arising from DFT molecular dynamics simulations [20]. We note that ChIMES force, stres, and energy evaluation routines have been checked for translation, rotational, and permutational invariance and that the ChIMES molecular dynamics code and force matching codes are kept in a version controlled repository with an integrated regression suite individually verifying 2, 3, and 4 body interactions. As discussed below, the training set in the present study contains data at multiple thermodynamic states and for the gas-phase equilibrium products. We will consider an objective function which includes terms for per-atom forces, per-system-configuration energies, and per-system-configuration stresses. For the gas species, only per-atom forces and per-molecule-configuration energies are used in the fit. ChIMES models are generated by minimizing the following objective function:

$$F_{\mathrm{obj}} = \left(\sum_{i=1}^{n_{\mathrm{l}}} \left[\sum_{j=1}^{n_{\mathrm{a}}^{(i)}} \sum_{k=1}^{3} w_{\mathrm{F}_{ijk}}^2 \left(\Delta F_{ijk} \right)^2 + w_{\mathrm{E}_i}^2 \left(\Delta E_i \right)^2 + \sum_{l=1}^{n_{\mathrm{t}}} w_{\sigma_{il}}^2 \left(\Delta \sigma_{il} \right)^2 \right] + \sum_{i=1}^{n_{\mathrm{g}}} \left[\sum_{j=1}^{n_{\mathrm{a}}^{(i)}} \sum_{k=1}^{3} w_{\mathrm{F}_{ijk}}^2 \left(\Delta F_{ijk} \right)^2 + w_{\mathrm{E}_i}^2 \left(\Delta E_i \right)^2 \right] \right) / N_d, \tag{4}$$

where N_d is the total number of data entries, given by

$$N_d = n_{\mathrm{l}}(n_{\mathrm{t}} + 1) + n_{\mathrm{g}} + \sum_{i=1}^{n_{\mathrm{l}}} 3 n_{\mathrm{a}}^{(i)} + \sum_{i=1}^{n_{\mathrm{g}}} 3 n_{\mathrm{a}}^{(i)}, \tag{5}$$

where F_{obj} and $\{c\}$ are the weighted mean-squared error and model coefficients, respectively, the number of frames and atoms are given by n_{f} and n_{a}, respectively, and F_{ijk} indicates the kth Cartesian component of the force on atom j in configuration i. Since ChIMES is entirely linear in its fitted parameters, the model optimization problem can be recast as the following over-determined matrix equation:

$$\boldsymbol{w} \boldsymbol{M} \boldsymbol{c} = \boldsymbol{w} \boldsymbol{F}_{\mathrm{DFT}}, \tag{6}$$

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

$$M_{ab} = \frac{\partial X_{a,\mathrm{ChIMES}\{c\}}}{\partial c_b}. \tag{7}$$

In the above equation, a represents a combined index over force and energy components, and b is the index over model parameters.

In order to avoid overfitting during determination of model parameters, c, we use the least absolute shrinkage and selection operator [21] (LASSO) method of regularization. The LASSO method minimizes the objective function

$$F_{\mathrm{LASSO}} = F_{\mathrm{obj}} + 2\lambda \sum_{i=1}^{n_{\mathrm{p}}} |c_i|, \tag{8}$$

where n_{p} is the total number of unique fitting parameters. The parameter λ controls the magnitude of the fitting coefficients, which reduces possible overfitting. The absolute values (L1 norm) used in the LASSO objective function have the added benefit of setting certain coefficient values c_i to 0, which can lead to substantial gains in computational efficiency for simulations utilizing these models. We have developed software that implements the Least Angle Regression [22, 23] (LARS) algorithm for LASSO. This algorithm minimizes the LASSO equations in stages, where variables are added or removed one at a time; each stage in the LARS algorithm is a solution of LASSO optimization for a value of λ larger than the requested value. When the requested value is reached, the algorithm terminates. This approach has better convergence properties than direct optimization of F_{LASSO}, and additionally allows the analysis of solutions for each iteration. Moreover, our code distributes the $\boldsymbol{M}$ matrix between computing nodes and uses rank 1 Cholesky decomposition updates to solve linear equations that arise in LARS. This allows us to efficiently calculate LASSO solutions involving several thousand parameters, with very large training data sets (i.e. matrices $\boldsymbol{M}$ in excess of 1 TB are possible, as discussed in Section V B 1). We note that ChIMES software includes a comprehensive test suite to ensures backwards compatibility and reproducibility, and that wherever possible, ChIMES software has been validated against external codes (e.g. LAMMPS [24], scikit-learn [25], scipy [26], etc.)

ChIMES exhibits a number of unique features which make it particularly well suited for the study of energetic materials. In particular:

1. Interactions are generated *via* explicit many-body expansion using an orthogonal polynomial description, which facilitates description of chemistry in organic materials.

2. Models are parameterically linear, which allows rapid generation *via* advanced linear solvers (e.g., in contrast to neural network potential energy models requiring relatively slow non-linear approaches).

3. The ChIMES framework is versatile and can be leveraged as a standalone IAP, a correction to quantum or semi-empirical quantum approaches, or in the future a coarse graining method.

Moreover, ChIMES models can be built to arbitrary complexity - and thus arbitrary accuracy, in principle - by increasing model bodiedness and/or polynomial order for a given bodiedness. While increasing either of these model hyperparameters substantially increases the number of candidate fitted parameters, the LASSO regression method can substantially decrease the number of final (i.e. non-zero) parameters and thus minimize the subsequent increase in computational expense. Moreover, we have found that the percentage of non-zero parameters typically decreases as model complexity is increased, indicating improved accuracy for a given computational expense.

Since its initial development, the ChIMES approach has been applied systems of increasing complexity, including molten carbon [11], liquid water [12], liquid carbon monoxide [13], and Pu/H systems [14], and has been used to study problems including hydriding in Pu [14] and carbon coagulation in carbon monoxide under extreme conditions [27]. In the remaining sections, we describe extension of the ChIMES approach for development of HN_3 and DNTF models. We note that each of these materials present unique challenges (e.g. HN_3 transitions from insulating to metallic under shock compression [28], DNTF is the first solid molecular material investigated with ChIMES and contains the largest number of atom types to date, etc.), requiring significant methodological advances which will be highlighted below.

IV. ACHIEVING QUANTUM ACCURATE SIMULATIONS OF SHOCK COMPRESSED HN_3

(Additional details for this work are available in Ref. 29)

Hydrazoic acid is a highly sensitive liquid energetic material that can be viewed as a simple analog for more commonly used azides, such as lead- and sodium-azide. This material exhibits rapid detonation chemistry (i.e. equilibrated within 10 ps), is chemically simple (i.e. comprised of only two atom types and is carbon-free, limiting the chemical complexity of resulting species), and has been previously studied *via* DFTB [28] (i.e. semi-empirical quantum data validation data are available for comparison and to help guide thermodynamic state points explored during training data generation), making it an ideal testbed for application of ChIMES to EM. In this section, we describe development of a quantum-accurate ChIMES model for HN_3. Though a relatively simple material, we show that achieving quantum accuracy over the broad range of temperature and pressure conditions relevant to shockwave driven detonation (i.e. ambient to 1000s of K and 10s of GPa) necessitates complex models.

The steady-state detonation velocity is determined by a particular thermodynamic state, called the Chapman-Jouguet state, that lies on the shock Hugoniot of the reacted high explosive. Therefore, calculation of the Chapman-Jouguet state allows for the detonation velocity to be predicted from atomistic simulations. The Cheetah thermochemical code calculates the Chapman-Jouguet state through an entirely different statistical mechanical approach. Therefore, determining the Chapman-Jouguet state with ChIMES is of considerable utility in the further development of Cheetah. The MultiScale Shock Technique (MSST) [30] is a simulation method that varies the cell volume and

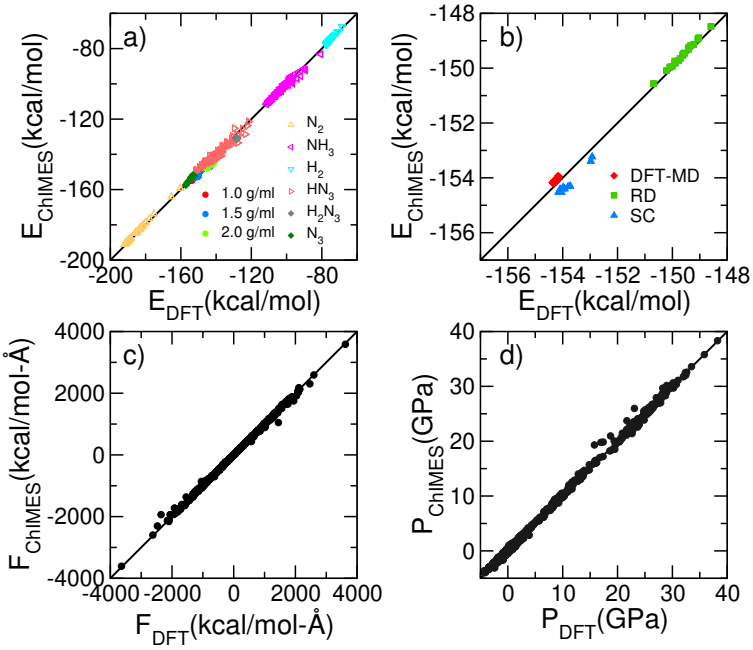

FIG. 1: Comparison of isolated molecule (a) and condensed phase (b) energies, atom forces (c), and pressures (d) predicted by DFT and the ChIMES HN_3 model. Labels "DFT-MD", "RD", and "SC" in panel b indicate data corresponding to training configurations generated *via* DFT-MD, random displacement, and self-consistent (i.e. iterative) refinement, respectively.

energy to obey the thermodynamic conditions imposed on a shocked material. The MSST can accelerate calculations many orders of magnitude over direct shock simulations, where a shock wave is driven by an atomic-scale impact event. Using ChIMES, we determined the Chapman-Jouguet state and related chemistry from MSST shock simulations for HN_3, and conducted preliminary DS simulations of bubble collapse in HN_3. These results represent the first ever shock simulations using ChIMES models.

A. ChIMES HN_3 Model Development and Validation

A DFT training database was generated by running 10 ps DFT-MD at (1.0, 300), (1.5, 4500), and (2.0, 4500) (g cm^{-3}/K)(i.e. previously determined relevant to shockwave-driven detonation of HN_3). An initial ChIMES model with explicitly two-, three-, and four-body interactions described through polynomials of respective orders $O_{2B}/O_{3B}/O_{4B} = 12, 7, 3$ was fit to 60 frames (i.e. system configurations) taken from this database. Consistent with previous ChIMES studies [11–13], resulting models were found unstable when leveraged in ChIMES-MD simulations. As will be discussed in detail in Section V A, this is due to inadequate sampling of model space during the course of the relatively short DFT-MD simulations, and can be efficiently remedied by including iteratively-generated configurations in the training set (e.g. taking configurations from ChIMES-MD simulations and assigning the corresponding forces/stresses, and energies through single point DFT calculations). A total of 22 such configurations were utilized for HN_3 model development. An additional, 60 frames were included, generated *via* random displacement of DFT-generated configurations to improve sampling of model space. Finally, approximately 400 configurations of isolated species (i.e. HN_3, N_2, NH_3, and H_2) were included to improve model description of their relative energetics. The final fitting problem culminated in a 8.8 GB design matrix for a model containing 3980 parameters. Using LASSO to select parameters, our final HN_3 model has only 1324 non-zero (out of 3980 possible) parameters. This is a substantial gain in computational efficiency over using all model parameters. In the following sections, we will refer to this method of parameterization as the "non-active" method, since active machine learning is not used to select configurations. The active machine learning algorithm was developed after most of the reported work on HN_3.

Fig. 1 shows that our HN_3 ChIMES model yields excellent agreement with DFT in terms of predicted forces, stresses, and energies, for training data. However, ML-IAPs are susceptible to serendipitous agreement and should always also be validated in terms of physical properties, e.g. system structure as measured by the radial pair distribution functions (Fig. 2, which provides the likelihood that an atom will be found at a given distance from another atom) and chemistry (Fig. 3). We note that HN_3 is known to form charged species (e.g. N_3, H_2N_3, and NH_4) during detonation [28]. We

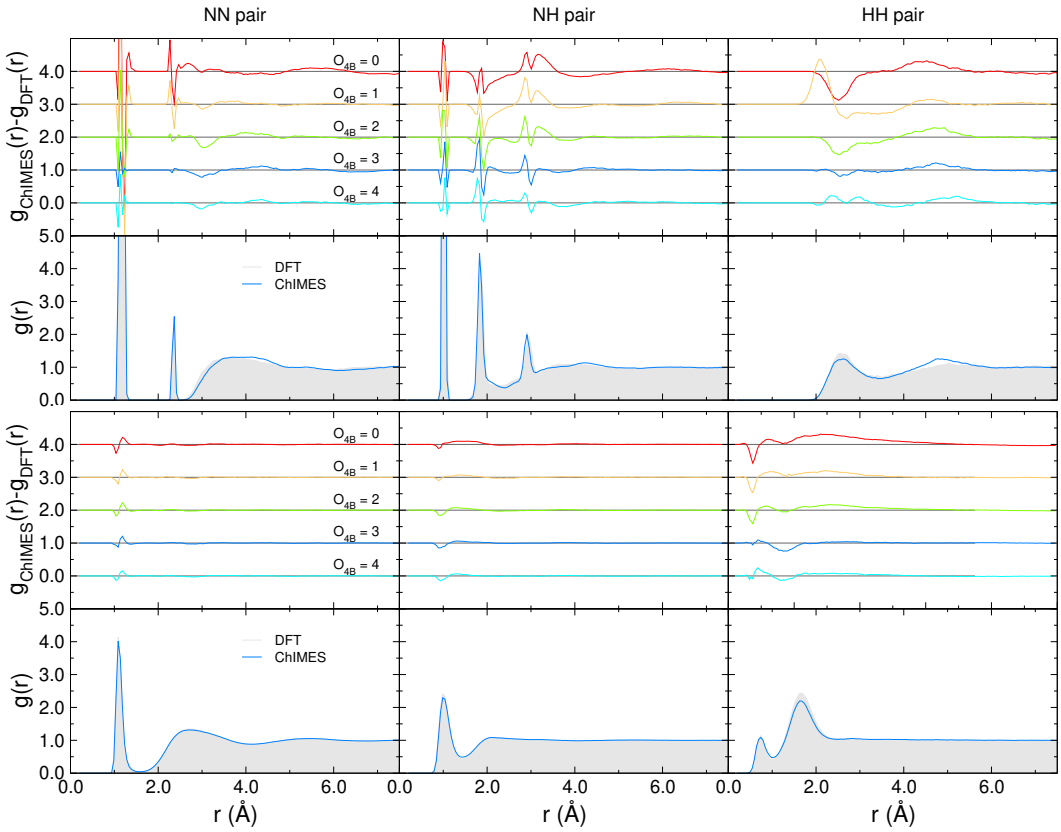

FIG. 2: Radial pair distribution functions (RDFs) predicted by DFT and ChIMES MD using the $O_{2B}/O_{3B}/O_{4B} = 12, 7, 3$ model at 1.0 g cm^{-3} and 300 K (top) and 2.0 g cm^{-3} and 4500 K (bottom). Differences between RDFs predicted using DFT and ChIMES models with differing 4-body polynomial orders are given in multicolored lines above each RDF, offset from each other for clarity. An improvement in the match of ChIMES MD to DFT is usually seen as the 4-body order is increased.

find that the HN_3 model developed in the present work is able to accurately recover configurations containing these species despite not explicitly resolving electrons or atom charge, suggesting that the charge state of any given species formed in detonating HN_3 is uniquely related to its relative atomic positions. Nevertheless, efforts are underway to allow explicit charge transfer in ChIMES models, for cases where species where relative atom coordinates do not uniquely determine species charge state. We find that the DFT-predicted system structure is well recovered for our final $O_{2B}/O_{3B}/O_{4B} = 12, 7, 3$ model, but that at both the ambient and dissociative state points, 4-body interactions (i.e. $O_{4B} \geq 3$) are needed to correctly recover the HH pair distribution functions. Inclusion of 4-body interactions is also found to systematically increase accuracy of chemistry predictions, most notably for species lifetimes at 1.5 g cm^{-3} and 4500 K. Overall, we find that HN_3 models exhibiting quantum accuracy in properties relevant to modeling shock induced detonation at a fraction of the computational expense can be developed as straightforward application of the ChIMES model and basic fitting framework.

B. Predicting the HN_3 Detonation State and Chemical Kinetics

We have used the machine learned ChIMES HN_3 model described above to determine Chapman-Jouguet (C-J) detonation velocity, investigate chemistry on timescales an order-of-magnitude greater than has previously been studied *via* atomistic methods, and for a preliminary investigation of how the presence of voids (i.e. bubbles) influence evolution of the shock-compressed material. The two former studies are conducted through multiscale shock technique [30] (MSST) simulations, which constrains the thermodynamic states of a MD simulation to those for a steady planar shockwave within continuum theory. MSST simulates steady shock waves by constraining the stress and energy of a MD simulation to the Rayleigh line and the Hugoniot energy relations. MSST allows shock simulations with smaller systems and lower computational cost than direct-shock simulations, (i.e. which can require orders-of-magnitude more atoms) and has been used to accurately reproduce the shock Hugoniot of many materials [31–33].

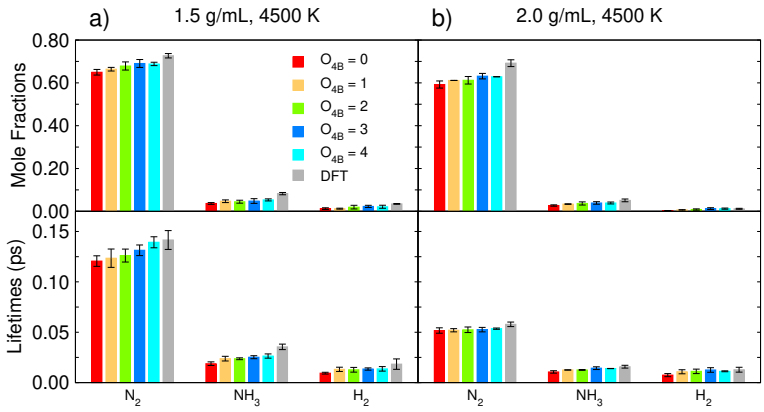

FIG. 3: Mole fractions and corresponding lifetimes for dominant small molecules predicted by DFT and ChIMES models with differing 4-body polynomial orders. Results at 1.5 g/mL improve more from a higher 4-body polynomial order than those at 2.0 g/mL.

Fig. 4 shows evolution of system density, temperature, and pressure during the the first 50 ps of a MSST simulation. In all cases, the system is compressed rapidly and then slowly expands and increases in temperature as the system reacts. The C-J point (i.e. the speed with which a steady state unsupported shockwave propagates through the material) can be taken as the as the lowest shock velocity for which a steady state solution exists (i.e. at which the MSST simulation does not diverge). Here, we find the C-J detonation velocity to fall between 6.5 and 6.6 km/s, which is closer to the experimentally reported [34, 35] range of 7.1 – 7.6 km/s than the DFTB predicted value [28] (6.0 km/s). The C-J pressure and temperature predicted here are approximately 14 GPa and 2900 K, respectively

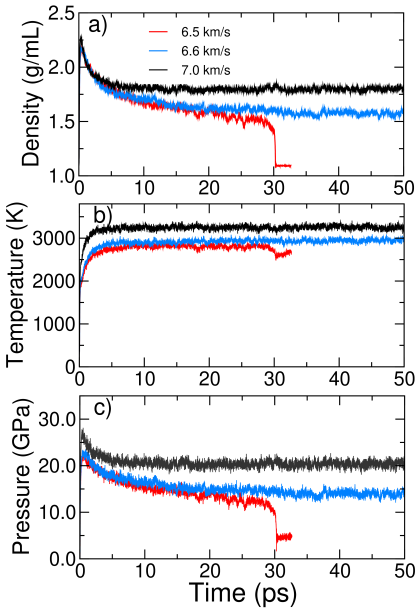

FIG. 4: Time evolution of system density (a), temperature (b), and pressure (c) during HN_3 MSST simulations. Corresponding shock velocities are given in the legend. The MSST simulations indicate that the Chapman-Jouguet detonation velocity is close to 6.6 km/s, which is the slowest shock speed that does not lead to a divergence in the density.

Fig. 5 shows how system chemistry evolves during the MSST simulation, which was run for a total of 1 ns (i.e. 2 orders of magnitude longer than possible with DFT). An earlier DFTB study [28] suggested HN_3 chemistry is equilibrated within 10 ps of shock compression, but we find that chemistry is still evolving after 50 ps. We note that the present 2048 atom 1 ns ChIMES simulation required a total of 288 CPU $\times$ 168 hrs = 4.8×10^4 CPU—hrs compared to DFT which was estimated to require roughly 5 orders of magnitude more CPU hrs.

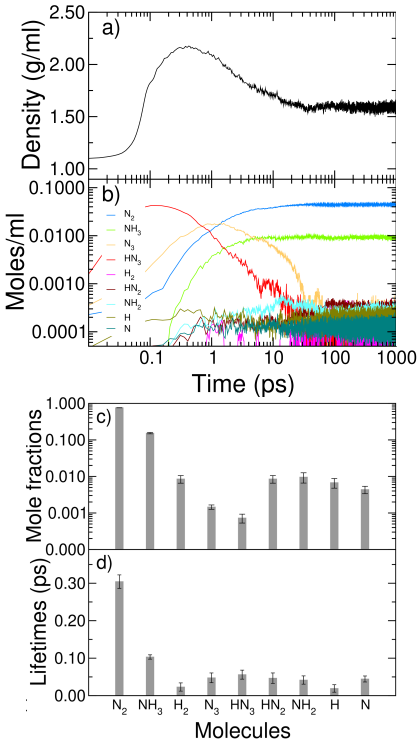

FIG. 5: Time evolution of system density (a) and mole fractions (b) during 1 ns long HN_3 MSST trajectories at 6.6 km/s, and chemistry (i.e. mole fractions (c) and lifetimes (d)) following equilibration in the shocked state.

The computational efficiency of ChIMES scales linearly with increasing system size, which also allows us to run much larger simulations. Here, we provide preliminary results for a 100,000 atom DS simulation with the presently developed ChIMES model. DS simulations involve driving a sample of material into a stationary piston at a given particle velocity and observing how the system evolves as the resulting shockwave passes through the material. Thus overall simulation length is set by the time required for the shockwave to traverse the material, (i.e., simulation length is determined by system size along the shock direction). Though this approach is generally more computationally demanding than MSST (i.e. typically requiring orders-of-magnitude more atoms), it is a powerful method for observing phenomena like defect-driven hotspot formation [9, 10] which require explicit resolution of the shock front and the ability to observe system evolution behind it. The computational efficiency exhibited by ChIMES for a fixed system size combined with its linear scalability with increasing system size make it ideal for use in DS simulations. Here, this is demonstrated through preliminary results from DS simulations of bubble collapse in HN_3.

Fig. 6 shows evolution of system density as a shockwave passes through the material. Unlike recently reported simulations of void collapse in crystalline EM [9, 10], no jetting is observed, likely due to low material strength (viscosity). Upon bubble collapse, density at the void location exceeds that of both the shocked an unshocked material. As the collapsed pore relaxes, rarefaction waves can be observed, momentarily creating a region of density that is once again lower than that of the surrounding materials.

These preliminary results indicate that the presence of voids significantly disturbs evolution of the shocked system, which is consistent with classic studies showing that liquid explosives can be sensitized through introduction of gas bubbles [36]. Future work will investigate whether detonation can be observed in weakly shocked HN_3 containing voids, and whether sensitization varies when voids are filled with various gases. Overall, the quantum-accuracy, transferability, and robustness demonstrated by the models leveraged in this section indicate that ChIMES is paving a path toward significantly improved simulation of EM.

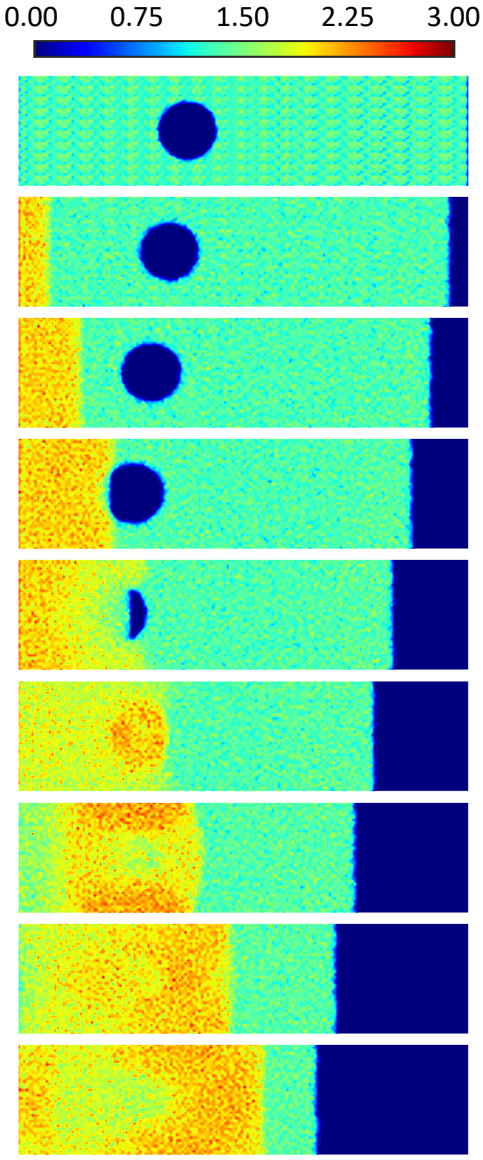

FIG. 6: Time evolution (i.e. from top to bottom in 2.5 ps increments) of a 100,000 atom HN_3 system containing a cylindrical void shocked *via* impact with a piston to the left of the system. System density is indicated in g cm^{-3} by colors in the color bar.

V. MODELING HIGHER COMPLEXITY EXPLOSIVES

The previous section demonstrated that a quantum-accurate ML-IAP could be generated for HN_3, a simple EM, through straightforward application of the non-active ChIMES fitting procedure. However, in this section we will show that generating models for the larger carbon-containing EM (i.e. similar to TATB or HMX) requires a significantly different fitting strategy. This necessity arises from the need for higher complexity models (i.e. because the number of candidate parameters for a model of fixed given bodiedness and polynomial orders increases with increased unique atom pair types); ML-IAP quality is inextricably linked to the data it was trained to and thus those fit for EM are highly susceptible to overfitting due to their high complexity. Ultimately, if the training data does not adequately sample model space (which may not directly map onto physicochemical space), the resulting model will be inaccurate at best, and unstable at worst.

Recalling that ML-IAP training data is typically generated from DFT-MD simulations, this is a particularly significant challenge for molecular materials (e.g. EM). This is because the finite time and length scales of DFT-MD simulations are such that the resulting trajectories primarily correspond to favorable regions of physicochemical space, and generally does not naturally sample the unfavorable regions that would otherwise inform the repulsive portion of the ML-IAP, leading to models that are inaccurate for rare events. Larger unreacted EM consist of clusters of atoms with strongly preferred orientations, meaning a significantly larger portion of model space will remain unsampled in a DFT-MD generated trajectory than a non-molecular material (e.g. a metal), which is typically of uniform atomic density throughout. Even when reaction occurs, sampling in a DFT-MD trajectory will be limited by the accessible timescales in DFT-MD. The practical implication is that effective ML-IAP development for EM requires intelligent training repository generation and maintenance approaches. To add to these challenges, the range of thermodynamic conditions relevant to a given EM is not always known *a priori*, and DFT-based exploration of this space (i.e. to guide efforts for ML-IAP development) is prohibitive.

In this section, we present new tools to overcome these challenges. Specifically, a new active learning framework, a multiscale ChIMES model development approach, and preliminary results for DNTF are presented. The simulations resulting from this study can be immediately used to improve thermochemical models in Cheetah and help resolve questions related to graphitic nano-onion formation from DNTF under shock compression [37]. In the following sections, we present a new active learning python tool for generation of robust, transferable, and reproducible ChIMES models for complicated chemical systems, preliminary results from application to DNTF, a ChIMES multiscale fitting framework for new EM, and insights arising from those studies.

A. Automating Development of ChIMES Models for Challenging Systems

(Additional details for this work are available in Ref. 38)

ML has gained significant attention in the IAP development community due to its versatility and ability to enable quantum-accurate simulation at a fraction of the computational cost. Unfortunately, the strength of ML-IAPs (i.e. their versatility) is also the source of their primary weakness: models can be enormously complex (e.g. $> 10^5$ parameters) and thus highly susceptible to overfitting. For this reason, generating training data that ensures resulting models are robust, accurate, and reproducible remains a central challenge of ML-IAP development [19, 39–42]. The iterative refinement framework described in the previous section helps overcome this challenge to an extent, and can be viewed as a kind of "enhanced sampling" technique due to the likelihood that early (e.g. poor) models will explore regions of model space unsampled by the original DFT-MD, but becomes impractical for complex molecular materials, since in practice it can take very many iterations to generate a stable model. Active learning [43], which centers around the idea of an automated machine learning algorithm to inspect unlabeled data (i.e. coordinates without DFT forces assigned), identify those that are expected to improve model performance, request labels, and add data to the training set, provides a promising alternative approach. However, the effectiveness and efficiency of a given AL framework relies on careful consideration of the target model and its implementation.

In this section, we present development of an automated active learning tool enabling reproducible development of robust parameterically linear ML-IAPs (e.g. ChIMES). We use a reactive C/O system under extreme T and p (i.e. 2400 K and 1.79 g cm^{-3}) as a testbed for this method, i.e. a system for which the non-active ChIMES parameterization approach has been shown to fail [13]. We note that a following section (i.e. Section V B) will present preliminary results arising from application of this active learning approach for development of a T/p transferable ChIMES model for DNTF.

1. The Non-Adaptive ChIMES Fitting Approach and its Shortcomings

The standard non-adaptive ChIMES iterative fitting approach is given below. We note that the success of this method hinges upon the notion that the relatively inaccurate models produced during early iterations are more likely to reach unsampled regions of physicochemical space and can thus be considered a means of rare-event sampling. As briefly mentioned above, this approach has been successful for systems of non-reactive small molecules, or those exhibiting rapid chemistry [11–13]. To generate these models, a simple iterative refinement framework was employed, where (i) training trajectories were obtained from short DFT-MD trajectories at the state points of interest for the system, (ii) a model was obtained by minimizing the objective function (i.e. Eq. 4), (iii) a MD trajectory was launched using this i^{th} force field, (iv) configurations from this simulation were periodically sent to DFT for single point calculation to be merged with the existing repository, and (v) steps ii through iv were repeated until the model exhibited the target level of accuracy.

The ChIMES model development scheme described above, (i.e. iterative fitting), has been found insufficient when at least one of the following two conditions are met [13]: the system undergoes relatively slow chemistry, with species lifetimes exceeding 1 ps (i.e. where kinetics are limited by relatively large reaction barriers), or the system contains structurally complex species, e.g. oligomers, heterocycles, fused rings, etc. where greater-than 3-body interactions (e.g. intramolecular torsion) have been shown [13] to play a significant role. As an example of this, we consider a system of 50/50% C/O at 2400 K at 1.79 g cm^{-3}. As shown in Fig. 7a and 7c, DFT simulations predict that both criteria are met for the above system. Previously, a ChIMES model was developed with the intention of transferability over $T = 9350$, 6500, and 2400 K, and $\rho = 2.56$, 2.5, and 1.79 g cm^{-3} [13]. This model was fit through the standard (i.e. non-active) iterative approach using 2+3-body interactions and was found to work well between the 9350 and 6500 K state points, but failed at 2400 K. As indicated in Fig. 7b and 7c, one of two structural themes emerged during ChIMES MD simulations at 2400 K using models derived from successive iterative iterations; either exclusively small linear species, or unphysical ring-like structures featuring highly coordinated oxygen. These two outcomes arose from an insufficiently complex ChIMES interaction model; to a model containing only 2- and 3-body interactions, the coordinated oxygen structure is reasonable, containing bond and angle distances reminiscent of those found in carbon dioxide, ethylene oxide, dimethyl ether, etc. However, when iteratively added to the training repository, DFT assigns a high energy, resulting in subsequent models that bias away for cyclic structures. As a consequence, predicted chemistry is far off from the DFT predictions.

The above issues can be resolved through addition of 4-body terms, which in the language of molecular mechanics force fields would allow for description of bonds (2-body), angles (3-body) and dihedrals/impropers (4-body). However, doing so substantially increases the number of parameters considered in the fitting process. For example, in a system with two atom types, polynomial orders of $\mathcal{O}_{2\text{B}}/\mathcal{O}_{3\text{B}}/\mathcal{O}_{4\text{B}} = 12/7/0$ yields a maximum of 806 parameters, whereas $\mathcal{O}_{2\text{B}}/\mathcal{O}_{3\text{B}}/\mathcal{O}_{4\text{B}} = 12/7/3$ (i.e. the polynomial orders used in the present work), yields a maximum of 3978 parameters. Increasing model complexity gives rise to an additional set of problems; beyond the concomitant increase in risk of over-fitting, a far greater number of iterative cycles are required to generate a converged force field, which is prohibitive for reactive carbon-containing systems (i.e. that exhibit significantly higher chemical complexity than those of only O, H, and/or N). In the following sections we describe strategies to overcome these challenges, through development of an active learning framework combining cluster analysis and concepts from Shannon [44] information theory.

2. The ChIMES Active Learning Framework

Development of the ChIMES active learning framework began with examination of our standard fitting approach. ChIMES training repositories contain both DFT- and iteratively-obtained frames from MD trajectories, comprised of $3n_{\text{a}}$ coordinates and forces, and a single overall system energy. The convoluted nature of this fitting problem becomes apparent if one considers fitting a $\mathcal{O}_{2\text{B}}/\mathcal{O}_{3\text{B}}/\mathcal{O}_{4\text{B}} = 12/7/3$ force field for a C/O system to only energies from DFT; doing so requires assignment of 3978 parameters such that the energies for each dimer, trimer, and tetramer sum to the single value for each frame. The total energy in a given frame arises from the sum of bonded and non-bonded interactions, inter- and intra-molecular interactions, etc. Thus, by decomposing a given frame into a collection of individual atom clusters (i.e. nominal molecules), computing the corresponding DFT forces and energies, and adding them to the training repositories, resulting fits contain greater information on how to assign the 10s to 1000s of parameters giving rise to different 2-, 3-, and 4-body energetic contributions. This concept is similar to generating training configurations from a direct cluster expansion, but can be more effective for molecular systems because resulting configurations are relevant to both the model and the target physicochemical space.

Note that the ChIMES model space can allow formation of unphysical many-body cluster configurations (e.g. 3 atoms forming an equilateral triangle with all distances equal to the corresponding 3-body inner cutoff). Thus, it is important that the training repository contain configurations of this nature to inform the fit of their unfavorability

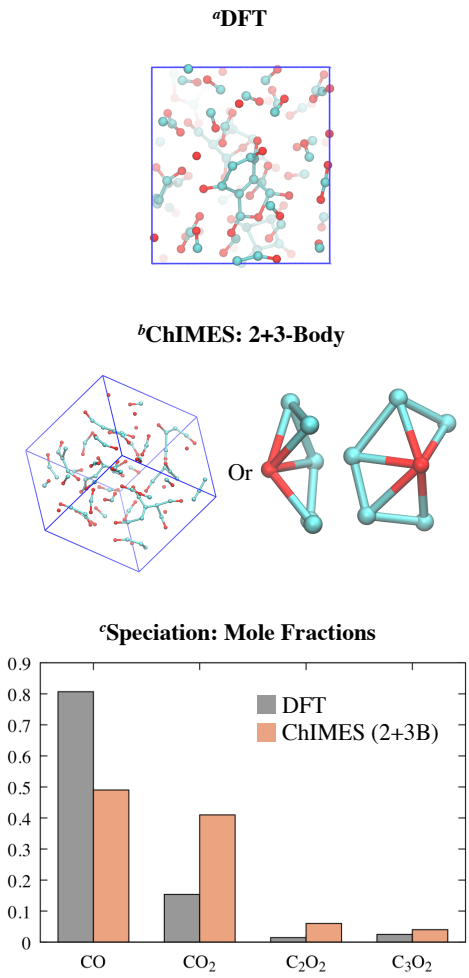

FIG. 7: Sample system configuration predicted by (a) DFT and (b) successive, iteratively generated $2+3-$body ChIMES force fields [13] for carbon monoxide at 2400 K and 1.79 g/cm^3; (c) comparison of corresponding mole fractions predicted by DFT and the 2-3-body ChIMES model.

and prevent their spurious appearance during MD simulations. However, this approach is still highly inefficient for several reasons:

1. Successive MD frames are highly correlated (e.g. due to the limited exploration of physicochemical space in short-time MD simulations at a fixed temperature(s)).

2. The species contained in each frame can be very similar (e.g. distributed about some chemical and/or conformational minimum energy).

3. The computational cost of evaluating the possible tens of thousands of species *via* DFT would be prohibitive from a practical standpoint.

In order to increase the efficiency of our fits, we aim to increase the information contained in the training repository while simultaneously maintaining a minimal size, by developing a method for selecting subsets of possible species. We do so by defining a "feature" for each species, simply taken to be $E_{\mathrm{ChIMES},i}$, the energy per species atom as computed by the ith ChIMES force field. According to Shannon information theory, the information contained in any given subset of species is maximized when the corresponding probability distribution of $E_{\mathrm{ChIMES},i}$ values is flattened. These concepts can be combined with the standard ChIMES iterative fitting procedure to form the active learning cycle shown in Fig. 8. We note the distinction between active learning ("AL"), represented by the orange components of the cycle in Fig. 8, and an active learning cycle ("ALC"), which involves *both* iterative refinement and active learning, i.e. the blue *and* orange components in Fig. 8.

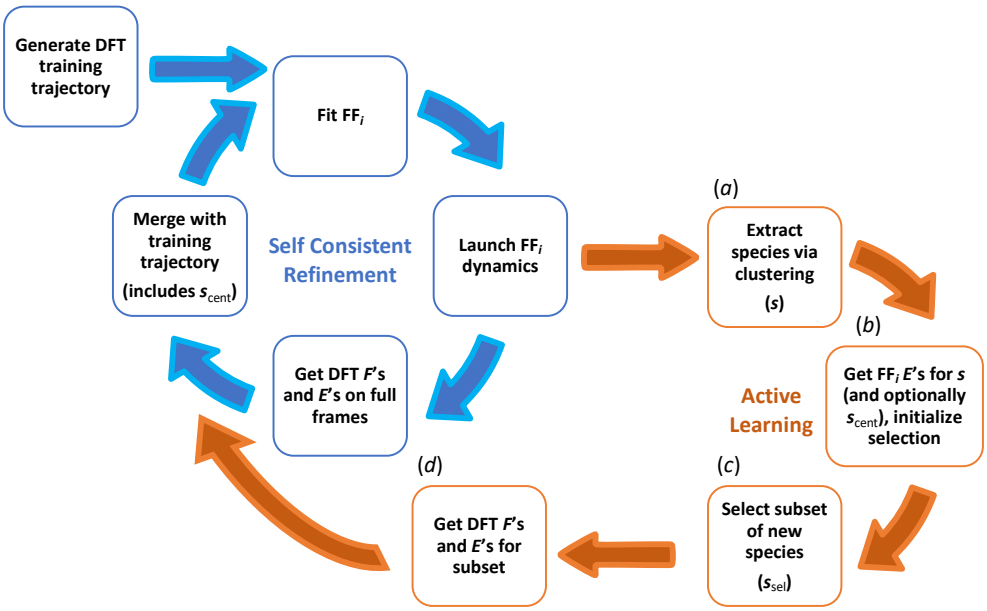

FIG. 8: The ChIMES active learning framework.

Fig. 9 provides an overview of how the AL process works in practice. First (a), atom clusters are extracted from ChIMES-MD trajectory frames. To seed selection of a subset of these clusters (b), i.e. for addition to the central training repository), these clusters are split to initialize a selection (s_{sel}) and a pool of remaining clusters (s_{pool}). A Monte Carlo scheme is then used to exchange clusters between s_{sel} and s_{pool} such that the probability histogram of cluster energies in s_{sel} is flattened, yielding a highly informative subset. Finally, single point DFT calculations are run on each of the selected clusters for assignment of force, stress, and energy, and the configurations are added to the training repository.

3. Sample Application: A C/O-Containing System Under Extreme Conditions

The 1:1 C/O-containing system at 2400 K and 1.79 g/cm^{-3} described in Section V A 1 was used as a testbed for development of the AL framework described above. Our goal was to show that the automated framework could be applied to develop a robust C/O model exhibiting improved description of system structure and chemistry (i.e. compared to the C/O model fit through the standard ChIMES approach shown in Fig. 7). Fig. 10 shows how prediction of small molecule chemistry (e.g. mole fractions and corresponding lifetimes) is improved with successive iterations of the AL cycle shown in Fig. 8. Notably, AL predictions are found to converge during late ALC, indicating resulting models are stable and consistent, and that the framework is generating reproducible results. As is shown in Fig. 11, an additional benefit of the ChIMES AL approach is that it allows accurate prediction of conformational energetics for large molecular species, which is important for accurate simulation of more complicated materials (e.g. TATB, 2,6-diamino-3,5-dinitropyrazine-1-oxide (LLM-105), etc). We find that models produced through the ChIMES AL framework yield quantum accurate predictions of condensed-phase system structure, as shown in Fig. 12.

Overall, the ChIMES AL framework represents a significant advance in ChIMES methodology and helps to overcome a challenge central to ML-IAP development in general (i.e. producing training sets enabling generation of robust, accurate, and reproducible ML-IAPs). Moreover, this framework has positioned us to more effectively generate models for high explosive materials (i.e. which are exhibit significantly greater structural and chemical complexity to any material studied with ChIMES to date). We note that this fully automated ChIMES AL framework has been implemented as a modular (i.e. easily extensible) python package.

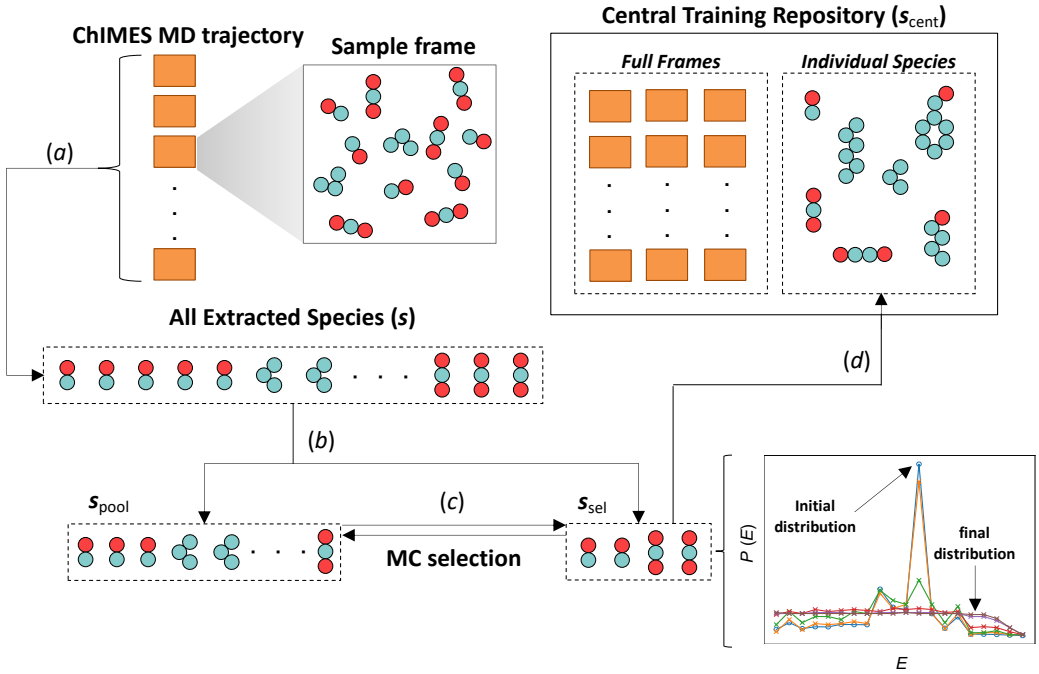

FIG. 9: Schematic representation of the active learning (i.e. orange) portion of an active learning cycle (i.e Fig. 8). Arrow a indicates species extraction *via* clustering and per-cluster energy assignment by the ith ChIMES force field. Arrow b indicates randomized initialization of the cluster selection process by splitting all extracted species into an initial selection subset and candidate pool. The c arrows indicate exchange of species between the pool and selection subset *via* MC (i.e. flattening the histogram of selected cluster energies), and arrow d indicates single point DFT evaluation for each selected cluster and subsequent addition to the ever-growing central repository.

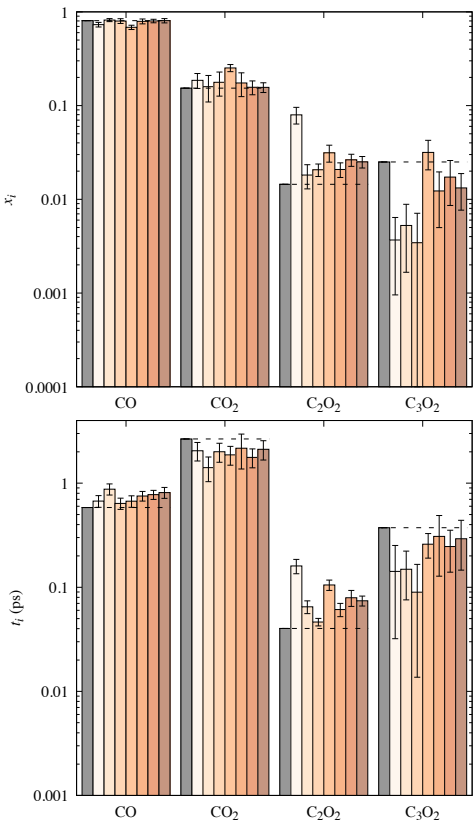

FIG. 10: Small molecule chemistry predicted by DFT (gray) and ChIMES models generated through successive iterations of the active learning framework (progressively darker orange bars) during simulations of C/O at 2400 K and 1.79 g cm^{-3}. Horizontal dashed lines serve as a guide to the eye providing DFT predictions for a given species.

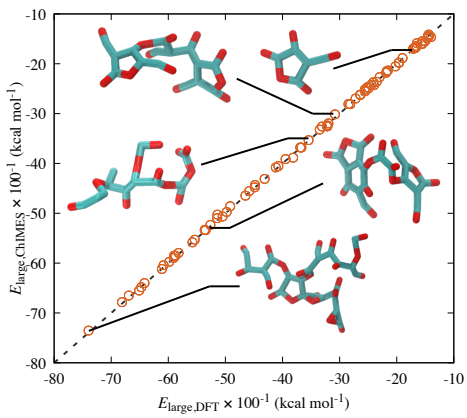

FIG. 11: Comparison of energies predicted by DFT and an actively learned ChIMES C/O model for large species formed during DFT-MD simulations of C/O at 2400 K and 1.79 g cm^{-3}.

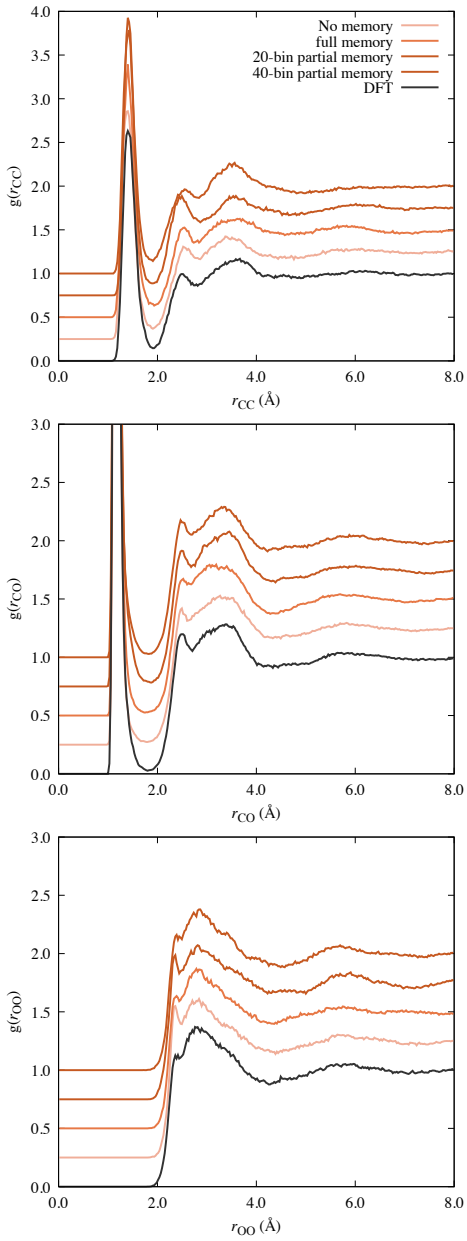

FIG. 12: Radial pair distribution functions predicted by DFT and actively learned ChIMES models varying by active learning driver user settings, for simulations of C/O at 2400 K and 1.79 g cm^{-3}. ChIMES predictions have been vertically offset for clarity.

TABLE I: Thermodynamic state points for DFT simulations and Hugoniot EOS predictions from the Cheetah [2] thermochemical package. $t_{\mathrm{sim,DFT}}$ is the total DFT simulation length. $X_{\mathrm{full,EOS}}$, $X_{\mathrm{part,EOS}}$, and $X_{\mathrm{un,EOS}}$ are properties (e.g. temperature, T, or pressure, p) predicted by Cheetah for the fully reacted, partially reacted, and unreacted DNTF systems. In this context, fully reacted corresponds to a system that has completely decomposed into small molecules with low heats of formation and carbon condensates, while partially reacted corresponds to exclusion of carbon condensates as a decomposition products.

ρ	T_{DFT}	P_{DFT}	$t_{\mathrm{sim,DFT}}$	$T_{\mathrm{full,EOS}}$	$P_{\mathrm{full,EOS}}$	$T_{\mathrm{part,EOS}}$	$P_{\mathrm{part,EOS}}$	$T_{\mathrm{un,EOS}}$	$P_{\mathrm{un,EOS}}$
g cm^{-3}	K	GPa	ps	K	GPa	K	GPa	K	GPa
1.86	300	0.03	5.00	4220	19.7	N/A	N/A	300	0.03
1.80	383	0.31	5.00						
2.00	2000	5.50	5.00						
2.00	4250	13.6	2.50	4250	20.7	4000	22	310	0.09
2.25	4440	21.0	2.50	4440	28.3	4000	32	360	4.25
2.50	4700	28.5	2.50	4700	39.0	4180	46	520	10.0
2.50	9000	35.6	0.25						

B. Generating Improved Reactive Models for DNTF

A central goal of this work was positioning ChIMES for agile development of quantum-accurate models for programatically relevant EM to enable prediction of material thermochemistry, shock Hugoniot equation of state, and to investigate phenomena including carbon coagulation and hot spot initiation. To do so, recovery of DFT-predicted structure, dynamics, and speciation across the range of temperatures and pressures representative of analogous material in both its common state and shocked states must be demonstrated. In the following sections, we describe our efforts toward this goal for DNTF. We describe initial training set generation, preliminary results arising from use of the AL driver for model generation, and compare preliminary model performance to a popular reactive model commonly used for simulations of EM (i.e. ReaxFF). Challenges in training set and model generation for complicated EM are discussed, as are the new approaches and new ChIMES active learning features established to overcome them.

1. Model Development Overview

To begin development of a ChIMES model for DNTF, a DFT training database was needed. In contrast to the HN_3 study discussed in Section IV, there are no previous shock simulation studies for this material and experimental data are scarce. Because our target is a model capable of describing chemistry in the shock compressed material, it was critical that the training set contain data at relevant thermodynamic conditions. The thermochemical code Cheetah [2] was used to predict a fully-reacted shock Hugoniot for DNTF up to a pressure of approximately 40 GPa, i.e. encompassing the range of conditions under which experiments predict carbon condensation occur. Three state points for generation of DFT-MD training data were identified from this Hugoniot. Three additional state points were utilized for training data generation, corresponding to the material under ambient conditions, at its melting point, and above its thermal decomposition temperature [45]. Finally, a high temperature training point was added to ensure adequate sampling of close interatomic contacts. These state points, along with the DFT-MD simulation length are given in Table I. We note that all of these simulations were initialized as the crystalline material so the resulting trajectory contained information on decomposition of the DNTF molecule and that the DFT simulations are short and have not necessarily reached a metastable reacted state within the explored simulation time. We note, however, that an additional benefit of iterative ChIMES fitting (e.g. 8) is sampling of - and thus generation of training data for - system evolution on timescales beyond that of the DFT training set.

The ChIMES AL-driver, which was extended to enable "parallel learning" (i.e. at multiple state points simultaneously) and include support for spin-polarized gas-phase calculations, was then used to generate DNTF models. We note that the latter extension was necessary for accurate description of DNTF decomposition species such as NO and NO_2. The present ChIMES models (which contain 3 unique atom types) included explicit 2-, 3-, and 4-body interactions described with respective polynomial orders of 12, 7, and 3, and contained 17,416 candidate parameters. Parameter determination required solution of up to 2 TB design matrices, made possible through development of the distributed solver described in Section IV A. Matrix solution requires approximately 24 hours of wall time on 12 36-core Quartz nodes, and a single AL cycle takes approximately 1.5 days of constant run time (i.e. not accounting for queueing). Several ChIMES test models were generated to establish the influence of hyperparameter (i.e. non-linear

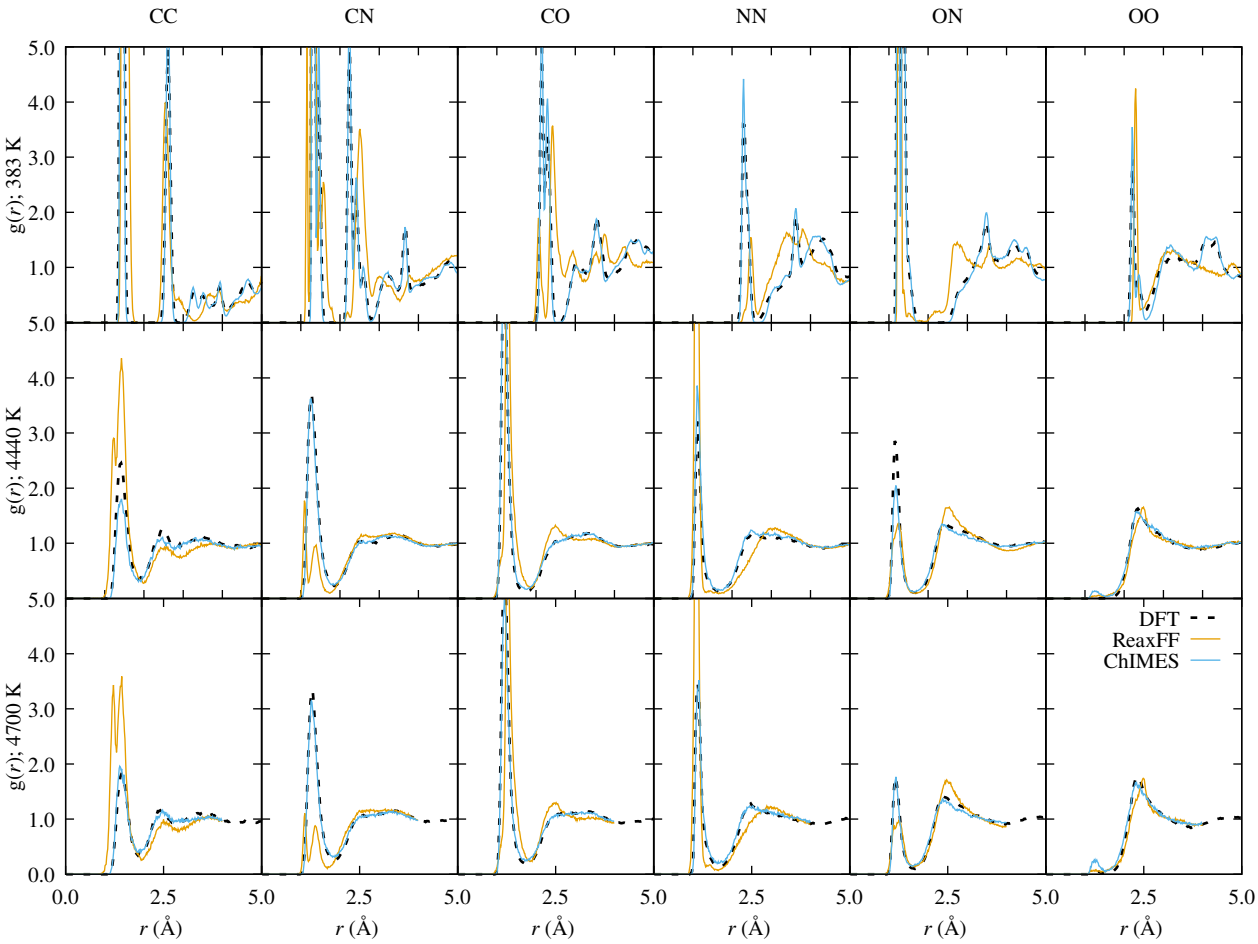

FIG. 13: Radial pair distribution functions for DNTF predicted by DFT, an existing DNTF reactive IAP (ReaxFF [46–49]), and the ChIMES IAP at select thermodynamic states spanning ambient shock compressed conditions.

model parameters specified by the user) choice on model performance, and test models were considered complete after 10 to 15 cycles.

2. *Preliminary Results*

Performance of the latest DNTF test model is shown here, in terms of the predicted EOS, small molecule speciation, and RDFs and compared to predictions from DFT, and ReaxFF [46–49]. Beginning with Fig. 13, we find that the ChIMES model is in good agreement with DFT, particularly relative to the ReaxFF predictions. We note that the ReaxFF model is highly reactive and was found to predict chemistry in DNTF under ambient conditions (e.g. ring-opening through N–O dissociation), which contributes significantly to poor recovery of the DFT RDFs at low temperature. Mole fractions and corresponding lifetimes for dominant small molecules (i.e. CO, CO_2, N_2, NO, N_2O, and O_2) predicted by ChIMES and ReaxFF are given in Fig. 14, relative to DFT. Both ReaxFF and ChIMES are yield excellent agreement with DFT for mole fractions of less than 0.2, but deviate non-negligibly at intermediate mole fractions (i.e. 0.2 to 0.7), though the ChIMES predictions are generally in closer agreement with DFT. The models yield similar performance for small molecule lifetime predictions, exhibiting good agreement with DFT for $t < 0.35$ ps and non-negligible deviations at larger t, though ChIMES values remain closer to the DFT prediction. As will be shown in Section V C 3, significant statistical uncertainty is expected in chemistry predictions, i.e. on the order of 0.15 - 0.20 in mole fractions and 0.25 - 0.5 ps in lifetimes.

Fig. 15 compares pressures predicted by the ChIMES test model, ReaxFF, and DFT at training thermodynamic state points. We find that the current ChIMES test model yields better agreement with DFT at lower temperatures and pressures (e.g. below 4440 K and 2.25 g/ cm^{-}3), while the ReaxFF model performs best at higher temperatures

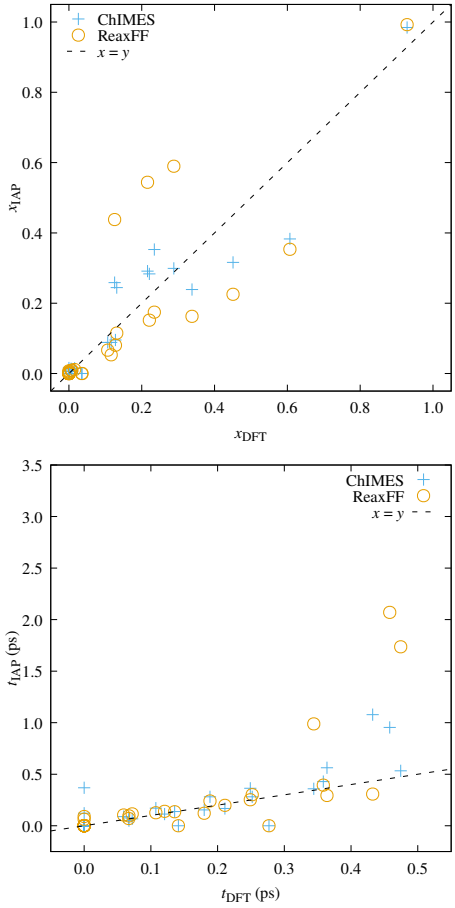

FIG. 14: Comparison of mole fractions (top) and lifetimes (bottom) for small species formed from DNTF over the range of training thermodynamic conditions, predicted by DFT, an existing DNTF reactive IAP (ReaxFF [46–49]), and the ChIMES IAP.

and pressures. It is interesting that the ReaxFF model yields a good material equation of state despite incorrect system structure and speciation predictions for this material. Essentially, the system structure and speciation are sensitive to specific many-body interactions, while the equation of state is an average over all interactions in the entire system. This behavior is likely caused by the analytical form underlying ReaxFF (i.e. the shape of the repulsive parts of the IAP); since it was not originally designed for materials under high pressure conditions, DFT pressure recovery comes at the expense of other properties (e.g. structure and chemistry).

While the present test model is not yet "quantum accurate," it already out-performs existing options (including standard semi-empirical quantum methods like DFTB, as will be shown in the next section). This result is very encouraging as many decisions (i.e. hyperparameter choices) during this fitting effort were intentionally made naively, since this study represents a conscious effort to expedite future ChIMES model development efforts for EM by aiming to establish "standard operating procedures," hence models were fit entirely through the automated active learning framework). However, we note that ChIMES model accuracy can be systematically improved by increasing model bodiedness and/or polynomial order, and model performance can be fine-tuned by through fit weighting, which is the focus of future efforts.

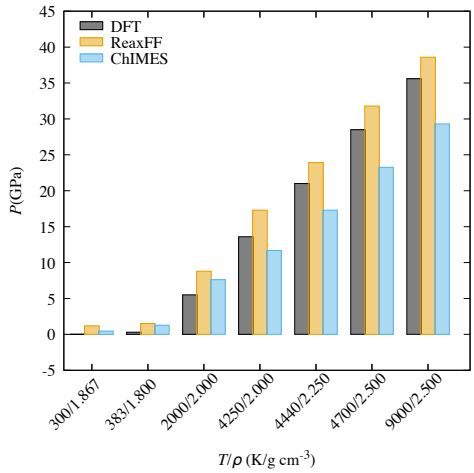

FIG. 15: DNTF pressures predicted by DFT, an existing DNTF reactive IAP (ReaxFF [46–49]), and the ChIMES IAP at thermodynamic states spanning ambient to shock compressed conditions.

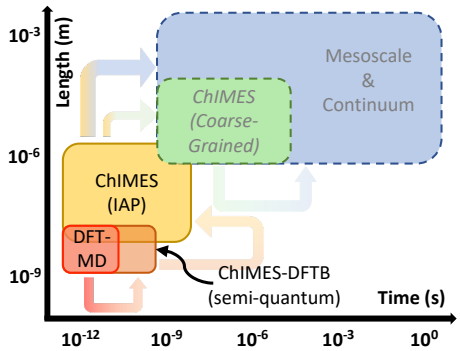

FIG. 16: Schematic of the ChIMES multiscale framework. Current efforts are focusing on generation of semi-empirical quantum ChIMES models (DFTB-ChIMES) which are trained to DFT and can be used as a DFT proxy to accelerate training data generation for ChIMES IAP development. ChIMES IAPs can be used to generate reference data for mesoscale and continuum models (e.g. thermochemistry, equation of state, etc). Future efforts will focus on establishing a particle-based coarse-grained ChIMES capability enabling reference data generation for properties requiring time and length scales impractical for ChIMES IAP simulation (e.g. material response in shocked systems containing extended defect structures, etc).

C. Multiscale ChIMES Model Development

As was discussed in above, few experimental data are available for DNTF, which complicates model development (e.g. determining relevant thermodynamic state points for training data generation). While the thermochemical predictions described in Section V B 1 were helpful in guiding these efforts, Cheetah models themselves must make assumptions when the experimental and/or simulation data generally used to fine-tune them are unavailable. Moreover, Cheetah predictions assume the system is some combination of the fully-unreacted material, carbon condensates, and small molecules possessing low heats of formation, which DFT-MD predictions may not, due to finite time and size effects (i.e. DFT- based MD simulations often settle in a metastable post-detonation state containing larger molecular species). Additionally, it is critical that the model training set includes information on the peak density/temperature the target material will reach at a given shock speed, which is difficult to determine *a priori*.

To overcome this challenge, we are establishing a multiscale ChIMES capability (see Fig. 16). In this framework, Cheetah calculations can be used to guide generation of an initial training repository for materials lacking previous experimental and/or simulation data, which can be used to develop a semi-empirical quantum (i.e. DFTB-) ChIMES model. This approach is feasible because of both the pre-tabulated Hamiltonian underlying the DFTB method and use of self-consistent charges, which significantly reduce complexity of the ChIMES correction required to recover quantum accuracy (i.e. compared to a standalone ChIMES-IAP) and by consequence, the training data required. Ultimately, this allows the DFTB-ChIMES model to be used as a proxy for DFT, making it possible to:

1. Run MSST simulations using a quantum-based approach (i.e. allowing identification of peak temperatures and densities and training data generation).

2. Efficiently generate training configurations from much longer simulations than possible with DFT.

3. Begin exploring the material EOS, thermochemistry, and even phenomena including carbon condensation while the relatively slower ChIMES-IAP fitting process proceeds.

In the following sections, we will discuss development of a DFTB-ChIMES model and compare its performance with existing DFTB IAP options, showing we can rapidly generate quantum-accurate models. Following, we show results from MSST simulations using this model, which:

1. Help identify the final T/p range to be targeted by for ChIMES IAP development,

2. Yield shock Hugoniot predictions,

3. Show chemical kinetic predictions from this model, the latter two of which can be immediately used to refine the existing Cheetah model for DNTF,

4. Present preliminary results related to formation of carbon condensates from shock compressed DNTF.

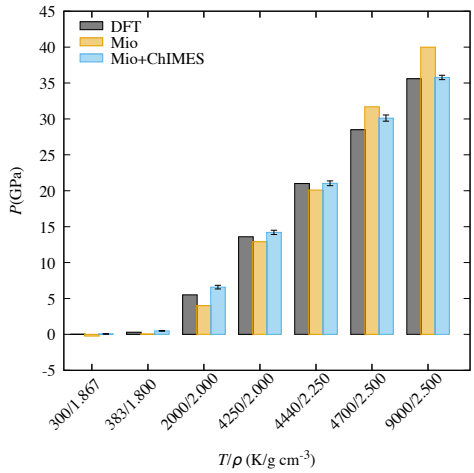

FIG. 17: DNTF pressures predicted by DFT, standard DFTB (Mio), and the DFTB-ChIMES model (Mio+ChIMES) at thermodynamic states spanning ambient to shock compressed conditions.

1. *Model generation and Validation*

The DFTB-SCC formalism has been discussed in detail elsewhere [50–52]. Briefly, the method assumes neutral, spherically symmetric charge densities on the atoms and expands the DFT Hamiltonian to second-order in charge fluctuations. The DFTB total energy is expressed as

$$E_{\mathrm{DFTB}} = E_{\mathrm{BS}} + E_{\mathrm{Coul}} + E_{\mathrm{corr}}, \tag{9}$$

where E_{BS} corresponds to the band structure energy, E_{Coul}, is a self-consistently computed charge fluctuation term, and E_{corr} is an empirical correction term accounting for ion–ion repulsions and Hartree and exchange-correlation double counting terms. The computational efficiency of DFTB is primarily realized through E_{BS}, which is taken as a sum over occupied electronic states from the approximate DFTB Hamiltonian. E_{BS} is usually computed from pre-tabulated Slater–Koster tables derived from DFT calculations with a minimal basis set, giving rise to an orders-of-magnitude increase in computational efficiency for a given system size, relative to DFT. Numerous DFTB Slater-Koster and E_{corr} parameter sets are available (i.e. at http://www.dftb.org), but these models are typically intended for biological systems (i.e. ambient T and p) or inorganic materials. Below, we will describe how the Chebyshev Interaction model for Efficient Simulation (ChIMES) can be used to tune E_{corr} for an existing DFTB model [8] (Mio-1-1), for description of DNTF under extreme T and p. We note that the ChIMES E_{corr} model developed in the present work includes 2 and 3-body interactions with respective polynomial orders of 12 and 7, affording a balance between accuracy and efficiency. All other ChIMES hyperparameters were set using protocols established in previous work [11–13].

The initial model was generated by training to 20 configurations from each of the DFT simulations described in Section V B 1. The resulting model exhibited excellent recovery of DFT structure, equation of state, and chemical kinetics and was stable when extended out to 10ps, but when leveraged in MSST simulations, was found to become unstable (i.e. substantial deviation in the conserved quantity was observed) due to exploration of temperature/density states more extreme than those explored in the original DFT training set. To correct this, a single iterative refinement was performed, where several configurations from the DFTB-ChIMES simulations were sent to DFT for single point evaluation, and combined with the training set from which a new model was generated. This model was found to be stable for the duration of the simulations.

Figures 17 and 18 provide an overview of DFTB-ChIMES performance relative to DFT and a standard DFTB model. In general, we find that the standard DFTB model yields a reasonable material equation of state, but poor speciation, much like the ReaxFF model described in Section V B 2. DFTB-ChIMES yields an improved material equation of state, and substantially improves upon predicted small molecule mole fractions and lifetimes. Ultimately, we find that DFTB-ChIMES models are quantum-accurate and can thus be viewed as a highly efficient proxy for DFT on longer time scales.

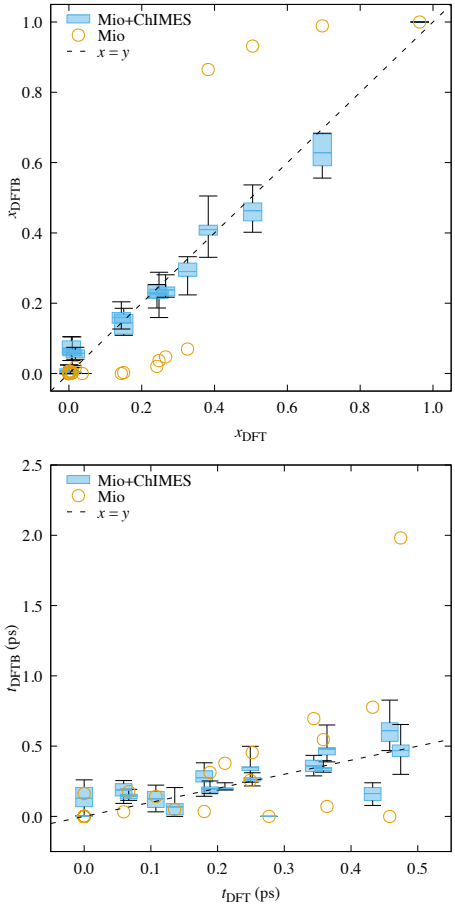

FIG. 18: Comparison of mole fractions (top) and lifetimes (bottom) for small species formed in DNTF over the range of training thermodynamic conditions, predicted by DFT and standard DFTB (Mio), and the DFTB-ChIMES model (Mio+ChIMES). The large error bar in species lifetime at t_{DFT} arise from the simulation at 2000 K/2.0 g cm^{-3}, where chemistry varies significantly across independent simulations. Boxplots are used to convey the DFTB-ChIMES data, and provide the max/min values (indicated by the error bars), 1st and 3rd quartiles (indicated by the range of the shaded box), and the average value (indicated by the horizontal line within each box).

2. *Hugoniot Equation of State Prediction*

As described in Section II, the relatively high efficiency exhibited by DFTB for a given system size makes increased statistical sampling feasible (i.e. through multiple independent simulations) and lends the method to application in longer time-scale simulation. Here, we use the presently developed DFTB-ChIMES model to predict the DNTF C-J detonation velocity and shock Hugoniot equation of state, and in doing so, generate training configurations for refining the ChIMES-IAP that would otherwise be unattainable through standard DFT approaches.

A total of 48 DFTB-ChIMES MSST simulations were run (e.g. 8 independent simulations at 6 shock velocities). Fig. 19 provides time evolution of a single independent simulation at each shock velocity, and shows that the predicted material C-J detonation velocity falls between 8.00 and 8.25 km s^{-1}. Fig. 20 shows the shock Hugoniot extracted from DFTB-ChIMES trajectories like those shown in 19, compared to the equation of state predicted by Cheetah. The equation of state for the unreacted material is in good agreement between the two methods (within error), but there is significant disagreement in the reacted Hugoniots for the two methods. These differences arise due to finite time and size effects in the DFTB-ChIMES MSST simulations, which allow the system to settle in a metastable state containing a mixture of the small molecule products predicted by Cheetah and larger species which will be shown in Section V C 3 to seed formation of carbon condensates.

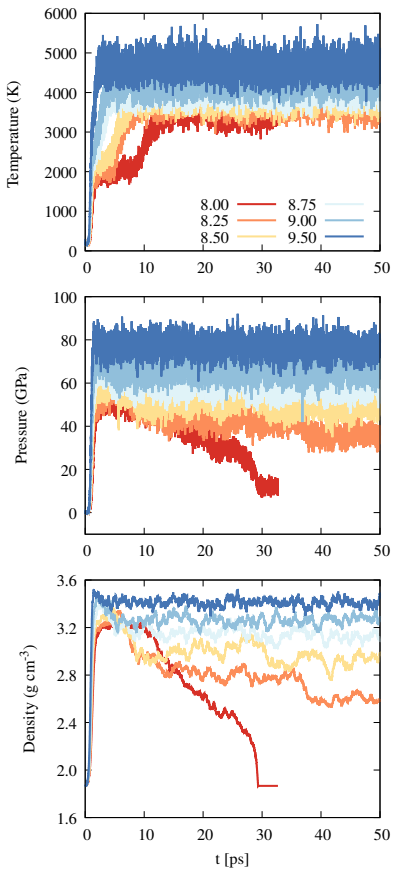

FIG. 19: Time evolution of temperature (top) pressure (middle), and density (bottom) for DNTF MSST simulations using the ChIMES DFTB model. Shock velocities are given by the key in km s^{-1}.

3. *Thermochemistry and Carbon Condensation*

The DFTB-ChIMES MSST simulations also provide a detailed view of chemical evolution in DNTF during during shock compression. As shown in Fig. 21, CO_2 and N_2 are the primary small molecules formed during shock compression, and a significant amount of other larger $C_xN_yO_z$ material is formed. This result is consistent with experimental observation of carbon condensates from shock-compressed DNTF [37], since the present results indicate O- and N-elimination from larger species occurs more rapidly than C-elimination. Table II provides an overview of these kinetics, by fitting the CO_2 and N_2 trajectories to an approximate rate equation of the form $x_{i,t} = x_{i,\infty}(1 - e^{-k_i \Delta t})$, where $x_{i,t}$ and $x_{i,\infty}$ are the mass fraction of species i at time t and at nominal equilibrium, k_i is the effective rate constant, $\Delta t = t - t_0$, and t_0 is the time point during the MSST trajectory at which chemistry begins. We note that DFTB-ChIMES rates at low shock velocity are consistent with those reported from simulation studies in HMX (i.e. 1,3,5,7-Tetranitro-1,3,5,7-tetrazoctan) under similar conditions [53].

Though DFTB-ChIMES simulations are small (i.e. approximately 200 - 300 atoms), there is sufficient $C_xN_yO_z$ material to begin investigating how carbon condensates evolve during the expansion and thermal quenching that follows shock compression. Fig. 22 shows how the system evolves during 50 ps simulations in which the system is adiabatically expanded from densities of approximately 2.6-3.25 g cm^{-3} (i.e. for systems compressed at 8.25 - 9.00 km s^{-1}) to 1 g cm^{-3}, leading to temperature decreases from approximately 3500-4250 to 1750-1900 K and pressure decreases of 38-65 to 1 GPa. We find that systems starting from higher shock velocities (i.e. which Fig. 21 indicates contains more $C_xN_yO_z$ material) tends to yield larger nominal clusters. These clusters are generally comprised of chains of alternating C and N atoms with occasional branch points terminated by oxygen. As time progresses, these chains begin to "zip" *via* elimination of oxygen, forming fused heterocycles (e.g., see snapshots in Fig. 22 labeled 8.75 and 9.00 km/s). This process is found to correlate with the slow formation of CO_x species, as opposed to O_2 or NO_x.

Overall, results from this study indicate that DFTB-ChIMES simulations can be used to efficiently explore chemistry,

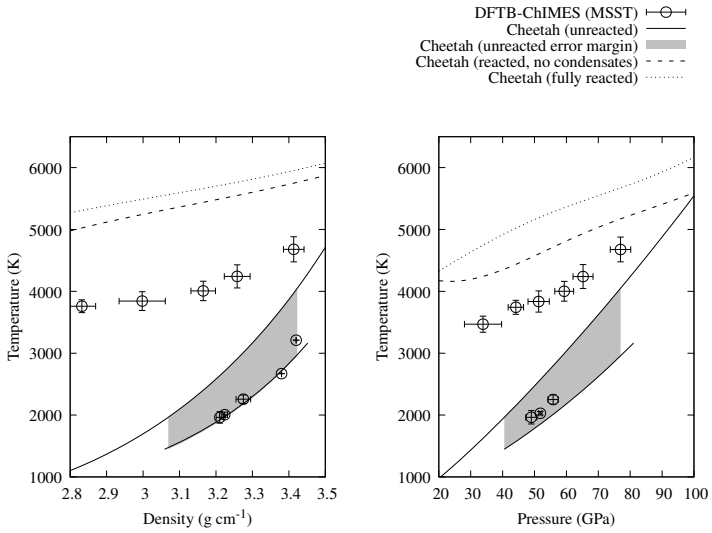

FIG. 20: DNTF shock Hugoniot determined from the DFTB-ChIMES MSST simulations compared with Cheetah predictions for a fully- and partially- reacted system, and upper/lower-bound estimates for the un-reacted system. The shaded region between the two solid unreacted provides an approximate error in the Cheetah prediction for the unreacted system.

the equation of state, and even preliminary carbon condensation for EM, generating reference data for thermochemical model development. Moreover, these simulations enable multiscale development of ChIMES-IAPs by generating informative training data (i.e. for the shock compressed and expanded system) that would otherwise be infeasible *via* DFT alone. Ultimately, the DFTB-ChIMES approach can be viewed as a proxy for DFT which allows us to extend accessible timescales by up to two orders of magnitude.

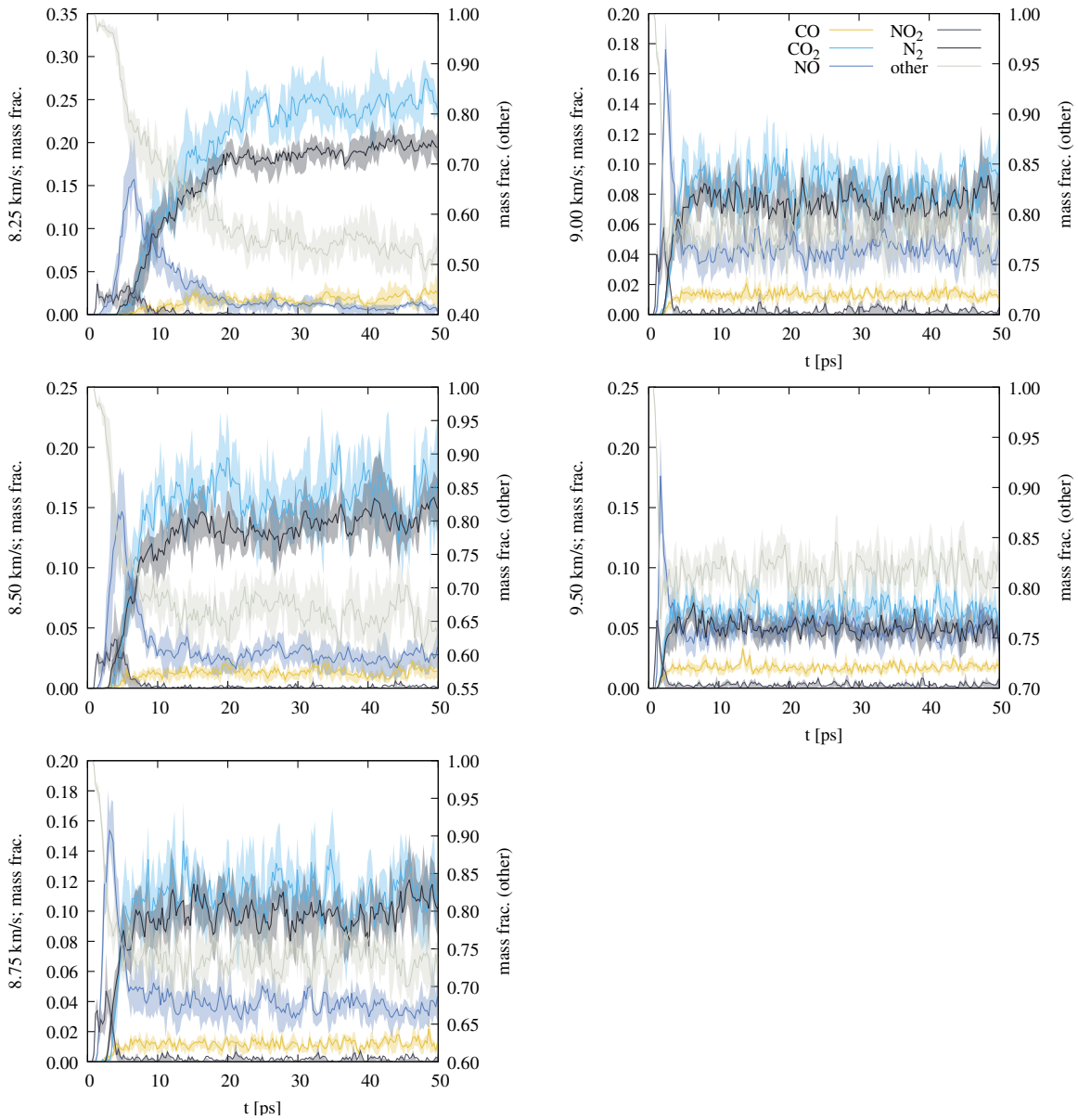

FIG. 21: Chemical evolution DNTF compressed at different shock velocities, predicted by the DFTB-ChIMES model. Shading gives the 1st and 3rd quartiles and solid line provide averages across independent simulations.

TABLE II: Kinetics for N_2 and CO_2 formation (i.e. dominant small molecule species) from shock compressed DNTF predicted from DFTB-ChIMES MSST simulations. Values for HMX [53], which were obtained from previous simulation studies, are provided for comparison.

	N_2		CO_2	
u_s (km s^{-1})	x_∞	k (ps^{-1})	x_∞	k (ps^{-1})
8.25	0.206	0.09	0.264	0.09
8.50	0.141	0.18	0.164	0.19
8.75	0.102	0.24	0.114	0.25
9.00	0.076	0.35	0.087	0.35
9.50	0.076	0.39	0.065	0.57
HMX	–	0.08	–	0.05

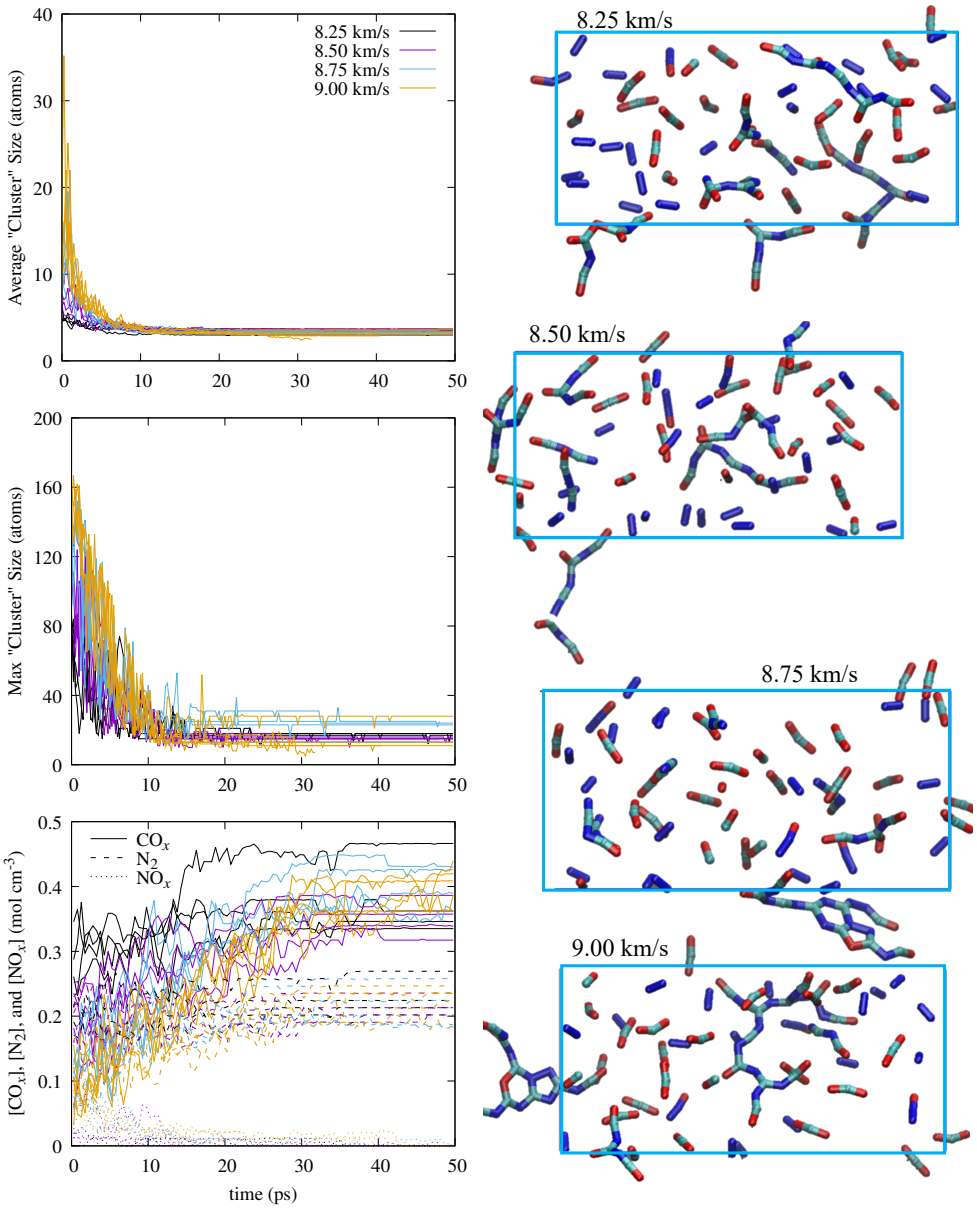

FIG. 22: Left: Evolution of average and maximum cluster sizes and $CO_x/N_2/NO_x$ concentrations during adiabatic expansion following MSST simulations. Line colors indicate shock velocity used in the prior MSST simulation, and multiple lines of the same color/dash type correspond to independent simulations initiated from the same shock velocity. Right: sample snapshots from the expanded systems. Cyan, red, and blue atoms correspond to carbon, oxygen, and nitrogen, respectively.

VI. AREAS FOR IMPROVEMENT

In this section we summarize our findings regarding the ChIMES model when applied to reacting EM at extreme conditions. For HN_3, we were able to obtain a high degree of accuracy using a ChIMES 4-body model, including configurations from condensed phase simulations, gas phase clusters extracted from the condensed phase, and random displacements of the gas phase clusters. The generation of the model, however, took more than two years of dedicated effort. Some of this is related to one-time research undertakings, such as development of the four-body interaction model, and the development of the cluster analysis method. In the future we would like to be much more agile in the development of IAPs. The development of more sophisticated training methods, such as the active learning approach, should accelerate IAP model development in the future. Ongoing work has also shown that the specific weighting scheme used during the fitting process can significantly impact performance of the resulting model, thus we are currently working to optimize and standardize our weighting approach. For the more complicated chemistry of DNTF, the model to date has challenges in quantitatively reproducing both chemistry and equation of state properties such as the pressure. Improved pressures could be obtained by developing an optimal balance between matching bulk properties like stresses and isolated molecule properties such as cluster energies.

It is likely that the higher coordination numbers common in carbon-containing compounds, as compared to nitrogen compounds like HN_3 requires a higher bodiedness than 4. This will be explored in the future with extensions to the current ChIMES implementation. It could be challenging to maintain numerical efficiency with high bodiedness. We believe that a GPU implementation of ChIMES could significantly accelerate calculations and make models with higher bodiedness affordable. There are also improvements possible in the evaluation of permutationally invariant polynomials that our computer codes do not yet take advantage of.

With higher bodiedness also comes more potential parameters for the LASSO algorithm to evaluate. Our current parallel LASSO algorithm works well for up to about 5,000 non-zero parameters, but further improvements are necessary to scale to many more parameters. The non-smooth nature of the LASSO objective function and the poor condition number of our design matrix preclude the use of generic scalable optimization methods. Our parallel LARS/LASSO implementation, however, could be made more scalable by distributing linear algebra solutions over multiple MPI processes. In addition, more model complexity will require better training set selection procedures. In addition to the active learning algorithms presented here, and systematic generation of molecular clusters is another way to efficiently generate training data that sample a broad range of environments.

VII. CONCLUSIONS

The goals of this work were to demonstrate the suitability of the ChIMES framework for atomistic description of EM. The significant challenges associated with the study of EM makes this a formidable goal. For example,

1. EM have a wide range of atom types and environments, and exhibit a complicated manifold of chemistry.
2. Shock initiation can be governed by interplay between material structure, strength, phase diagram, thermochemistry and chemical kinetics, requiring accurate prediction of many properties in many environments from the IAP.
3. Carbon coagulation involves chemistry during post-shock expansion, which is another unique environment with chemistry occurring on many ns.

From the standpoint of atomistic modeling, the practical implications are that quantum accurate models capable of describing chemistry, structure, and dynamics under both ambient and extreme thermodynamic conditions on time and length scales approaching 1 us and 1 um, respectively, are needed. When this capability is achieved, the direct atomistic simulation of energetic materials initiation processes will be possible, enabling the development of more accurate continuum models for HE shock response.

For many years, atomistic simulations of EM have relied on one of: DFT (i.e. limiting problem space to $\mathcal{O}$ 100 atoms and 10 ps), DFTB (extending problem space to 100s of ps with less accuracy than DFT), or classic reactive IAP methods (which are less accurate than DFTB). Advances in ML are allowing us to begin overcoming these limitations through development of ML-IAPs retaining the accuracy of first-principles at the computational expense of classic reactive IAPs. Though this is an active area of research, ChIMES is the only ML-IAP model specifically targeting this challenging problem, to the best of our knowledge.

Through the HN_3 work described in Section IV, we have shown that quantum-accurate modeling of the detonation state of simple EM is possible through development of robust and temperature/pressure transferable ChIMES models through non-active fitting approaches. At the same time, the many orders of magnitude increase in efficiency afforded

by ChIMES allows application to large-scale direct shock simulations, a mainstay of atomistic EM research and development.

In Section V A, we discussed how the non-active ChIMES fitting approach can break down when applied to carbon-containing species (i.e. which undergo convoluted chemistry and can yield complicated species including large molecules and carbon condensates), when subject to extreme conditions. Two new ChIMES approaches were developed to overcome these challenges, including (1) a new active-learning approach which greatly reduced the human effort required to generate ChIMES models and which increases reproducibility, and (2) ChIMES-corrected semi-empirical quantum models (i.e. DFTB) for EM which can be viewed as a proxy for DFT on orders-of-magnitude longer timescales, enabling efficient generation of meaningful ChIMES training data, and preliminary exploration of material equation of state, chemical kinetics, and even carbon coagulation. These two advances are being combined to establish a multiscale ChIMES capability.

Section V B showed preliminary results from application of the active learning framework to DNTF, where the resulting ChIMES IAP was shown to outperform both standard reactive models and standard semi-empirical quantum DFTB models for prediction of system structure and chemistry relative to DFT, and Section V C demonstrated development of a quantum accurate ChIMES DFTB model and its application for predicting the Hugoniot equation of state of DNTF, chemical kinetics in the shocked systems, and a preliminary exploration of carbon condensation during post-shock expansion. In doing so, each of our L2 milestone goals were achieved, i.e. to develop, validate, and document a new reactive molecular dynamics method for EM, based on machine learning by:

1. Generating databases of first-principles- (e.g. DFT) derived forces, stresses, and energies for HN_3 and DNTF.
2. Generating atomistic force fields from these databases *via* ML.
3. Benchmarking model performance against first principles calculations.

The efforts documented in the report have positioned us to generate highly-accurate and efficient atomistic models for energetic materials. We note that the Laboratory Directed Research and Development program was instrumental in establishing ChIMES. Specifically, initial ChIMES development (i.e. which included models for molten carbon and liquid water) were funded by LLNL LDRD 16-LW-020, while development a carbon monoxide model and its subsequent application to study carbon nucleation in the shock compressed carbon monoxide were funded by LLNL LDRD 17-ERD-011. These efforts paved the way for development of models and tools for substantially more complex systems, such as those discussed herein. Looking forward, we will: (1) extend ChIMES to increasingly complex EM such as TATB and LLM-105, and (2) continue improvement of ChIMES and our ML approaches. We are particularly interested in determining the mechanism of carbon cluster (soot) formation in TATB and LLM-105, since we believe that the accuracy of the Cheetah-based detonation models used in continuum simulations could be substantially improved with a more accurate carbon cluster model. Carbon condensation simulations in these materials will necessarily be large (i.e. 10s to 100s of millions of atoms) and long (i.e. 100s of ns) and thus require access to large high-performance computing (HPC) allocations such as those provided through the Computing Grand Challenge program. This is in contrast to HN_3, for which rapid chemistry and small characteristic length scales allow converged predictions from standard computer allocation. DNTF also undergoes rapid chemistry but exhibits large characteristic length scales (i.e. due to formation of detonation soot), and thus will similarly necessitate a large HPC allocation. We note our current ChIMES implementation can, in principle, achieve capability-type simulations on CPU-based systems (e.g. Quartz at Livermore Computing). Our implementation does not currently support simulation on GPU or Phi architectures, but efforts toward that goal are underway.

Experimental validation of our ChIMES models is critical to these efforts since our training data correspond to time and length scales orders of magnitude smaller than those we plan to use the models on (e.g. less than 10 ps and nm for training versus 100s of ns and nm for our target simulations). Our initial ChIMES model for carbon monoxide was validated *via* ultra-fast laser-driven shock experiments and subsequent transmission electron microscopy, which allowed us to confirm that the size, shape, and composition of carbon condensates observered in our ChIMES simulations, as well as the timescales over which they formed were accurate (LLNL LDRD 17-ERD-011) [13, 27]. We note that recent time-resolved, small-angle X-ray scattering experiments [37] provide similar data (e.g. condensate sizes, shapes, and approximate composition and formation timescales) against which our ChIMES DNTF model can be validated. Overall, continued experiments of this nature are critical to ChIMES validation efforts.

Toward our first goal (i.e. ChIMES extension to increasingly complex materials), we will begin developing a model for mixtures of hydrogen peroxide and nitromethane in FY'21 and a model for TATB in FY'21 and FY'22. The former mixtures provide a testbed for investigating the feasibility of a universal ChIMES training repository and transferable parameter sets (i.e. describing C, H, O, and N-containing systems), both of which would greatly reduce the model development time for future ChIMES simulation studies involving new EM. The TATB efforts will target description of the formation of carbon condensates in TATB, which presents a number of challenges that our L2 milestone efforts have positioned us to overcome. For example, the material comprises a large number of unique

atom types (i.e. C, H, O, N) and thus exhibits a complex manifold of decomposition chemistry, description of which requires high-complexity models and large, informative training datasets. The extensibility of ChIMES demonstrated in section IV A in combination with our distributed linear solver has ensured we can construct models of the necessary complexity, while development of our active learning tool has ensured appropriate training sets can be generated efficiently and generated models are robust and consistent. Additionally, TATB is inherently multiscaled (e.g. it is known to exhibit defect-driven sensitization and form nm-scale carbon condensates on 100s of ns timescales). This poses challenges for both training data generation (e.g. because DFT cannot access the relevant timescales) and model development (i.e. because resulting models must achieve a balance of accuracy and efficiency to enable description of atom-scale reactive phenomena driving material evolution on scales approaching 1 μm and μs. We have overcome the DFT training data generation challenge by developing a capability for producing DFTB-ChIMES models (i.e. offering the predictive power of DFT on significantly longer timescales), and have enabled development of accurate yet efficient models by use of the LASSO approach in our distributed linear solver (i.e. which automatically removes unnecessary model parameters, significantly reducing model size and thus computational expense). Nevertheless, we anticipate that TATB model generation will require additional developmental efforts.

Remaining challenges surrounding development of ChIMES models for increasingly complex EM will be overcome through our second thrust, i.e. extension of our active learning framework and further improvement of ChIMES to make higher levels of accuracy possible. We are investigating automatic methods for cluster generation to parameterize ChIMES, and in addition we are exploring increasing the maximum bodiedness in the ChIMES-MD code while also adding explicit treatments of charge transfer and metallic many-body interactions to the model. Future efforts will also explore more sophisticated tools for data management, to facilitate repeatability and data reuse.

Acknowledgments

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

---

[1] R. Menikoff, in *AIP Conf. Proc.* (American Institute of Physics, 2009), vol. 1195, pp. 18–25.
[2] S. Bastea and L. E. Fried, J. Chem. Phys. **128**, 174502 (2008).
[3] J. A. Hammons et al., J. Phys. Chem. C **123**, 19153 (2019).
[4] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[5] N. Goldman, E. J. Reed, I.-F. W. Kuo, L. E. Fried, C. J. Mundy, and A. Curioni, J. Chem. Phys. **130**, 124517 (2009).
[6] N. Goldman, E. J. Reed, and L. E. Fried, J. Chem. Phys. **131**, 204103 (2009).
[7] N. Goldman, L. E. Fried, and L. Koziol, J. Chem. Theory Comput. **11**, 4530 (2015).
[8] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Fraunheim, S. Suhai, and G. Seifert, Phys. Rev. B **58**, 7260 (1998).
[9] M. A. Wood, M. J. Cherukara, E. M. Kober, and A. Strachan, J. Phys. Chem. C **119**, 22008 (2015).
[10] M. A. Wood, D. E. Kittell, C. D. Yarrington, and A. P. Thompson, Phys. Rev. B **97**, 014109 (2018).
[11] R. K. Lindsey, L. E. Fried, and N. Goldman, J. Chem. Theory Comput. **13**, 6222 (2017).
[12] R. K. Lindsey, L. E. Fried, and N. Goldman, J. Chem. Theory Comput. **15**, 436 (2019).
[13] R. K. Lindsey, N. Goldman, L. E. Fried, and S. Bastea, The Journal of Chemical Physics **153**, 054103 (2020).
[14] N. Goldman, B. Aradi, R. K. Lindsey, and L. E. Fried, J. Chem. Theory Comput. **14**, 2652 (2018).
[15] N. R. Greiner, D. Phillips, J. Johnson, and F. Volk, Nature **333**, 440 (1988).
[16] V. N. Mochalin, O. Shenderova, D. Ho, and Y. Gogotsi, Nature nanotechnol. **7**, 11 (2012).
[17] M. Bagge-Hansen, L. Lauderbach, R. Hodgin, S. Bastea, L. Fried, A. Jones, T. van Buuren, D. Hansen, J. Benterou, T. May, C aand Graber, et al., J. Appl. Phys. **117**, 245902 (2015).
[18] A. P. Thompson, L. P. Swiler, C. R. Trott, S. M. Foiles, and G. J. Tucker, J.Comput. Phys. **285**, 316 (2015).
[19] J. S. Smith, B. Nebgen, N. Lubbers, O. Isayev, and A. E. Roitberg, J. Chem. Phys. **148**, 241733 (2018).
[20] F. Ercolessi and J. B. Adams, Europhys. Lett. **28**, 583 (1994).
[21] R. Tibshirani, J. R. Statist. Soc. B **58**, 267 (1996).
[22] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, Ann. Stat. **32**, 407 (2004).
[23] J. Friedman, T. Hastie, and R. Tibshirani, J. Stat. Softw. **33**, 1 (2010).
[24] S. Plimpton, Journal of computational physics **117**, 1 (1995).
[25] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al., Journal of Machine Learning Research **12**, 2825 (2011).
[26] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, et al., Nature Methods **17**, 261 (2020).

[27] M. R. Armstrong, R. K. Lindsey, N. Goldman, M. H. Nielsen, E. Stavrou, L. E. Fried, J. M. Zaug, , and S. Bastea, Nat. Commun. **11**, 1 (2020).
[28] E. J. Reed, A. W. Rodriguez, M. R. Manaa, L. E. Fried, and C. M. Tarver, Phys. Rev. Lett. **109**, 038301 (2012).
[29] C. H. Pham, R. K. Lindsey, L. E. Fried, and N. Goldman, Submitted: J. Chem. Phys. (2020).
[30] E. J. Reed, L. E. Fried, and J. Joannopoulos, Phys. Rev. Lett. **90**, 235503 (2003).
[31] E. J. Reed, L. E. Fried, W. D. Henshaw, and C. M. Tarver, Phys. Rev. E **74**, 056706 (2006).
[32] M. R. Armstrong, J. M. Zaug, N. Goldman, I.-F. W. Kuo, J. C. Crowhurst, W. M. Howard, J. A. Carter, M. Kashgarian, J. M. Chesser, T. W. Barbee, et al., J. Phys. Chem. A **117**, 13051 (2013).
[33] N. Goldman and S. Bastea, J. Phys. Chem. A **118**, 2897 (2014).
[34] R. K. Kurbangalina, E. Patskov, L. Stesik, and G. Yakovleva, J. Appl. Mech. Tech. Phys. **11**, 672 (1970).
[35] G. Yakovleva and R. K. Kurbangalina, Combust. Explos. Shock Waves **12**, 692 (1976).
[36] F. Bowden, M. F. R. Mulcahy, R. G. Vines, and A. Yogge, Proc. Roy. Soc. of London Series A **188**, 291 (1947).
[37] M. Bagge-Hansen, S. Bastea, J. A. Hammons, M. H. Nielsen, L. Lauderbach, R. Hodgin, P. Pagoria, C. May, S. Aloni, A. Jones, et al., Nat. Commun. **10**, 1 (2019).
[38] R. K. Lindsey, L. E. Fried, N. Goldman, and S. Bastea, Just Accepted: J. Chem. Phys.; https://doi.org/10.26434/chemrxiv.12660083.v1 (2020).
[39] S. L. Frederiksen, K. W. Jacobsen, K. S. Brown, and J. P. Sethna, Phys. Rev. Lett. **93**, 165501 (2004).
[40] V. Botu and R. Ramprasad, Int. J. Quantum Chem. **115**, 1074 (2015).
[41] E. V. Podryabinkin and A. V. Shapeev, Comput. Mater. Sci. **140**, 171 (2017).
[42] L. Zhang, D.-Y. Lin, H. Wang, R. Car, and E. Weinan, Phys. Rev. Mater. **3**, 023804 (2019).
[43] B. Settles, Tech. Rep., University of Wisconsin-Madison Department of Computer Sciences (2009).
[44] C. E. Shannon, Bell Labs Tech. J. **27**, 379 (1948).
[45] V. P. Sinditskii, A. V. Burzhava, A. B. Sheremetev, and N. S. Aleksandrova, Propellants Explos. Pyrotech. **37**, 575 (2012).
[46] A. Strachan, A. C. van Duin, D. Chakraborty, S. Dasgupta, and W. A. Goddard III, Phys. Rev. Lett. **91**, 098301 (2003).
[47] A. Strachan, E. M. Kober, A. C. van Duin, J. Oxgaard, and W. A. Goddard III, J. Chem. Phys. **122**, 054502 (2005).
[48] L. Zhang, A. C. v. Duin, S. V. Zybin, and W. A. Goddard III, J. Phys. Chem. B **113**, 10770 (2009).
[49] L. Zhang, S. V. Zybin, A. C. van Duin, S. Dasgupta, W. A. Goddard III, and E. M. Kober, J. Phys. Chem. A **113**, 10619 (2009).
[50] P. Koskinen and V. Mäkinen, Comput. Mater. Sci. **47**, 237 (2009).
[51] M. Gaus, Q. Cui, and M. Elstner, J. Chem. Theory Comput. **7**, 931 (2011).
[52] B. Aradi, B. Hourahine, and T. Frauenheim, J. Phys. Chem. A **111**, 5678 (2007).
[53] M. R. Manaa, L. E. Fried, C. F. Melius, M. Elstner, and T. Frauenheim, J. Phys. Chem. A **106**, 9024 (2002).